



OPEN ACCESS

EDITED BY

Carol H. Miao,
Seattle Children's Research Institute,
United States

REVIEWED BY

Fiji Shridhar,
Mayo Clinic, United States
Mark Nachtigal,
University of Manitoba, Canada

*CORRESPONDENCE

Kellie Dean,
✉ k.dean@ucc.ie

SPECIALTY SECTION

This article was submitted
to Cancer Cell Biology,
a section of the journal
Frontiers in Cell and
Developmental Biology

RECEIVED 21 November 2022

ACCEPTED 31 January 2023

PUBLISHED 13 February 2023

CITATION

McCabe A, Zaheed O, McDade SS and
Dean K (2023), Investigating the suitability
of *in vitro* cell lines as models for the
major subtypes of epithelial
ovarian cancer.
Front. Cell Dev. Biol. 11:1104514.
doi: 10.3389/fcell.2023.1104514

COPYRIGHT

© 2023 McCabe, Zaheed, McDade and
Dean. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Investigating the suitability of *in vitro* cell lines as models for the major subtypes of epithelial ovarian cancer

Aideen McCabe^{1,2}, Oza Zaheed^{1,2}, Simon Samuel McDade³ and Kellie Dean^{1*}

¹School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland, ²The SFI Centre for Research Training in Genomics Data Science, Galway, Ireland, ³The Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

Epithelial ovarian cancer (EOC) is the most fatal gynaecological malignancy, accounting for over 200,000 deaths worldwide per year. EOC is a highly heterogeneous disease, classified into five major histological subtypes—high-grade serous (HGSOC), clear cell (CCOC), endometrioid (ENOC), mucinous (MOC) and low-grade serous (LGSOC) ovarian carcinomas. Classification of EOCs is clinically beneficial, as the various subtypes respond differently to chemotherapy and have distinct prognoses. Cell lines are often used as *in vitro* models for cancer, allowing researchers to explore pathophysiology in a relatively cheap and easy to manipulate system. However, most studies that make use of EOC cell lines fail to recognize the importance of subtype. Furthermore, the similarity of cell lines to their cognate primary tumors is often ignored. Identification of cell lines with high molecular similarity to primary tumors is needed in order to better guide pre-clinical EOC research and to improve development of targeted therapeutics and diagnostics for each distinctive subtype. This study aims to generate a reference dataset of cell lines representative of the major EOC subtypes. We found that non-negative matrix factorization (NMF) optimally clustered fifty-six cell lines into five groups, putatively corresponding to each of the five EOC subtypes. These clusters validated previous histological groupings, while also classifying other previously unannotated cell lines. We analysed the mutational and copy number landscapes of these lines to investigate whether they harboured the characteristic genomic alterations of each subtype. Finally we compared the gene expression profiles of cell lines with 93 primary tumor samples stratified by subtype, to identify lines with the highest molecular similarity to HGSOC, CCOC, ENOC, and MOC. In summary, we examined the molecular features of both EOC cell lines and primary tumors of multiple subtypes. We recommend a reference set of cell lines most suited to represent four different subtypes of EOC for both *in silico* and *in vitro* studies. We also identify lines displaying poor overall molecular similarity to EOC tumors, which we argue should be avoided in pre-clinical studies. Ultimately, our work emphasizes the importance of choosing suitable cell line models to maximise clinical relevance of experiments.

KEYWORDS

Ovarian cancer, cell line, subtype classification, RNA-Seq—RNA sequencing, Non-negative matrix factorization (NMF), correlation

Introduction

Ovarian cancer (OC) is the most deadly cancer of gynaecological origin, with over 300,000 new cases and 200,000 deaths occurring in 2020 (Sung et al., 2021). In the United States alone, OC accounts for approximately 20,000 new cases and 13,000 deaths per year (Siegel et al., 2022). Epithelial ovarian cancer (EOC) represents the majority of ovarian malignancies (Jayson et al., 2014) and is further classified into five major subtypes—high grade serous (HGSOC), clear cell (CCOC), endometrioid (ENOC), mucinous (MOC) and low grade serous (LGSOC) ovarian carcinomas (Prat et al., 2018). These

subtypes differ in their genetic profiles, precursor lesions, response to therapy and clinical outcome (Table 1). For example, LGOSC, CCOC and MOC display poor responses to traditional platinum based chemotherapies, whereas ENOC and HGSOC display initial chemosensitivity (Sugiyama et al., 2000; Schmelzer and Gershenson, 2008; Gershenson et al., 2009; Mackay et al., 2010; Kim et al., 2012; Prat et al., 2018; Lheureux et al., 2019a; Babaier and Ghatage, 2020). Furthermore, the lack of *BRCA1/2* mutations and hormone receptor (HR) deficiencies in non-HGSOC/high grade ENOC subtypes limits the use of poly (ADP-ribose) polymerase (PARP) inhibitors in disease management (Lheureux et al., 2019a;

TABLE 1 Incidence, mutational landscape, precursor lesions, chemotherapy response, and prognoses of the five major subtypes of EOC.

	HGSOC	CCOC	ENOC	MOC	LGSOC	References
Incidence	70%	10%	10%	3%	<5%	Prat et al. (2018)
Commonly mutated genes	Near-ubiquitous <i>TP53</i> mutations	<i>ARID1A</i> most frequently mutated (50%–60% of cases)	<i>ARID1A</i> mutated in up to 30% of cases	<i>KRAS</i> and <i>NRAS</i> frequently mutated	<i>BRAF</i> , <i>KRAS</i> and <i>NRAS</i> commonly mutated	Shaw et al. (2002), Risch et al. (2006), Jones et al. (2010), The Cancer Genome Atlas Research Network, (2011), McAlpine et al. (2012), Emmanuel et al. (2014), Kanchi et al. (2014), Hunter et al. (2015), Friedlander et al. (2016), Itamochi et al. (2017), Maru et al. (2017), Murakami et al. (2017), Kim et al. (2018), Shibuya et al. (2018), Cybulska et al. (2019), Cheasley et al. (2019), Morice, et al. (2019), Gorringer et al. (2020), Pierson et al. (2020), Cheasley et al. (2022)
	<i>BRCA1/2</i> germline mutations or inactivation	<i>PIK3CA</i> , <i>KRAS</i> , <i>PTEN</i> , <i>SMARCA4</i> also mutated	<i>PIK3CA</i> , <i>KRAS</i> , <i>PTEN</i> , <i>CTNNB1</i> also mutated	Mutations in <i>ARID1A</i> and <i>PIK3CA</i>	<i>TP53</i> mutations rare	
		<i>TP53</i> mutations in up to 20% of cases	<i>TP53</i> mutations in ~17% of cases	Amplification of <i>ERBB2</i> and copy number loss of <i>CDKN2A</i>		
				<i>TP53</i> mutations in up to 64% of cases, often confined to later stages		
Precursor lesions	Serous tubal intraepithelial carcinoma (STIC) lesions in the fallopian tube	Endometriosis and retrograde menstruation, adenofibroma	Endometriosis and retrograde menstruation, adenofibroma	Mucinous benign cystadenomas and mucinous borderline ovarian tumors (MBTs)	Serous borderline tumors (SBTs) in 60%–80% of cases	Lee et al. (2007), Medeiros et al. (2006), Kurman and Shih, (2010), Wiegand et al. (2010), Erickson et al. (2013), Ahn et al. (2016), Prat et al. (2018), Cheasley et al. (2019), Shih et al. (2021)
Response to chemotherapy	Initially responsive (80%) but 20%–30% of tumors relapse 6 months after treatment and can develop resistance	Poor response (15%) and resistance to platinum chemotherapy noted	High grade ENOCs initially sensitive to platinum chemotherapy	Intermediate response (15%–60%) and resistance to platinum chemotherapy noted	Poor response (26%–28%)	Sugiyama et al. (2000), Gershenson et al. (2009), Alexandre et al. (2010), Berns and Bowtell, (2012), Kim et al. (2012), Prat et al. (2018), Lheureux et al. (2019a), Morice et al. (2019)
Clinical outcome	Poor survival rates due to overall later stage diagnosis and recurrent tumors after chemotherapy	Intermediate, with a greater risk of venous thromboembolism than other subtypes. Late stage CCOC has a poorer prognosis than late stage HGSOC	Low grade ENOC more prevalent and favourable, poorer outcomes in high grade ENOC cases	As 65%–80% are diagnosed at an early stage, survival is favourable, later stages have a poorer prognosis than late stage HGSOC	Most favourable outcomes of all subtypes, along with low-grade endometrioid carcinoma	Mackay et al. (2010), Berns and Bowtell, (2012), Peres et al. (2018), Prat et al. (2018), Ricciardi et al. (2018), Soyama et al. (2018), Morice et al. (2019)

Babaier and Ghatage, 2020). Targeted therapies aimed at exploiting subtype-specific mutations are currently being developed (Coward, et al., 2015; Lheureux et al., 2019a) but have yet to be widely applied clinically. As such, in order to develop novel and targeted therapies for HGSOC and non-HGSOC EOCs and to effectively study the pathogenesis of these heterogeneous subtypes, it follows that pre-clinical studies should make use of models that are both stratified by subtype and reflective of their respective primary tumors.

Cell lines are often utilized to study cancer biology, as a cheaper and less time-consuming alternative to patient derived explants or animal models. There exists over 100 OC cell lines, with around 70 commercially available through the American Type Culture Collection (ATCC), European Collection of Authenticated Cell Cultures (ECACC), Rikagaku Kenkyusho (Institute of Physical and Chemical Research, RIKEN), Leibniz Institute DSMZ, CellBank Australia (CBA) and/or the Japanese Collection of Research Biosources Cell Bank (JCRB) (Beaufort et al., 2014; Ciucci et al., 2022). Cancer cell lines are relatively cheap, easy to maintain, and facilitate rapid results compared to more complex organoid, animal and tumor-on-a-chip platforms (Ciucci et al., 2022). However, the utility of cell lines in translational and pre-clinical research has been questioned. Studies in liver cancer (Chen et al., 2015), breast cancer (Jiang et al., 2016) and in particular, ovarian cancer (Anglesio et al., 2013; Domcke et al., 2013; Beaufort et al., 2014; Barnes et al., 2021) have demonstrated the limited capabilities of cell lines in accurately representing their corresponding tumor types. Importantly, these studies highlight that this poor applicability of models is exacerbated when histological subtype is not taken into account, emphasizing the importance of choosing suitable cell lines to maximize translation into the clinic. Indeed, a number of clinical trials have failed due to a lack of consideration for EOC subtype in pre-clinical studies (Coward et al., 2011) and the need for subtype-specific research in EOC to improve drug development and patient outcome has been emphasized (Alvarez et al., 2016; Lheureux et al., 2019b). This lack of subtype-specific research has been perpetuated by a lack of accurate cell line annotations, with the origin and subtype of most OC cell lines being poorly defined. Furthermore, the similarity of these models to their respective primary tumors is often not investigated or taken into account.

Since cell lines do not possess the morphological traits necessary for histological subtyping, current studies utilize molecular characteristics in order to determine the subtypes of these EOC models. A seminal study by Domcke et al. (2013) observed striking molecular differences between the most commonly cited cell lines and HGSOC tumor samples. By comparing copy number changes, mutational landscape and gene expression between 47 EOC cell lines and thousands of HGSOC tumor samples, the authors recommend a set of cell lines that best represent HGSOC primary tumors. Yu et al. (2019) also correlated molecular data of cell lines and tumors for 22 cancer types, including HGSOC, identifying the cell lines most correlated to their respective primary tumors. Efforts have also been made to identify CCOC cell lines (Anglesio et al., 2013). In this study, Anglesio and colleagues combined a panel of immunohistochemical biomarkers and a predictive algorithm to establish a subset of well-suited CCOC cell lines, as well as analyzing some HGSOC, ENOC and MOC models. Taking a wet-lab based approach, Beaufort et al. (2014) characterized 39 EOC cell lines in

terms of gene expression and therapeutic response, partitioning cell lines in terms of both putative histological subtype and morphology. More recently, Barnes et al. (2021) expanded on the aforementioned work by utilizing non-negative matrix factorization (NMF) to separate 44 EOC cell lines into five groups that were suggested to correspond to all five EOC subtypes, expanding knowledge on cell line classification to other, non-HGSOC subtypes.

These studies have made remarkable progress in identifying the most suitable *in vitro* models to study different subtypes of ovarian cancer. However, there exists no study that compares the molecular profiles of EOC cell lines to primary tumor tissue of multiple EOC subtypes. Here, we integrate and analyse EOC cell line molecular data from multiple different sources, with the aim of generating a reference dataset of lines representative of the major EOC subtypes. We apply non-negative matrix factorisation (NMF) to a panel of 56 EOC cell lines. This resulted in five stable clusters that largely validated results obtained in previous studies, while also suggesting new subtype annotations for five previously unannotated cell lines. Additionally, genetic profiles of EOC cell lines were compared to 93 EOC primary tumor samples, from multiple independent datasets, stratified by subtype. We investigate the molecular similarity of cell lines to not only HGSOC primary tumors, but also CCOC, ENOC, and MOC. Ultimately, we identify the most and least representative cell lines to represent these EOC subtypes, generating a reference dataset for future *in silico* and *in vitro* studies. Consequently, our work highlights the need for the generation of additional, subtype specific datasets, in particular for the LGSOC and ENOC subtypes.

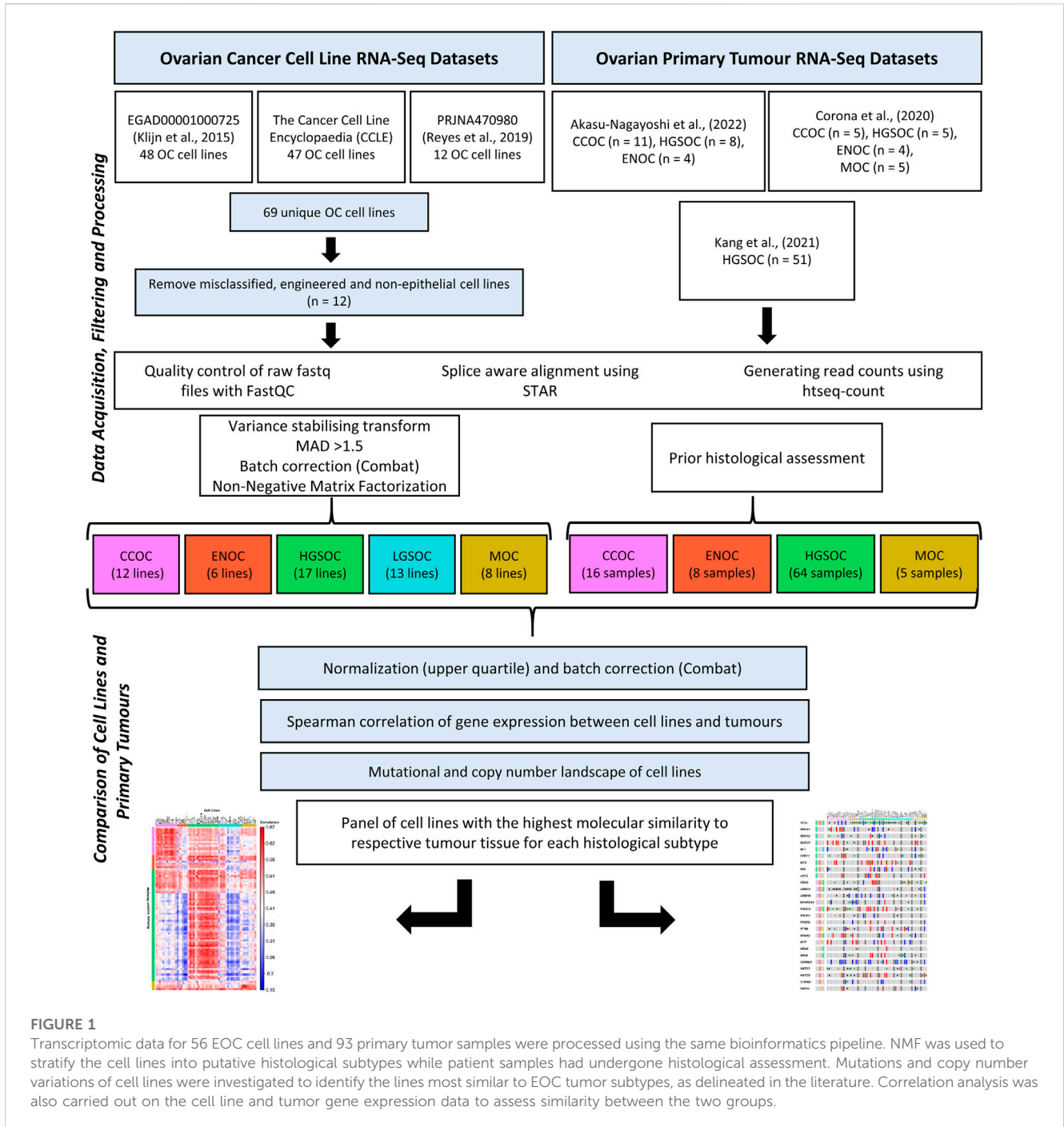
Materials and methods

An overview of the methodology followed is displayed in Figure 1.

RNA-seq data acquisition and filtering

Ovarian cancer cell line RNA sequencing (RNA-seq) data was obtained from three sources: Klijn et al. (2015), the Cancer Cell Line Encyclopedia (CCLE, Ghandi et al., 2019) and Reyes et al. (2019). Raw sequence files from Klijn et al. (2015) were obtained with permission from the Genentech Data Access Committee (dataset ID EGAD00001000725) and downloaded from the EMBL-EGA servers using the pyEGA3 download client. For the CCLE and Reyes datasets, raw sequence files in FASTQ format were obtained from the European Nucleotide Archive (accessions PRJNA523380 and PRJNA470980). Klijn et al. (2015), CCLE and Reyes et al. (2019) provide RNA-seq data for 48, 47 and 12 OC cell lines respectively. In total, RNA-seq reads were available for 69 unique OC cell lines. An overview of datasets used and associated metadata are detailed in Supplementary Table S1.

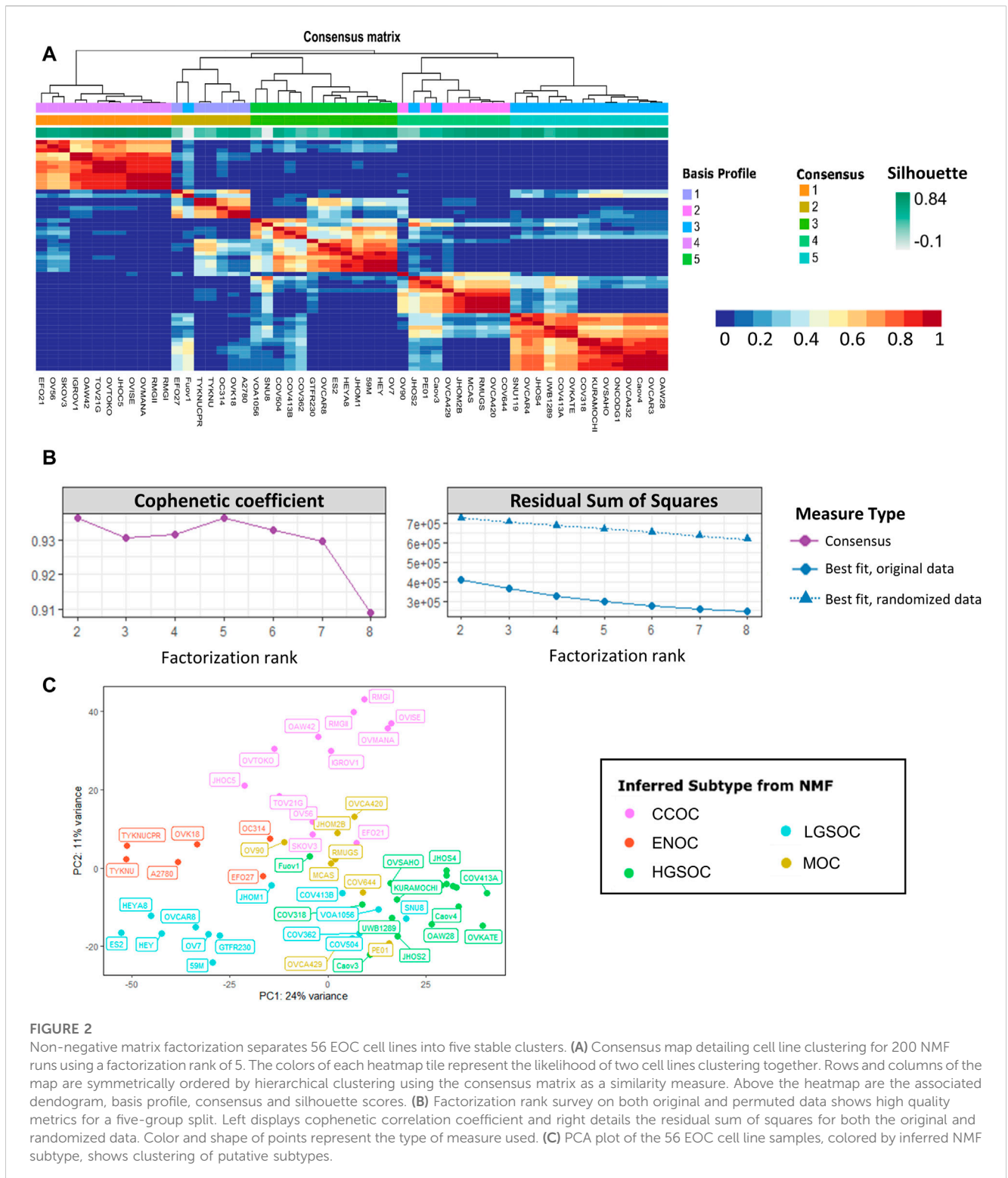
OC cell lines that were misclassified, engineered or non-epithelial in origin were removed from analysis ($n = 11$, Supplementary Table S2). Two cell lines (DOV13 and OVCAR433) could not be processed due to errors with the EGA download client pyEGA3. 56 EOC cell lines remained for further



analysis. A literature search on site of origin, original subtype annotation, *TP53* mutational status and treatment for each cell line and is detailed in [Supplementary Table S3](#).

Ovarian primary tumor tissue data was obtained from three independent datasets (Corona et al., 2020; Kang et al., 2021; Akasu-Nagayoshi et al., 2022). Corona et al. (2020) provides both RNA-seq and chromatin immunoprecipitation sequencing (ChIP-seq) data on EOC tumors of various subtypes: CCOC (n = 5), HGSOC (n = 5), ENOC (n = 4) and MOC (n = 5). RNA-seq data for these 19 primary

tumors were analyzed in this study. Kang et al. (2021) generated transcriptomic profiles for 51 HGSOC tumors of various stages and sensitivities to chemotherapy. Finally Akasu-Nagayoshi et al. (2022) conducted RNA-seq on stage III and IV primary tumor tissue of various subtypes: CCOC (n = 11), HGSOC (n = 8) and ENOC (n = 4). There is no publicly available RNA-seq data for LGSOC primary tumors (although a number of array-based sequencing datasets are available). In total, transcriptomic data for 16 CCOCs, 64 HGSOCs, 8 ENOCs and 5 MOCs were used.



RNA-seq data processing

Quality control checks were performed on all FASTQ files using the FastQC package (Andrews, 2010). Forward and reverse reads for each cell line were mapped using the Spliced Transcripts Alignment to a Reference (STAR) software, version 2.7.9a (Dobin et al., 2013). The reads generated by Akasu-Nagayoshi et al. (2022), were

single-end, which was taken into account during the alignment step. Reads were mapped to the human reference genome sequence (GRCh38) from Gencode (Frankish et al., 2021). Genome indices were generated using the comprehensive Gencode v39 gene annotation. Separate indices were required for each read length (Supplementary Table S1). Transcript level counts were generated using htseq-count (Anders et al., 2015). Strandedness of each dataset

was taken into account during this step. Counts were compiled into a single matrix using R studio (v4.1.1). ComBat (Leek et al., 2012) was used to correct for batch effects between the three independent sources of data.

Non-negative matrix factorization

NMF was carried out on the processed count matrix, filtered to retain only cell lines, in R studio (v4.1.1). A variance stabilizing transform was applied to read counts using the DESeq2 package (Love et al., 2014). In order to retain the most variable genes, transcripts were filtered to exclude those with a median absolute deviation of less than 1.5. There were 2233 transcripts with median absolute deviation of ≥ 1.5 ; these were used as input for the NMF R package (Gaujoux and Seoighe, 2010). Estimation of factorization rank (r) was carried out by running NMF for values of r from 2 to 8, using 50 random initiation points. Quality measures including the cophenetic correlation coefficient, residual sum of squares and dispersion were used to select the most suitable value of r . In the presented data, clustering for r values of 2 and 5 displayed high quality metrics (Figure 2). After selection of a suitable r value, 200 runs of NMF were carried out with a factorization rank of 5 and a fixed random seed. Samples were assigned to one of five clusters based on maximum coefficient matrix values. The most likely subtype of each NMF cluster was inferred by comparison with previous studies and literature annotations.

Mutational landscape of EOC cell lines and tumors

For cell lines present in the CCLE study, mutational and copy number landscape was originally determined by Ghandi et al. (2018) and visualized using cBioportal (Cerami et al., 2012; Gao et al., 2013). Mutational profiles of cell lines not present in the CCLE were determined from multiple sources (Anglesio et al., 2013; Beaufort et al., 2014; Tate et al., 2019). Copy number data was not available for 59M, UWB1289, OVCA420, OVCA432, PE01, HEY, RMGII, TYKNUCPR, COV413A, COV413B, GTFR230, VOA1056 or OVCA429. Mutational data was not available for a subset of genes in the RMGII, TYKNUCPR, COV413A, OVCA431, COV413B, GTFR230, VOA1056, and OVCA429 cell lines. A literature search was also performed to investigate the mutation frequency of 26 cancer driver/tumor suppressor genes in each of the five subtypes of EOC (Supplementary Table S4)

Correlation analysis

The uniformly processed EOC cell line and tumor tissue counts were upper quartile normalized. ComBat (Leek et al., 2012) was then used to correct for batch effects between different sources of data. The normalized and batch corrected counts were filtered to keep the 2233 most variable transcripts in EOC cell lines (as identified earlier). The Spearman correlation was then calculated between cell lines and tumor tissues. Results were plotted as a correlation matrix using the corrplot package (Wei et al., 2021) and as boxplots

using the ggplot2 package (Wickham, 2016), ordered by median correlation.

Results

NMF separates EOC cell lines into five clusters that reflect histological subtyping

Non-negative matrix factorization is a method of unsupervised learning often applied to gene expression data to extract biologically relevant information (Brunet et al., 2004; Gaujoux and Seoighe, 2010; Barnes et al., 2021). It functions by transforming a large, non-negative matrix (such as read counts) into a lower dimensional space: two smaller matrices W and H . The basis matrix W delineates the contribution of a small subset of genes to what are termed 'metagenes'. This is essentially a decomposition of genes into those whose co-expression influences cluster assignment (Brunet et al., 2004). The coefficient matrix H details the co-expression pattern of metagenes in each sample, and can be used to cluster samples into a defined number of groups (r). Seeding is used to initialize the starting point for the NMF algorithm (i.e., to provide starting values for the basis and consensus matrices). The NMF package contains a number of built in seeding methods. Here the 'random' method is used, whereby these initial values are obtained from a uniform distribution with the same range as the input matrix. For reproducibility, a numerical value is passed into the seed function to seed the random number generator. In order to achieve a stable result from this random seeding method, multiple runs of NMF are required.

The number of metagenes and sample clusters are defined by the factorization rank (r), a critical parameter that is selected by the user-ideally, r should be small enough to reduce noise but large enough to preserve biologically relevant information. Brunet et al. (2004) suggests that the factorization rank should be chosen as the smallest value of r for which the cophenetic correlation coefficient starts decreasing. The cophenetic correlation coefficient indicates the robustness of clusters for a given choice of r . Frigyesi and Höglund (2008) suggest investigating methods other than the cophenetic correlation coefficient. It is argued that an increase in r is only relevant if the information captured by factorization is greater than that derived from permuted data, otherwise this increase will lead to overfitting. They suggest using the smallest value of r where the decrease in residual sum of squares (RSS) is lower than the decrease observed in randomized data. By comparing the residual error of NMF from the original data to that of permuted data, a solution is identified that is more information than noise, while preventing overfitting.

The consensus matrix also provides insight into the stability of the clusters. The entries of the consensus matrix reflect the probability two samples belong to the same cluster. The dispersion of the consensus matrix measures the reproducibility of clusters obtained from NMF; 1 for a perfect consensus matrix, and between 0 and 1 for a scattered consensus matrix. A consensus map can be plotted, which is a consensus matrix computed over multiple independent NMF runs, which is the average of the connectivity matrices of each separate run (Figure 2A). Quality of clustering can also be assessed using silhouette scores. This takes into account how



similar a particular sample is to the other samples in its assigned cluster (mean intra-cluster distance), as well as the similarity to samples in other clusters (mean neighboring-cluster distance). A silhouette score close to 0 indicates the sample is on or near the decision boundary of two clusters whereas a silhouette score of 1 shows that the sample is distinct from clusters to which it does not belong. A negative silhouette score indicates a given sample has likely been erroneously assigned to a given cluster.

r was estimated by performing NMF with 50 runs of each value of r from 2 to 8. This was also completed for variance stabilized counts that had been permuted using the randomize function from the NMF package (Gaujoux and Seighe, 2010). In the presented data, clustering for r values of 2 and 5 displayed high quality metrics (Figures 2A, B). Cophenetic correlation coefficients for $r = 2$ and 5 show robustness, and a drop in cophenetic correlation coefficient for $r = 3$ and $r = 6$ indicates less stability. For $r = 6$ the decrease in RSS in the observed data is less than the decrease in RSS in the randomized data. Therefore, a factorization rank of 5 was chosen for NMF analysis. The most likely subtype of each of the 5 NMF

clusters was inferred by comparison with previous studies and literature annotations. In order to investigate the intra-cluster similarity, principal component analysis was carried out on the variance stabilized counts (Figure 2C). Each cluster assigned by NMF largely grouped together on the PCA plot. The ENOC and CCOC lines clustered together from the other samples, whereas there was some overlap between certain MOC, LGSOC and HGSOC cell lines. This suggested that some cell lines may have been misannotated by NMF or may display transcriptional characteristics of multiple EOC subtypes.

NMF validates previous subtype classifications and assigns subtypes to five previously unannotated cell lines

The results obtained in using NMF analysis predominantly validated subtypes assigned in other studies (Supplementary Table S3). This study analyzed 43, 42, 22, and 20 EOC cell lines

in common with Barnes et al. (2021), Domcke et al. (2013), Beaufort et al. (2014) and Anglesio et al. (2013), with subtype assignment agreeing in 39, 36, 15, and 14 cases respectively. Disagreements between our classifications and those of other studies added clarity to certain subtype assignments. We also utilized NMF to assign subtypes to five previously unannotated cell lines.

There were a number of differences between the subtypes assigned to certain cell lines in this study compared to other previous works, often adding clarity to stratifications. OAW42, while classified as serous by Beaufort et al. (2014), was found to cluster with CCOC lines both here and by Barnes et al., (2021). Domcke et al. (2013) also classed this line as unlikely to be of HGSOc origin. The fact that OAW42 is *TP53* wild type, and possesses characteristic CCOC mutations (*ARID1A* and *PIK3CA*, Figure 3), suggests that this line is likely a model of CCOC. OVCAR8 was classified as LGSOC both here and in Barnes et al. (2021), unlikely to be HGSOc by Domcke et al. (2013) and as HGSOc by Anglesio et al. (2013). Although this line has a *TP53* mutation and *MYC* amplification (both of which are common to HGSOc), it also possesses *KRAS* and *ERRB2* mutations (Figure 3), suggesting it is more likely a model of LGSOC, in agreement with Barnes et al. (2021) and Domcke et al. (2013). HEY was classified here as LGSOC, which directly contradicted with Anglesio et al. (2013), where this line was determined to be of HGSOc origin. HEYA8 is another cell line analyzed, which was derived from the peritoneal cavity of mice injected with HEY cells (Supplementary Table S3). HEYA8 was categorized as a LGSOC line both here and by Barnes et al. (2021). It is likely that HEY is of LGSOC origin, as it possesses characteristic LGSOC mutations (*KRAS* and *BRAF*, Figure 3), and it is the clonal ancestor of the HEYA8 cell line. Finally, OV56 was classified as LGSOC in Barnes et al. (2021), unlikely HGSOc in Domcke et al. (2013) and ENOC/CC in Beaufort et al. (2014). It is important to note that although classified as LGSOC by Barnes et al. (2021), the authors suggest that OV56 more likely represents CCOC, due to the presence of *KRAS*, *ARID1A*, *PIK3RI* and *PTEN* mutations (Figure 3). Indeed, we found this to be true in our analysis, with OV56 clustering with the CCOC cell lines.

Five cell lines (GTFR230, OVCA420, OVCA429, OVCA432, and TYKNUCPR) have not been investigated in terms of EOC subtype since their creation, with original annotation data unavailable for four of these lines (OVCA420, OVCA429, OVCA432, and TYKNUCPR). Additionally, these cell lines are rarely cited in literature and as such, mutational and copy number data was largely unavailable (Figure 3). TYKNUCPR is from a population of TYKNU cells exposed to cisplatin chemotherapy, so it is likely that these cell lines share a common subtype. Indeed, both lines were shown to cluster together, with putative ENOC cell lines. GTFR230 was originally derived from a stage IC MOC, although we show here that it clustered with LGSOC lines with a relatively high silhouette score (Figure 2A). As no mutational or copy number data was available for this line, an investigation into the similarity of GTFR230 to LGSOC and MOC tumors could not be carried out. This highlights the importance of other methods to determine subtype molecular similarity. OVCA432 was assigned as HGSOc in this study, while OVCA420 and OVCA429 both clustered with MOC lines. Mutational data was available for OVCA420, which showed that

this line may possess characteristics of both HGSOc (*TP53*, *BRCA1*, *CDK12*, *RBI* mutations) and ENOC/CCOC/MOC (*ARID1A*) (Figure 3). Ultimately, we assign potential subtypes to these five cell lines based on NMF cluster membership. However, we note that due to the lack of additional mutational and copy number data, further investigation is needed to confirm classification of not only these lines, but all lines analyzed in this study.

Integration of cell line molecular data highlights the most and least suitable models for subtype specific research

As discussed previously, subtype assignment of certain cell lines can prove difficult to resolve using transcriptional data, especially when mutational or copy number data is unavailable. Comparison of cell lines to tumor tissue is required to determine the suitability of these *in vitro* models. Some progress has been made in identifying suitable cell lines to utilize as HGSOc models. Both Domcke et al. (2013) and Yu et al. (2019) compared various molecular characteristics of cell lines to primary tumors of HGSOc origin, providing researchers with recommendations of the most suitable HGSOc lines. As of yet however, no study has extended analysis of this nature to multiple subtypes, to compile a reference set of cell lines representative of other EOC subtypes.

Publicly available transcriptomic data was available for primary tumors of HGSOc, CCOC, ENOC and MOC origins (93 samples in total, Supplementary Table S1). No RNA-seq data was available for LGSOC primary tumors, therefore this subtype could not be included in correlation analysis. Following normalization and batch corrections, we calculated correlation of gene expression profiles between EOC cell lines and primary EOC tumors (Figures 4, 5). Median correlations ranged from 0.28 to 0.59, with a similar range observed in Yu et al. (2019). In general, we observed that cell lines we assigned as HGSOc, CCOC and MOC were highly correlated to their respective primary tumor subtypes, whereas ENOC lines were poorly correlated to EOC overall (Figures 4, 5). We also note that most cell lines classified as LGSOC correlated poorly with tumors of HGSOc, CCOC, MOC and ENOC origin. Our data suggests that these lines may be useful models of LGSOC, although further analysis is needed to compare similarity to actual LGSOC tumor samples.

All HGSOc cell lines (apart from Fuov1) were within the top 20 most correlated cell lines to HGSOc primary tumors, with Caov4, COV413A, OVCA432, OVCAR3 and OVKATE representing the most highly correlated cell lines (Figure 5A). This study represents the first instance of subtype assignment of OVCA432, demonstrating its high similarity to HGSOc tumors. To validate that our ranking of HGSOc lines was consistent with both Domcke et al. (2013) and Yu et al. (2019), we first calculated the mean correlation of all cell lines with the 56 primary HGSOc tumors. We filtered this to include only the 39 cell lines analyzed by both Domcke et al. (2013) and Yu et al. (2019), and ranked cell lines in order of mean correlation with HGSOc tumors. Our suitability ranking was highly correlated with that observed by both Yu et al. (2019) (Spearman's $\rho = 0.89$, p -value $< 2.2e-16$) and Domcke et al. (2013) order (Spearman's $\rho = 0.58$, p -value $= 9.592e-05$). A lower ρ value was observed between our ranking,

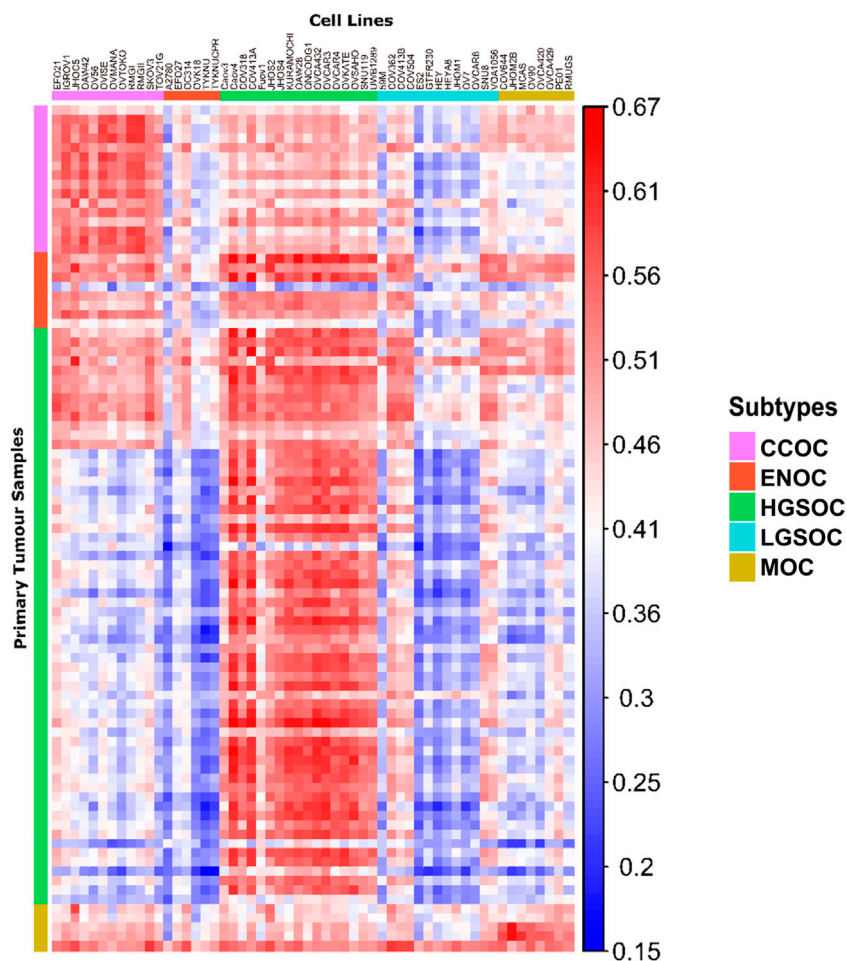


FIGURE 4

Correlation matrix comparing similarity of cell line models to primary tumor tissue of each subtype. Each column represents a cell line, with the putative subtype assigned by NMF delineated by colored boxes (top). Each row represents a tumor sample, with the subtype of samples as identified by histological assessment shown by colored boxes (left). Each tile shows the correlation of gene expression between a cell line and tumor sample, ranging from 0.15 to 0.67 (blue to red).

produced using gene expression data only, compared to that produced by Domcke et al. (2013), which took into account copy number alteration and mutational landscape. A higher rho value was observed between our ranking and that detailed in Yu et al. (2019), which was expected, as both studies also analyze transcriptomic data. However, in this study we analyze a completely independent set of 56 HGSOE tumors yet produce a similar suitability ranking.

Similar to HGSOE, cell lines classified here as CCOC possessed the highest ranking median correlations to CCOC tumors (Figure 5B). The 12 CCOC lines fell within the top 13 most correlated lines to CCOC tumors, with OVCAR4 (a HGSOE cell line) ranking 11th. The top 5 cell lines with highest mean correlation to CCOC tumors are RMGII, RMGI, OVISE, IGROV1 and OAW42. Although no mutational data is available for RMGII, its high correlation to CCOC tumors warrants its use as a CCOC model; however, mutational and copy number analysis on this line would be extremely useful to confirm this similarity. TOV21G and JHOC5 are also highly correlated to CCOC tumors, supporting recommendations for their use as CCOC models from Anglesio

et al. (2013). It is important to note that Domcke et al. (2013) observed hypermutated genomes in TOV21G and IGROV1. Therefore, although these cell lines are highly correlated with CCOC tumors, caution should be taken when using these in experiments due to their hypermutated genomes. High correlation of OVCAR4 with CCOC tumors could be explained by the fact that this line harbors mutations that are seen in both HGSOE and CCOC (*BRCA2*, *MYC* amplifications), as well as those that are generally not observed in HGSOE (*MET* and *PTEN* amplifications, mutations in *CDKN2A*) (Figure 3). This high mutational overlap with CCOC may influence the high position of OVCAR4 in this ranking.

None of the six lines classified as ENOC ranked within the top 20 most correlated lines to ENOC tumors (Figure 5C). In fact, these cell lines were poorly correlated to tumors of HGSOE, CCOC and MOC. The most striking observation relates to A2780, a cell line reported to represent over 90% of citations in EOC studies (Domcke et al., 2013). A2780 has been previously reported to poorly model HGSOE tumors, and our results show that is a poor model of

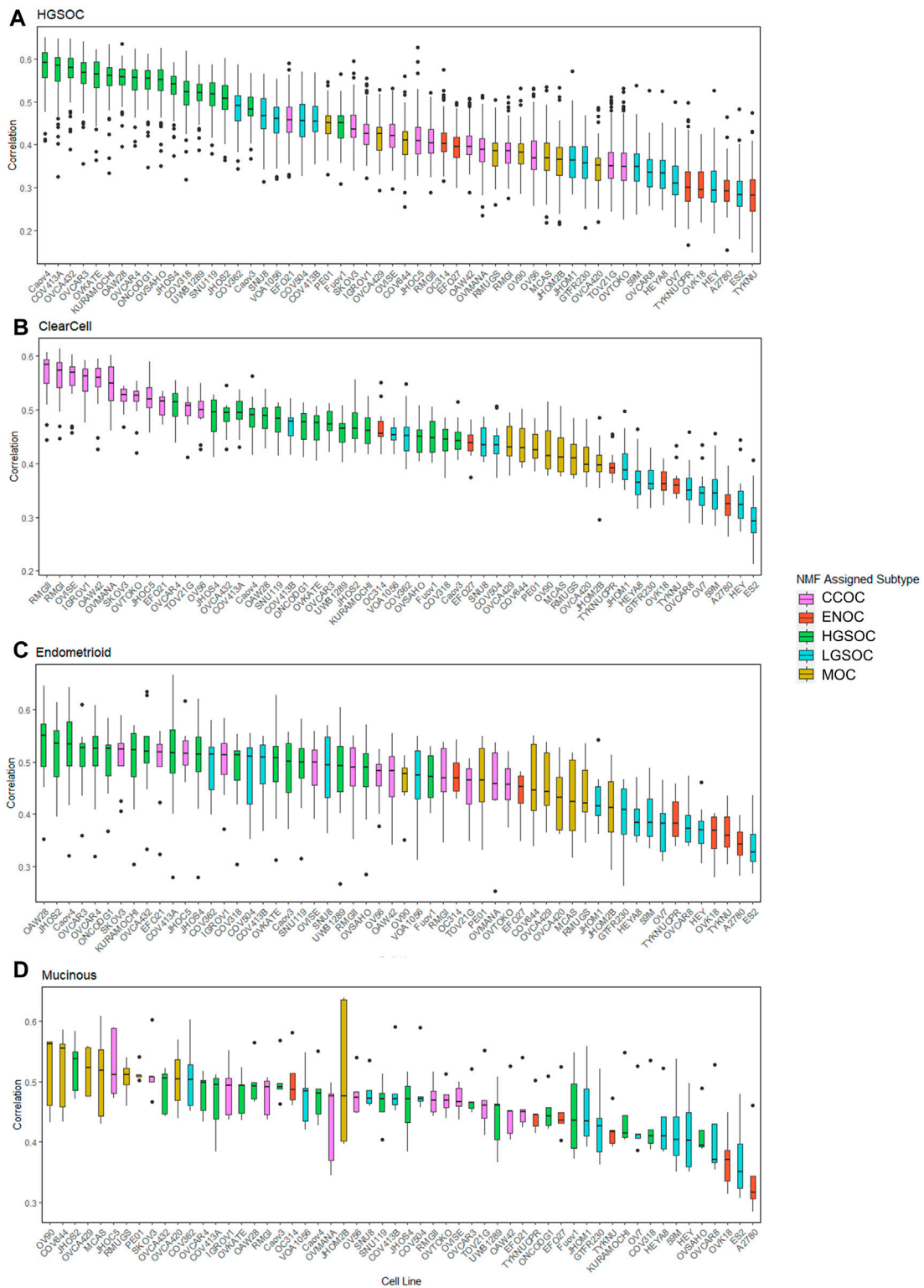


FIGURE 5

Distribution of correlations between cell lines and primary tumors of varying subtypes. Each boxplot is ordered in decreasing rank of median correlation between a particular cell line and all tumors of each subtype, (A) HGSOc tumors, (B) CCOC tumors, (C) ENOC tumors and (D) MOC tumors. Each box is colored based on subtype assigned to each cell line by NMF.

Cell Line	Subtype Assigned	Barnes	Domcke	Beaufort	Anglesio	Recommendations for Use
A2780	ENOC	ENOC	Not within the top 20 most likely HGSOC	ENOC/CCOC	ENOC Carcinoma	Not recommended for use
Caov3	HGSOC	HGSOC	Within the top 20 most likely HGSOC	Serous	HGSOC	HGSOC Model
OVCA83	HGSOC	HGSOC	Within the top 20 most likely HGSOC	HGSOC	HGSOC	HGSOC Model
SKOV3	CCOC	CCOC	Not within the top 20 most likely HGSOC	ENOC/CCOC	Non-serous	CCOC Model
TOV21G	CCOC	CCOC	Hypermutated	CCOC	CCOC	CCOC Model*
Caov4	HGSOC	HGSOC	Within the top 20 most likely HGSOC	HGSOC	NA	HGSOC Model
COV318	HGSOC	HGSOC	Within the top 20 most likely HGSOC	HGSOC	NA	HGSOC Model
COV644	MOC	MOC	Not within the top 20 most likely HGSOC	MOC	NA	MOC Model
IGROV1	CCOC	CCOC	Hypermutated	Mixed	ENOC Carcinoma	CCOC Model*
OAW28	HGSOC	HGSOC	Within the top 20 most likely HGSOC	HGSOC	NA	HGSOC Model
JHOC5	CCOC	CCOC	Not within the top 20 most likely HGSOC	NA	CCOC	CCOC Model
KURAMOCHI	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	HGSOC	HGSOC Model
MCAS	MOC	MOC	Not within the top 20 most likely HGSOC	NA	MOC	MOC Model
OVCA4	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	HGSOC	HGSOC Model
OVISE	CCOC	CCOC	Not within the top 20 most likely HGSOC	NA	Non-serous	CCOC Model
OVMANA	CCOC	CCOC	Not within the top 20 most likely HGSOC	NA	CCOC	CCOC Model
OVTOKO	CCOC	CCOC	Not within the top 20 most likely HGSOC	NA	CCOC	CCOC Model
EFO21	CCOC	CCOC	Not within the top 20 most likely HGSOC	NA	NA	CCOC Model
Fuov1	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	Not recommended for use
HEYA8	LGSOC	LGSOC	Not within the top 20 most likely HGSOC	NA	NA	Potential LGSOC Model
JHOM1	LGSOC	LGSOC	Not within the top 20 most likely HGSOC	NA	NA	Potential LGSOC Model
JHOS2	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	HGSOC Model
JHOS4	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	HGSOC Model
OAW42	CCOC	CCOC	Not within the top 20 most likely HGSOC	Serous	NA	CCOC Model
OVCA88	LGSOC	LGSOC	Not within the top 20 most likely HGSOC	NA	HGSOC	Potential LGSOC Model
OVK18	ENOC	ENOC	Hypermutated	NA	NA	Not recommended for use
OVKATE	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	HGSOC Model
OVSAHO	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	HGSOC Model
RMGI	CCOC	CCOC	Not within the top 20 most likely HGSOC	NA	No validated cell line histotype	CCOC Model
RMUGS	MOC	MOC	Not within the top 20 most likely HGSOC	NA	NA	MOC Model
ONCODG1	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	HGSOC Model
SNU119	HGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	HGSOC Model
EFO27	ENOC	ENOC	Hypermutated	NA	NA	Not recommended for use
OV90	MOC	MOC	Not within the top 20 most likely HGSOC	HGSOC	No validated cell line histotype	MOC
59M	LGSOC	LGSOC	Within the top 20 most likely HGSOC	Mixed	NA	Potential LGSOC Model
OV7	LGSOC	LGSOC	NA	Mixed	NA	Potential LGSOC Model
ES2	LGSOC	LGSOC	Within the top 20 most likely HGSOC	CCOC	Non-serous	Potential LGSOC Model
JHOM2B	MOC	MOC	Within the top 20 most likely HGSOC	NA	NA	Not recommended for use
OC314	ENOC	ENOC	NA	NA	NA	Not recommended for use
COV504	LGSOC	NA	Not within the top 20 most likely HGSOC	Serous	NA	Not recommended for use
OV56	CCOC	LGSOC	Not within the top 20 most likely HGSOC	ENOC/CCOC	NA	CCOC Model
COV362	LGSOC	HGSOC	Within the top 20 most likely HGSOC	ENOC/HGSOC	NA	Not recommended for use
UWB1289	HGSOC	NA	NA	Serous	NA	HGSOC Model
RMGII	CCOC	NA	NA	NA	CCOC	CCOC Model
VOA1056	LGSOC	NA	NA	NA	LGSOC	Not recommended for use
SNU8	LGSOC	HGSOC	Within the top 20 most likely HGSOC	NA	NA	Not recommended for use
COV413A	HGSOC	NA	NA	ENOC/CCOC	NA	HGSOC Model
COV413B	LGSOC	NA	NA	ENOC/CCOC	NA	Not recommended for use
GTFR230	LGSOC	NA	NA	NA	NA	Not recommended for use
HEY	LGSOC	NA	NA	NA	HGSOC	Potential LGSOC Model
OVCA420	MOC	NA	NA	NA	NA	MOC Model
OVCA429	MOC	NA	NA	NA	NA	MOC Model
OVCA432	HGSOC	NA	NA	NA	NA	HGSOC Model
PE01	MOC	NA	NA	HGSOC	NA	MOC Model
TYKNU	ENOC	LGSOC	Within the top 20 most likely HGSOC	NA	NA	Not recommended for use
TYKNUCPR	ENOC	NA	NA	NA	NA	Not recommended for use

FIGURE 6

Consensus table displaying cell line subtype classification in this study, Barnes et al. (2021), Domcke et al. (2013), Beaufort et al. (2014) and Anglesio et al. (2013), and recommendations for future use. Overall, there are 16 models of HGSOC, 12 models of CCOC and seven models of MOC. Seven lines represent potential models of LGSOC, although further comparisons to tumor samples are needed to confirm this. Fourteen lines should be avoided in future *in vitro* and *in silico* studies due to i) poor correlation to EOC tumors overall (A2780, OVK18, OC314, EFO27, TYKNU, TYKNUCPR), ii) classification as a certain subtype but displaying poor correlation to respective primary tumors (Fuov1, JHOM2B), iii) high correlations to multiple EOC subtypes (SNU8), iv) classification as LGSOC but displaying high correlations to other EOC subtypes (COV504, VOA1056, COV413B, COV362) and GTFR230 due to origination from a MOC tumor but classification as LGSOC. *Caution should be taken when utilizing TOV21G and IGROV1 as models of CCOC due to their hyper mutated genomes.

ENOC, the EOC tumor type it is expected to represent. Expression based clustering by Domcke et al. (2013) also showed that A2780 clusters closer to cancers of non-ovarian origin (such as lung, liver, stomach and small intestine) than to those of ovarian origin. Our analysis bolsters this observation, demonstrating that not only does A2780 display poor correlation with HGSOC tumors, but of EOC tumors of multiple subtypes. Based on our evidence, we recommend that use of A2780 for *in vitro* EOC studies should be avoided, and importantly, there is a striking need to develop additional ENOC cell lines. We observe a similar pattern for other lines classified as ENOC. Cell lines classified as HSGOC and CCOC may represent the best

available models of this subtype, as they possess the highest median correlations to ENOC primary tumors (Figure 5C). Indeed HGSOC tumors have been noted to be highly similar to ENOC tumors of higher grades (Gilks et al., 2008; Karnezis et al., 2013; Lim et al., 2016) and CCOC tumors display a high mutational overlap with characteristic ENOC genomic alterations (Supplementary Table S4).

All lines classified as MOC (apart from JHOM2B) fell within the top 20 most highly correlated lines to MOC primary tumors (Figure 5D). COV644, MCAS, OVCA420, OVCA429, and RMUGS were amongst the lines with highest similarity. Again, this is the first reported instance of subtype assignment for

OVCA420 and OVCA429, showing that these lines possess high molecular similarity to MOC tumors, despite a lack of mutational and copy number data. We also note here that GTF2R230, while classified in this study as LGSOC, originated from a stage IC MOC tumor (Supplementary Table S2). We observed that this line ranks 48th (out of 56 lines) in terms of median correlation to MOC tumors, much lower than any putative MOC cell line (Figure 5D). We therefore suggest that this cell line is more representative of LGSOC, due to clustering with LGSOC lines. Further studies are needed to confirm this.

Most cell lines classified as LGSOC (HEY, HEYA8, JHOM1, OVCAR8, 59M, OV7, and ES2) were poorly correlated to primary tumors of HGSOE, CCOC, ENOC, and MOC origin (Figures 3–5). Additionally, these lines possess the major genomic alterations of LGSOC (Figure 3), and have been suggested to represent the LGSOC subtype by Barnes et al. (2021). We therefore hypothesize that these lines may potentially represent models of LGSOC. However, a comparison of these lines to primary LGSOC tumors is needed to confirm this, highlighting the need for additional datasets to be generated for this rarer EOC subtype.

Overall, we show that correlation of gene expression patterns between EOC cell lines and primary tumors can identify models with high molecular similarity to specific subtypes and aid in subtype identification, with the use of techniques such as NMF. In summary, we generated a reference dataset of cell lines most representative of HGSOE, CCOC, and MOC (Figure 6). We also highlight potential LGSOC models and those that are unsuitable for use in subtype specific studies.

Discussion

In this study, we utilized publicly available transcriptomic, mutational and copy number data to investigate the molecular characteristics of 56 EOC cell lines. We demonstrated that cell lines optimally clustered into five stable groups that likely represent the five histological subtypes of EOC. We also compared the gene expression profiles of these cell lines to primary tumors of four EOC subtypes, generally observing that HGSOE, CCOC and MOC cell lines were highly correlated to their respective primary tumors. We noted that ENOC cell lines, especially A2780, were poorly correlated overall to not only ENOC tumors but EOC overall. We identified cell lines that feature the major genomic and transcriptomic features of each particular subtype and are therefore the most suitable *in vitro* models, at least for gene expression-based studies. Potential cell line models of LGSOC were also reported. Finally, we highlight the need for generation of additional subtype-specific datasets to help identify suitable models of less studied EOC subtypes, including LGSOC.

While this study is the first to correlate gene expression between EOC cells and primary tumor tissues stratified by multiple subtypes, it is not without its limitations. This study is purely an *in silico* investigation that lacks experimental validation. However, we note that our results are in agreement with other studies of this nature (Domcke et al., 2013; Yu et al., 2019) and are similar to other experimentally based studies of cell line subtype identification (Anglesio et al., 2013; Beaufort et al., 2014). Another potential limitation of this study is that we did not adjust for tumor purity in our correlation analysis, which has been identified as a confounder in previous studies (Yu et al. 2019). Finally, as our analysis

was conducted using available genomic profiles, we did not include any LGSOC tumors. We also noted an overrepresentation of HGSOE tumor datasets compared to other subtypes. Since HGSOE is the most common and aggressive form of EOC (Prat et al., 2018), there is a disproportionate number of datasets representing this subtype, and comparably fewer examples of similar datasets for rarer, more indolent subtypes.

This evaluation on suitability of cell lines is by no means exhaustive, and is solely tailored to research questions involving gene expression. As our analysis was directed at correlation of transcriptomic profiles, we anticipate our results are most applicable in designing studies aimed at identifying biomarkers with elevated or reduced expression levels in certain subtypes of EOC. In general, determining the optimal tumor model depends on a myriad of considerations and circumstances. For example, future work should focus on identifying the most suitable set of cell lines for methylation or proteomic profiling studies, based on similarity with corresponding tumor genomic data. Additionally, the mutational status of genes other than BRCA1/2 included in clinical lab panels for genetic testing of heritable cancer syndromes could prove useful in identifying histotype-specific models. These include genes commonly mutated in Lynch Syndrome (*MLH1*, *MSH2*, *MSH6*, *PMS2*, and *EPCAM*), genes involved in homologous repair (*RAD51C*, *RAD51D*, *BRIP1*), or genes found to be associated with an increased risk of developing EOC overall (*STK11*, *CHEK2*, *PALB2*, *NBN*, *MRE11A*, and *RAD50*) (Harley et al., 2008; Li et al., 2019; Wagner et al., 2019; Amin et al., 2020).

It is possible that the use of two-dimensional, monocultures of individual cell lines may become greatly reduced or outdated in cancer research, due to the rising popularity of three dimensional *in vitro* models such as spheroids, multicellular organoid systems and tumor-on-a-chip models (Ciucci et al., 2022). Cell lines underrepresent the heterogeneity of tumors and the involvement of the tumor microenvironment, including other cells such as stromal cells (Wangsa et al., 2018, a colorectal cancer example). Cell lines also have higher somatic mutation rates than tumors, acquiring mutations through the culturing process (Kasai et al., 2016). In any case, there will be distinct disadvantages associated with any model system, and the ease of use and affordability of cell lines solidifies them as attractive model systems for lab researchers as starting points for drug screening or experimental optimization. Therefore, it is imperative to maximize the utility of these models, by selecting those that most accurately represent the tumor type to address the research question at hand.

In conclusion, although major technological progress and healthcare improvements have facilitated more effective cancer treatments, the last 20 years have only seen modest improvements in OC survival rates (Allemani et al., 2018; Arnold et al., 2019). Indeed, OC is a poor prognosis cancer, due in part to non-specific symptoms, a primarily imperceptible pre-invasive phase and development of resistance to traditional chemotherapies (Goff et al., 2004). Solutions to these problems have yet to be developed, owing to a lack of consideration for EOC subtypes in pre-clinical studies (Alvarez et al., 2016). In this work, we provide a potential reference for gene expression-based EOC studies to assist researchers in selecting appropriate cell lines to represent EOC subtypes. We concurrently highlight the need for more datasets, representing all subtypes to contribute towards the overall goal of expanding subtype-specific research in the field of EOC. A greater understanding of these diverse subtypes will no doubt lead to more targeted therapies, novel diagnostics and increased survival for patients affected by EOC.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: CCLE (<https://www.ebi.ac.uk/ena/browser/view/PRJNA523380>), Reyes et al. (2019) (<https://www.ebi.ac.uk/ena/browser/view/PRJNA470980>), Akasu-Nagayoshi et al. (2022) (<https://www.ebi.ac.uk/ena/browser/view/PRJNA783540>), Corona et al. (2020) (<https://www.ebi.ac.uk/ena/browser/view/PRJNA495805>), Kang et al. (2021) (<https://www.ebi.ac.uk/ena/browser/view/PRJNA697947>). Data from Klijn et al. (2015) was obtained with permission from the Genentech Data Access Committee (dataset ID EGAD00001000725). R scripts used in the NMF and correlation analyses can be found at https://github.com/aideenmc/FrontiersPaper_EOC_CellLines.

Author contributions

AM, OZ, SM, and KD contributed to the conception and design of the study. AM analysed the data and wrote the original manuscript. AM, SM, and OZ interpreted the data. AM, OZ, SM, and KD edited the manuscript. KD provided funding and supervised the project. SM supervised the project. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Science Foundation Ireland (SFI) through the SFI Centre for Research Training in Genomics Data Science under grant number 18/CRT/6214.

References

- Ahn, G., Folkins, A. K., McKenney, J. K., and Longacre, T. A. (2016). Low-grade serous carcinoma of the ovary: Clinicopathologic analysis of 52 invasive cases and identification of a possible noninvasive intermediate lesion. *Am. J. Surg. Pathology* 40 (9), 1165–1176. doi:10.1097/PAS.0000000000000693
- Akasu-Nagayoshi, Y., Hayashi, T., Kawabata, A., Shimizu, N., Yamada, A., Yokota, N., et al. (2022). PHOSPHATE exporter XPR1/SLC53A1 is required for the tumorigenicity of epithelial ovarian cancer. *Cancer Sci.* 113 (6), 2034–2043. doi:10.1111/cas.15358
- Alexandre, J., Ray-Coquard, I., Selle, F., Floquet, A., Cottu, P., Weber, B., et al. (2010). Mucinous advanced epithelial ovarian carcinoma: Clinical presentation and sensitivity to platinum-paclitaxel-based chemotherapy, the GINECO experience. *Ann. Oncol.* 21 (12), 2377–2381. doi:10.1093/annonc/mdq257
- Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M., et al. (2018). Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): Analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 391 (10125), 1023–1075. doi:10.1016/S0140-6736(17)33326-3
- Alvarez, R. D., Karlan, B. Y., and Strauss, J. F. (2016). Ovarian cancers: Evolving paradigms in research and care[®]. Report from the Institute of medicine. *Gynecol. Oncol.* 141 (3), 413–415. doi:10.1016/j.ygyno.2016.04.541
- Amin, N., Chaabouni, N., and George, A. (2020). Genetic testing for epithelial ovarian cancer. *Best Pract. Res. Clin. Obstetrics Gynaecol.* 65, 125–138. doi:10.1016/j.bpobgyn.2020.01.005
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2), 166–169. doi:10.1093/bioinformatics/btu638
- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Anglesio, M. S., Wiegand, K. C., Melnyk, N., Chow, C., Salamanca, C., Prentice, L. M., et al. (2013). Type-specific cell line models for type-specific ovarian cancer research. *PLoS One* 8 (9), e72162. doi:10.1371/journal.pone.0072162
- Arnold, M., Rutherford, M. J., Bardot, A., Ferlay, J., Andersson, T. M., Myklebust, T. Å., et al. (2019). Progress in cancer survival, mortality, and incidence in seven high-

Acknowledgments

The authors would like to thank Dr Antoinette Perry, Asia Jordan, Adele Connor, and Claire Hughes (University College Dublin) for their helpful discussions and advice.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2023.1104514/full#supplementary-material>

income countries 1995–2014 (ICBP SURVMARK-2): A population-based study. *Lancet Oncol.* 20 (11), 1493–1505. doi:10.1016/S1470-2045(19)30456-5

Babaier, A., and Ghatage, P. (2020). Mucinous cancer of the ovary: Overview and current status. *Diagnostics* 10 (1), 52. doi:10.3390/diagnostics10010052

Barnes, B. M., Nelson, L., Tighe, A., Burghel, G. J., Lin, I., Desai, S., et al. (2021). Distinct transcriptional programs stratify ovarian cancer cell lines into the five major histological subtypes. *Genome Med.* 13 (1), 1–19. doi:10.1186/s13073-021-00952-5

Beaufort, C. M., Helmijr, J. C., Piskorz, A. M., Hoogstraat, M., Ruigrok-Ritstier, K., Besselink, N., et al. (2014). Ovarian cancer cell line panel (OCCP): Clinical importance of *in vitro* morphological subtypes. *PLoS One* 9 (9), e103988. doi:10.1371/journal.pone.0103988

Berns, E. M., and Bowtell, D. D. (2012). The changing view of high-grade serous ovarian cancer. *Cancer Res.* 72 (11), 2701–2704. doi:10.1158/0008-5472.CAN-11-3911

Brunet, J. P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci.* 101 (12), 4164–4169. doi:10.1073/pnas.0308531101

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095

Cheasley, D., Fernandez, M. L., Köbel, M., Kim, H., Dawson, A., Hoenisch, J., et al. (2022). Molecular characterization of low-grade serous ovarian carcinoma identifies genomic aberrations according to hormone receptor expression. *NPJ Precis. Oncol.* 6 (1), 47–48. doi:10.1038/s41698-022-00288-2

Cheasley, D., Wakefield, M. J., Ryland, G. L., Allan, P. E., Alsop, K., Amarasinghe, K. C., et al. (2019). The molecular origin and taxonomy of mucinous ovarian carcinoma. *Nat. Commun.* 10 (1), 3935. doi:10.1038/s41467-019-11862-x

Chen, B., Sirota, M., Fan-Minogue, H., Hadley, D., and Butte, A. J. (2015). Relating hepatocellular carcinoma tumor samples and cell lines using gene expression data in translational research. *BMC Med. Genomics* 8 (2), S5–S10. doi:10.1186/1755-8794-8-S2-55

- Ciucci, A., Buttarelli, M., Fagotti, A., Scambia, G., and Gallo, D. (2022). Preclinical models of epithelial ovarian cancer: Practical considerations and challenges for a meaningful application. *Cell. Mol. Life Sci.* 79 (7), 364. doi:10.1007/s00118-022-04395-y
- Corona, R. I., Seo, J. H., Lin, X., Hazelett, D. J., Reddy, J., Fonseca, M. A., et al. (2020). Non-coding somatic mutations converge on the PAX8 pathway in ovarian cancer. *Nat. Commun.* 11 (1), 2020. doi:10.1038/s41467-020-15951-0
- Coward, J. I., Middleton, K., and Murphy, F. (2015). New perspectives on targeted therapy in ovarian cancer. *Int. J. Women's Health* 7, 189–203. doi:10.2147/IJWH.S52379
- Coward, J., Kulbe, H., Chakravarty, P., Leader, D., Vassileva, V., Leinster, D. A., et al. (2011). Interleukin-6 as a therapeutic target in human ovarian cancer. *Clin. Cancer Res.* 17 (18), 6083–6096. doi:10.1158/1078-0432.CCR-11-0945
- Cybulska, P., Paula, A. D. C., Tseng, J., Leitao, M. M., Jr, Bashashati, A., Huntsman, D. G., et al. (2019). Molecular profiling and molecular classification of endometrioid ovarian carcinomas. *Gynecol. Oncol.* 154 (3), 516–523. doi:10.1016/j.ygyno.2019.07.012
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635
- Domcke, S., Sinha, R., Levine, D. A., Sander, C., and Schultz, N. (2013). Evaluating cell lines as tumor models by comparison of genomic profiles. *Nat. Commun.* 4 (1), 1–10.
- Emmanuel, C., Chiew, Y. E., George, J., Etemadmoghadam, D., Anglesio, M. S., Sharma, R., et al. (2014). Genomic classification of serous ovarian cancer with adjacent borderline differentiates RAS pathway and TP53-mutant tumors and identifies NRAS as an oncogenic driver. *Clin. Cancer Res.* 20 (24), 6618–6630. doi:10.1158/1078-0432.CCR-14-1292
- Erickson, B. K., Conner, M. G., and Landen, C. N., Jr (2013). The role of the fallopian tube in the origin of ovarian cancer. *Am. J. Obstetrics Gynecol.* 209 (5), 409–414. doi:10.1016/j.ajog.2013.04.019
- Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., et al. (2021). Gencode 2021. *Nucleic acids Res.* 49 (D1), D916–D923. doi:10.1093/nar/gkaa1087
- Friedlander, M. L., Russell, K., Millis, S., Gatalica, Z., Bender, R., and Voss, A. (2016). Molecular profiling of clear cell ovarian cancers: Identifying potential treatment targets for clinical trials. *Int. J. Gynecol. Cancer* 26 (4), 648–654. doi:10.1097/IGC.0000000000000677
- Frigyesi, A., and Höglund, M. (2008). Non-negative matrix factorization for the analysis of complex gene expression data: Identification of clinically relevant tumor subtypes. *Cancer Inf.* 6, 275S606–292. doi:10.4137/cin.s606
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Signal.* 6 (269), pii. doi:10.1126/scisignal.2004088
- Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinforma.* 11 (1), 367–369. doi:10.1186/1471-2105-11-367
- Gershenson, D. M., Sun, C. C., Bodurka, D., Coleman, R. L., Lu, K. H., Sood, A. K., et al. (2009). Recurrent low-grade serous ovarian carcinoma is relatively chemoresistant. *Gynecol. Oncol.* 114 (1), 48–52. doi:10.1016/j.ygyno.2009.03.001
- Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature* 569 (7757), 503–508. doi:10.1038/s41586-019-1186-3
- Gilks, C. B., Ionescu, D. N., Kalloger, S. E., Köbel, M., Irving, J., Clarke, B., et al. (2008). Tumor cell type can be reproducibly diagnosed and is of independent prognostic significance in patients with maximally debulked ovarian carcinoma. *Hum. Pathol.* 39 (8), 1239–1251. doi:10.1016/j.humphath.2008.01.003
- Goff, B. A., Mandel, L. S., Melancon, C. H., and Muntz, H. G. (2004). Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *Jama* 291 (22), 2705–2712. doi:10.1001/jama.291.22.2705
- Gorringe, K. L., Cheasley, D., Wakefield, M. J., Ryland, G. L., Allan, P. E., Alsop, K., et al. (2020). Therapeutic options for mucinous ovarian carcinoma. *Gynecol. Oncol.* 156 (3), 552–560. doi:10.1016/j.ygyno.2019.12.015
- Harley, I., Rosen, B., Risch, H. A., Siminovitsh, K., Beiner, M. E., McLaughlin, J., et al. (2008). Ovarian cancer risk is associated with a common variant in the promoter sequence of the mismatch repair gene MLH1. *Gynecol. Oncol.* 109 (3), 384–387. doi:10.1016/j.ygyno.2007.11.046
- Hunter, S. M., Anglesio, M. S., Ryland, G. L., Sharma, R., Chiew, Y. E., Rowley, S. M., et al. (2015). Molecular profiling of low grade serous ovarian tumours identifies novel candidate driver genes. *Oncotarget* 6 (35), 37663–37677. doi:10.18632/oncotarget.5438
- Itamochi, H., Oishi, T., Oumi, N., Takeuchi, S., Yoshihara, K., Mikami, M., et al. (2017). Whole-genome sequencing revealed novel prognostic biomarkers and promising targets for therapy of ovarian clear cell carcinoma. *Br. J. Cancer* 117 (5), 717–724. doi:10.1038/bjc.2017.228
- Jayson, G. C., Kohn, E. C., Kitchener, H. C., and Ledermann, J. A. (2014). Ovarian cancer. *Lancet* 384 (9951), 1376–1388. doi:10.1016/S0140-6736(13)62146-7
- Jiang, G., Zhang, S., Yazdanparast, A., Li, M., Pawar, A. V., Liu, Y., et al. (2016). Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. *BMC Genomics* 17 (7), 525–301. doi:10.1186/s12864-016-2911-z
- Jones, S., Wang, T. L., Shih, I. M., Mao, T. L., Nakayama, K., Roden, R., et al. (2010). Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* 330 (6001), 228–231. doi:10.1126/science.1196333
- Kanchi, K. L., Johnson, K. J., Lu, C., McLellan, M. D., Leiserson, M. D., Wendl, M. C., et al. (2014). Integrated analysis of germline and somatic variants in ovarian cancer. *Nat. Commun.* 5 (1), 3156. doi:10.1038/ncomms4156
- Kang, H., Choi, M. C., Kim, S., Jeong, J. Y., Kwon, A. Y., Kim, T. H., et al. (2021). USP19 and RPL23 as candidate prognostic markers for advanced-stage high-grade serous ovarian carcinoma. *Cancers* 13 (16), 3976. doi:10.3390/cancers13163976
- Karnezis, A. N., Aysal, A., Zaloudek, C. J., and Rabban, J. T. (2013). Transitional cell-like morphology in ovarian endometrioid carcinoma: Morphologic, immunohistochemical, and behavioral features distinguishing it from high-grade serous carcinoma. *Am. J. Surg. Pathology* 37 (1), 24–37. doi:10.1097/PAS.0b013e31826a5399
- Kasai, F., Hirayama, N., Ozawa, M., Iemura, M., and Kohara, A. (2016). Changes of heterogeneous cell populations in the ishikawa cell line during long-term culture: Proposal for an *in vitro* clonal evolution model of tumor cells. *Genomics* 107 (6), 259–266. doi:10.1016/j.ygeno.2016.04.003
- Kim, A., Ueda, Y., Naka, T., and Enomoto, T. (2012). Therapeutic strategies in epithelial ovarian cancer. *J. Exp. Clin. cancer Res.* 31 (1), 14–18. doi:10.1186/1756-9966-31-14
- Kim, S. I., Lee, J. W., Lee, M., Kim, H. S., Chung, H. H., Kim, J. W., et al. (2018). Genomic landscape of ovarian clear cell carcinoma via whole exome sequencing. *Gynecol. Oncol.* 148 (2), 375–382. doi:10.1016/j.ygyno.2017.12.005
- Klijn, C., Durinck, S., Stawiski, E. W., Haverty, P. M., Jiang, Z., Liu, H., et al. (2015). A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* 33 (3), 306–312. doi:10.1038/nbt.3080
- Kurman, R. J., and Shih, I. M. (2010). The origin and pathogenesis of epithelial ovarian cancer—a proposed unifying theory. *Am. J. Surg. Pathology* 34 (3), 433–443. doi:10.1097/PAS.0b013e3181cf3d79
- Lee, Y., Miron, A., Drapkin, R., Nucci, M. R., Medeiros, F., Saleemuddin, A., et al. (2007). A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *J. Pathology A J. Pathological Soc. G. B. Irel.* 211 (1), 26–35. doi:10.1002/path.2091
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034
- Lheureux, S., Braunstein, M., and Oza, A. M. (2019a). Epithelial ovarian cancer: Evolution of management in the era of precision medicine. *CA A Cancer J. Clin.* 69 (4), 280–304. doi:10.3322/caac.21559
- Lheureux, S., Gourley, C., Vergote, I., and Oza, A. M. (2019b). Epithelial ovarian cancer. *Lancet* 393 (10177), 1240–1253. doi:10.1016/S0140-6736(18)32552-2
- Li, W., Shao, D., Li, L., Wu, M., Ma, S., Tan, X., et al. (2019). Germline and somatic mutations of multi-gene panel in Chinese patients with epithelial ovarian cancer: A prospective cohort study. *J. Ovarian Res.* 12 (1), 80–89. doi:10.1186/s13048-019-0560-y
- Lim, D., Murali, R., Murray, M. P., Veras, E., Park, K. J., and Soslow, R. A. (2016). Morphological and immunohistochemical re-evaluation of tumors initially diagnosed as ovarian endometrioid carcinoma with emphasis on high-grade tumors. *Am. J. Surg. pathology* 40 (3), 302–312. doi:10.1097/PAS.0000000000000550
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Mackay, H. J., Brady, M. F., Oza, A. M., Reuss, A., Pujade-Lauraine, E., Swart, A. M., et al. (2010). Prognostic relevance of uncommon ovarian histology in women with stage III/IV epithelial ovarian cancer. *Int. J. Gynecol. Cancer* 20 (6), 945–952. doi:10.1111/IGC.0b013e3181dd0110
- Maru, Y., Tanaka, N., Ohira, M., Itami, M., Hippo, Y., and Nagase, H. (2017). Identification of novel mutations in Japanese ovarian clear cell carcinoma patients using optimized targeted NGS for clinical diagnosis. *Gynecol. Oncol.* 144 (2), 377–383. doi:10.1016/j.ygyno.2016.11.045
- McAlpine, J. N., Porter, H., Köbel, M., Nelson, B. H., Prentice, L. M., Kalloger, S. E., et al. (2012). BRCA1 and BRCA2 mutations correlate with TP53 abnormalities and presence of immune cell infiltrates in ovarian high-grade serous carcinoma. *Mod. Pathol.* 25 (5), 740–750. doi:10.1038/modpathol.2011.211
- Medeiros, F., Muto, M. G., Lee, Y., Elvin, J. A., Callahan, M. J., Feltmate, C., et al. (2006). The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. *Am. J. Surg. Pathology* 30 (2), 230–236. doi:10.1097/01.pas.0000180854.28831.77
- Morice, P., Gouy, S., and Leary, A. (2019). Mucinous ovarian carcinoma. *N. Engl. J. Med.* 380 (13), 1256–1266. doi:10.1056/NEJMra1813254
- Murakami, R., Matsumura, N., Brown, J. B., Higasa, K., Tsutsumi, T., Kamada, M., et al. (2017). Exome sequencing landscape analysis in ovarian clear cell carcinoma shed light on key chromosomal regions and mutation gene networks. *Am. J. Pathology* 187 (10), 2246–2258. doi:10.1016/j.ajpath.2017.06.012

- Peres, L. C., Cushing-Haugen, K. L., Anglesio, M., Wicklund, K., Bentley, R., Berchuck, A., et al. (2018). Histotype classification of ovarian carcinoma: A comparison of approaches. *Gynecol. Oncol.* 151 (1), 53–60. doi:10.1016/j.ygyno.2018.08.016
- Pierson, W. E., Peters, P. N., Chang, M. T., Chen, L. M., Quigley, D. A., Ashworth, A., et al. (2020). An integrated molecular profile of endometrioid ovarian cancer. *Gynecol. Oncol.* 157 (1), 55–61. doi:10.1016/j.ygyno.2020.02.011
- Prat, J., D'Angelo, E., and Espinosa, I. (2018). Ovarian carcinomas: At least five different diseases with distinct histological features and molecular genetics. *Hum. Pathol.* 80, 11–27. doi:10.1016/j.humpath.2018.06.018
- Reyes, A. L. P., Silva, T. C., Coetzee, S. G., Plummer, J. T., Davis, B. D., Chen, S., et al. (2019). GENAVi: A shiny web application for gene expression normalization, analysis and visualization. *BMC Genomics* 20 (1), 745–749. doi:10.1186/s12864-019-6073-7
- Ricciardi, E., Baert, T., Ataseven, B., Heitz, F., Prader, S., Bommert, M., et al. (2018). Low-grade serous ovarian carcinoma. *Geburtshilfe Frauenheilkd.* 78 (10), 972–976. doi:10.1055/a-0717-5411
- Risch, H. A., Bale, A. E., Beck, P. A., and Zheng, W. (2006). PGR+ 331 A/G and increased risk of epithelial ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* 15 (9), 1738–1741. doi:10.1158/1055-9965.EPI-06-0272
- Schmeler, K. M., and Gershenson, D. M. (2008). Low-grade serous ovarian cancer: A unique disease. *Curr. Oncol. Rep.* 10 (6), 519–523. doi:10.1007/s11912-008-0078-8
- Shaw, P. A., McLaughlin, J. R., Zweemer, R. P., Narod, S. A., Risch, H., Verheijen, R. H. M., et al. (2002). Histopathologic features of genetically determined ovarian cancer. *Int. J. Gynecol. Pathology* 21 (4), 407–411. doi:10.1097/00004347-200210000-00011
- Shibuya, Y., Tokunaga, H., Saito, S., Shimokawa, K., Katsuoka, F., Bin, L., et al. (2018). Identification of somatic genetic alterations in ovarian clear cell carcinoma with next generation sequencing. *Genes, Chromosomes Cancer* 57 (2), 51–60. doi:10.1002/gcc.22507
- Shih, I. M., Wang, Y., and Wang, T. L. (2021). The origin of ovarian cancer species and precancerous landscape. *Am. J. Pathology* 191 (1), 26–39. doi:10.1016/j.ajpath.2020.09.006
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Ahmedin, J. (2022). Cancer statistics, 2022. *CA A Cancer J. Clin.* 71 (3), 7–33. doi:10.3322/caac.21708
- Soyama, H., Miyamoto, M., Takano, M., Iwahashi, H., Kato, K., Sakamoto, T., et al. (2018). A pathological study using 2014 WHO criteria reveals poor prognosis of grade 3 ovarian endometrioid carcinomas. *Vivo* 32 (3), 597–602. doi:10.21873/invivo.11281
- Sugiyama, T., Kamura, T., Kigawa, J., Terakawa, N., Kikuchi, Y., Kita, T., et al. (2000). Clinical characteristics of clear cell carcinoma of the ovary: A distinct histologic type with poor prognosis and resistance to platinum-based chemotherapy. *Cancer Interdiscip. Int. J. Am. Cancer Soc.* 88 (11), 2584–2589. doi:10.1002/1097-0142(20000601)88:11<2584:aid-cnrcr22>3.0.co;2-5
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). Cosmic: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47 (D1), D941–D947. doi:10.1093/nar/gky1015
- The Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474 (7353), 609–615. doi:10.1038/nature10166
- Wagner, A. F., Shulman, L. P., and Dungan, J. S. (2019). “Cancer genetics: Risks and mechanisms of cancer in women with hereditary predisposition to epithelial ovarian cancer,” in *Textbook of oncofertility research and practice* (Cham: Springer), 29–43.
- Wangsa, D., Braun, R., Schiefer, M., Gertz, E. M., Bronder, D., Quintanilla, I., et al. (2018). The evolution of single cell-derived colorectal cancer cell lines is dominated by the continued selection of tumor-specific genomic imbalances, despite random chromosomal instability. *Carcinogenesis* 39 (8), 993–1005. doi:10.1093/carcin/bgy068
- Wei, T., Simko, V. R., and Levy, M. (2021). *Package “corrplot”: Visualization of a correlation matrix*. Version 0.84.
- Wickham, H. (2016). “Data analysis,” in *ggplot2* (Cham: Springer), 189–201.
- Wiegand, K. C., Shah, S. P., Al-Agha, O. M., Zhao, Y., Tse, K., Zeng, T., et al. (2010). ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* 363 (16), 1532–1543. doi:10.1056/NEJMoa1008433
- Yu, K., Chen, B., Aran, D., Charalel, J., Yau, C., Wolf, D. M., et al. (2019). Comprehensive transcriptomic analysis of cell lines as models of primary tumors across 22 tumor types. *Nat. Commun.* 10 (1), 3574. doi:10.1038/s41467-019-11415-2