THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

# High Content Profiling in Oesophageal Adenocarcinoma to Inform Drug Discovery and Patient Stratification

Rebecca E Hughes

Doctor of Philosophy

The University of Edinburgh

2021

# *Declaration*

This thesis presents my own work, and has not been submitted in fulfilment of any other degree or professional qualification. Where results have been obtained in collaboration with others, I have clearly stated it in the text. Any information derived from the published work of others has been cited in the text, and a complete list of references can be found in the bibliography. Published papers arising from the work described in this thesis can be found in the appendices.

– Rebecca Hughes, 2021

# *Acknowledgements*

First and foremost, I would like to thank my supervisor, Neil Carragher. It goes without saying that I could not have done this without your guidance and your confidence in my abilities, allowing me to dive head first into a subject I had little knowledge of. You have provided me with invaluable research opportunities, the chance to attend conferences and workshops around the world, and the freedom to pursue my own ideas.

If the last year has taught me anything it's that a great job is nothing without good company and I must therefore thank my colleagues in the Carragher lab: Alison, John, Rich, Kenny, Scott, Karen, Jan and many more; and all the students at the IGMM, without whose laughter, company, and baked goods the last four years would not have been the same. I must also thank each and every one of them individually for helping me in many numerous ways throughout this work.

A significant thank you must go out to Scott, Dom and Jon without whom I would never have attempted to use a computer to do more than check my emails. What a world you helped me discover. I must also thank my friends who have balanced out the last 10 years at University, always providing an adventure in the great outdoors to look forward to and reminding me there is more to life than work.

On a more personal level, I owe everything to my parents for inspiring me to pursue science in the first place and encouraging me to never stop asking questions- to their own detriment on many occasions I am sure. Finally, to Craig, you have kept my life whole, and full of happiness and adventure all these years. Thank you.

# *Abstract*

Oesophageal adenocarcinoma (OAC) is a highly heterogeneous disease, dominated by copy number alterations and large-scale genomic rearrangements. Such characteristics have hampered both therapeutic target discovery and the clinical success of targeted therapies, contributing to its status as an area of unmet therapeutic need. Phenotypic drug discovery describes the screening and selection of hit or lead compounds based on quantifiable phenotypic endpoints from cell-based assays or model organisms without prior knowledge of the drug target. Thus, this may prove to be a beneficial strategy for complex, heterogeneous diseases where target biology is poorly understood and where modern, target directed drug discovery strategies have made little impact on patient care, as exemplified by OAC.

In the chapters that follow, I describe the development and validation of a high content profiling assay and machine learning pipelines to identify repurposing opportunities in OAC. Following on from this I focus on one group of compounds which show high potency and selectivity for targeting OAC cell lines relative to tissue-matched controls. I further explore the mechanism of action of these compounds and potential patient stratification hypotheses to support clinical translation.

In the first chapter I present an in-depth study of compound induced cell morphology in the context of a heterogeneous panel of eight oesophageal cell lines. To achieve this I profiled a small number of reference compounds and built a library of phenotypic fingerprints for each mechanistic class. I then interrogated

the phenotypic fingerprints using a variety of supervised and unsupervised machine learning techniques. This work demonstrates that compound induced morphologies are reproducible across cell lines, despite underlying morphological heterogeneity, allowing machine learning classifiers to be applied to previously unseen cell lines. The second chapter describes the application of the high content assay to a comprehensive small-molecule screen encompassing 19,555 small molecules, eight cell lines, 512x384 well plates, 3.9 million images and 36 TB of data for the identification of new drug discovery and drug repurposing opportunities in OAC. Following on from this, in the third chapter, I focus on one particular group of compounds identified as highly selective for OAC and explore their mechanism of action. In the final chapter, through the integration of transcriptomic pathway analyses we gain insight into drug selectively and the basis for future biomarker-based clinical trials in OAC.

# *Lay Summary*

All cells are made up of a complex mixture of biological molecules that work together to carry out a myriad of cellular processes, from cell growth and division, to cell communication and movement. Disease occurs when normal cellular processes become dysregulated. Cancer is one such disease. I am studying cancer that is found in our food pipe (the oesophagus). Oesophageal cancer is one of numerous diseases for which there is currently no targeted treatment. If you are diagnosed with oesophageal cancer today, you have the same treatment options and chance of survival as you would have done 50 years ago - just 12 %.

Therapeutic drugs act by altering the dysfunctional processes within diseased cells, trying to reverse a dysfunctional cellular process to a healthy one. These changes in cellular processes can be visualised at the microscopic level through a microscope, using fluorescent dyes to light-up the different structures within the cells; blue for DNA, green for the cell skeleton, red for energy producing machinery.

In order to find new treatments for oesophageal cancer I have taken cancer cells and normal cells and systematically treated thousands of them each with different drugs, painted them with the fluorescent dyes and taken photographs. Then, just as image recognition software to unlock your phone takes measurements of the geometry of your face to create a unique code or facial “fingerprint”, I have taken nearly a thousand different measurements of each cell - including measurements describing the cell’s size, shape, location, and the quantities of certain biomolecules within it, to generate a cellular “fingerprint”. Using this technique I

have identified a number of potential new treatments for oesophageal cancer that only affect the fingerprints of the cancer cells and not the normal cells.

Following the identification of new treatments it is important to elucidate how they work - identifying the weakness that makes the treatment effective on the cancer cells but not the normal cells. We can do this in several ways. In this work I compared the drug induced fingerprints of the new treatments to known therapies to group those that act in the same way, and identify how the new drugs may work. I have also measured all the molecules within a cell and identified the ones that change in response to the therapy. This can give us a deeper understanding of which processes and pathways within a cell the therapy affects.

In oesophageal cancer, each patient's cancer is genetically very different from that within other patients. This is known as heterogeneity, and means that we have to identify which patients will benefit from which treatments. In order to do this we group cancers by the processes that are dysregulated in them or by the quantities of certain molecules within them. In the final section of this work I have attempted to explain what makes some cancers susceptible to the new treatments we have discovered, and identify the specific molecules within a cell that make it sensitive. These could be used in the future to predict which patients will most effectively respond to the treatment.

# *Contents*

Declaration ..... i

Acknowledgements ..... iii

Abstract ..... v

Lay Summary ..... vii

List of Figures ..... xiii

List of Tables ..... xv

1 | Introduction ..... 1

1.1 Oesophageal Cancer ..... 1

1.1.1 Previous screening studies in OAC ..... 5

1.2 Drug Discovery ..... 6

1.2.1 Target-based drug discovery ..... 7

1.2.2 Phenotypic drug discovery ..... 8

1.3 High Content Phenotypic Profiling ..... 9

1.3.1 Cell painting ..... 11

1.3.2 Image analysis ..... 12

1.3.3 Phenotypic fingerprints and mechanism of action studies ..... 14

1.3.4 JUMP-CP ..... 16

1.4 Phenotype to Genotype ..... 16

1.4.1 Biomarkers ..... 17
1.4.2 Functional genomic approaches ..... 18
1.4.3 Pharmacological screening in cell panels ..... 19
1.4.4 Pathway analysis ..... 20
1.4.5 Compound induced gene expression signatures ..... 21
1.5 Aims of the Project ..... 22
2 | Methods ..... 23
2.1 General Methods ..... 23
2.1.1 Cell culture ..... 23
2.1.2 Compound management ..... 25
2.1.3 Image analysis ..... 29
2.1.4 Data analysis ..... 30
2.1.5 Published datasets ..... 34
2.1.6 Seahorse ..... 35
2.1.7 ICP-MS ..... 38
2.2 Chapter 3: Morphology Predicts Mechanism ..... 39
2.2.1 Unsupervised machine learning: Clustering ..... 39
2.2.2 Supervised machine learning: Classification ..... 39
2.3 Chapter 4: High Content Profiling in Oesophageal Adenocarcinoma. ..... 41
2.3.1 Hit identification ..... 41
2.3.2 Hierarchical clustering ..... 42
2.3.3 Validation ..... 42_Toc65757428
2.3.6 Machine learning ..... 43
2.4 Chapter 5: Ammonium pyrrolidinedithiocarbamate, Disulfiram, and Elesclomol: Towards a unified theory of drug mechanism ..... 43

2.4.1 Patient derived organoids ..... 43
2.4.2 Differential gene expression analysis ..... 43
2.4.3 Copper chelator bathocuproinedisulfonic acid (BCS) ..... 44
2.4.4 N-acetyl-L-cysteine dose response ..... 44
2.4.5 $H_2O_2$ dose response ..... 44
2.5 Chapter 6: Pharmacogenomics of Disulfiram and Elesclomol ..... 44
2.5.1 Differential gene expression analysis ..... 44
2.5.2 Calculation of OAC $IC_{50}$ and NanoString gene expression analysis ..... 45
2.5.3 Calculation of pan cancer AUC and gene expression correlation ..... 45
2.5.4 Gene set enrichment analysis ..... 45
3| Morphology Predicts Mechanism ..... 47
3.1 Introduction ..... 47
3.2 Results ..... 49
3.2.1 Building a high content panel of OAC cell lines ..... 49
3.2.2 Predicting mechanism of action from phenotypic information ..... 51
3.3 Discussion ..... 61
4| High Content Profiling in Oesophageal Adenocarcinoma ..... 65
4.1 Introduction ..... 65
4.2 Results ..... 68
4.2.1 Primary hit identification ..... 68
4.2.2 Dose response validation ..... 71
4.2.3 Validation of machine learning and class prediction ..... 74
4.3 Discussion ..... 80
5| Ammonium pyrrolidinedithiocarbamate, Disulfiram, and Elesclomol: Towards a unified theory of drug mechanism ..... 85

5.1 Introduction ..... 85
5.2 Results and discussion ..... 87
5.2.1 Validation of copper chelating compounds in OAC ..... 87
5.2.2 Elesclomol, disulfiram and ammonium pyrrolidinedithiocarbamate act through a unified mechanism of action ..... 93
5.2.3 Mechanism of action ..... 97
5.3 Discussion ..... 113
6| Pharmacogenomics of Disulfiram and Elesclomol ..... 119
6.1 Introduction ..... 119
6.2 Results ..... 121
6.2.1 Selectivity over tissue-matched control cells ..... 121
6.2.2 Sensitivity to Disulfiram and elesclomol ..... 130
6.3 Discussion ..... 140
7| Conclusions ..... 143
8| References ..... 151
A.1| Appendices ..... 173

# *List of Figures*

1.1 Oesophageal cancer statistics....2
1.2 Oesophageal adenocarcinoma milestones. ....3
1.3 Drug discovery workflow. ....7
1.4 High content screening milestones. ....10
1.5 Cell Painting assay. ....11
1.6 Image analysis pipeline ....14
1.7 Compound class specific fingerprints. ....15
2.1 Random forests. ....32
2.2 Seahorse assay. ....36
3.1 Oesophageal adenocarcinoma cell panel. ....51
3.2 Library of reference compounds unsupervised machine learning. ....54
3.3 Hierarchical clustering heatmap for library of reference compounds. ....55
3.4 Unsupervised clustering of subtle phenotypic effects. ....57
3.5 Reference library supervised machine learning. ....59
3.6 Leave one cell line out-Reference Library Supervised Machine Learning.. 60
4.1 Hit analysis. ....69
4.2 Hierarchical clustering heatmap for the cell panel across the repurposing libraries. ....70
4.3 Univariate dose responses for 11 OAC selective compounds. ....72
4.4 Multivariate phenotypic dose responses for DNA damaging hit compounds. ....77
4.5 Colour combined images of reference classes and antimetabolites. ....79

4.6 Machine learning class probabilities for the antimetabolites. ..... 80
5.1 Dose responses for ammonium pyrrolidinedithiocarbamate, disulfiram and elesclomol. ..... 88
5.2 Phenotypic dose responses for ammonium pyrrolidinedithiocarbamate, disulfiram and elesclomol. ..... 90
5.3 Patient derived organoid dose responses for elesclomol. ..... 92
5.4 Correlation of disulfiram and elesclomol $IC_{50}$s. ..... 94
5.5 Disulfiram and elesclomol gene expression changes. ..... 97
5.6 Role of copper in compound activity. ..... 99
5.7 Role of reactive oxygen species in compound activity. ..... 101
5.8 Effect of elesclomol treatment on mitochondrial function. ..... 105
5.9 Elesclomol induced gene expression signature. ..... 107
5.10 Effect of elesclomol on highlighted cellular pathways. ..... 109
5.11 Comparison of phenotypes induced by compounds affecting proteostasis. ..... 111
5.12 Correlation of bortezomib and elesclomol induced gene expression changes. ..... 113
6.1 Selectivity over tissue-matched control cell lines. ..... 122
6.2 Treatment induced gene expression studies in tissue-matched control cell lines. ..... 124
6.3 ICP-MS copper. ..... 126
6.4 OAC versus tissue-matched control basal differential expression analysis 127
6.5 OAC versus tissue-matched control basal gene expression IPA pathway analysis. ..... 129
6.6 OAC comparison to pan cancer elesclomol sensitivity. ..... 131
6.7 Gene expression defining elesclomol sensitivity. ..... 133
6.8 GSEA Hallmarks analysis. ..... 135
6.9 GSEA KEGG analysis. ..... 136
6.10 Basal mitochondrial function. ..... 139

# *List of Tables*

2.1 Seeding densities. ..... 24
2.2 Mechanism of action reference compounds ..... 25
2.3 Library concentrations. ..... 27
2.4 Cell Painting reagents. ..... 29
2.5 Seahorse assay conditions. ..... 37
3.1 High content cell panel criteria. ..... 50
3.2 Random forest classifier accuracy for individually trained and leave-one-out classifiers. ..... 61
4.1 OAC selective compound hits. ..... 73
4.2 Screen and validation summary. ..... 74
5.1 Compound $IC_{50}$ values across panel of cell lines. ..... 88
5.2 Patient derived organoids. ..... 92

# *List of Acronyms*

| | |
|---|---|
| **2D** | Two-dimensional |
| **3D** | Three-dimensional |
| **5-FU** | 5-Fluorouracil |
| **ALDH** | Aldehyde Dehydrogenase |
| **AUC** | Area Under the Curve |
| **BCS** | Bathocuproinedisulfonic Acid |
| **CAT** | Catalase |
| **CCLE** | Cancer Cell Line Encyclopedia |
| **CPX** | Ciclopirox Olamine |
| **CRISPR** | Clustered Regularly Interspaced Short Palindrome Repeats |
| **DEPMAP** | Cancer Dependency Map Project |
| **DETC** | Diethyldithiocarbamate |
| **DFO** | Deferoxamine |
| **DMSO** | Dimethyl sulfoxide |
| **DSF** | Disulfiram |
| **EGFR** | Epidermal Growth Factor Receptor |
| **EMA** | European Medicines Agency |
| **ER** | Endoplasmic Reticulum |
| **FC** | Fold Change |
| **FCCP** | Trifluoromethoxy carbonylcyanide phenylhydrazone |

| | |
|---|---|
| **FDA** | Food and Drug Administration |
| **Fe-S** | Iron-Sulfur |
| **GDSC** | Genomics of Drug Sensitivity in Cancer |
| **GORD** | Gastro-Oesophageal Reflux Disease |
| **GSEA** | Geneset Enrichment Analysis |
| **GSH** | Glutathione |
| **HDAC** | Histone Deacetylase |
| **HER2** | Human Epidermal Growth Factor Receptor 2 |
| **HSR** | Heat Shock Response |
| **ICP-MS** | Inductively Coupled Plasma Mass Spectrometry |
| **IPA** | Ingenuity Pathway Analysis |
| **KEGG** | Kyoto Encyclopedia of Genes and Genomes |
| **LOPAC** | Library of Pharmacologically Active Compounds |
| **LXR/RXR** | Liver X Receptor-Retinoid X Receptor |
| **MOA** | Mechanism of Action |
| **MRSB2/3** | Methionine sulfoxide reductase B2/3 |
| **NAC** | N-acetyl-L-cysteine |
| **OAC** | Oesophageal Adenocarcinoma |
| **OCR** | Oxygen Consumption Rate |
| **PBS** | Phosphate Buffered Saline |
| **PCA** | Principal Component Analysis |
| **ROS** | Reactive Oxygen Species |
| **RTK** | Receptor Tyrosine Kinase |
| **T-SNE** | T-Distributed Stochastic Neighbour Embedding |
| **UPR** | Unfolded Protein Response |

# *1| Introduction*

## 1.1 Oesophageal Cancer

Combined, the two major histological subtypes of oesophageal cancer; oesophageal adenocarcinoma (OAC) and oesophageal squamous cell carcinoma (OSCC), represent the sixth leading cause of cancer deaths worldwide (Fitzmaurice *et al.*, 2015). Oesophageal cancer is an aggressive malignancy with inherent resistance to chemotherapy. This results in a high incidence of recurrence and metastasis, and is associated with poor prognosis and a 5-year survival rate of <20% (Herszényi and Tulassay, 2010). While the incidence of many cancers are predicted to decrease, by 2025 the incidence of OAC is expected to increase by 140% (Napier, Scheerer and Misra, 2014).

The two distinctive subtypes of oesophageal cancer differ with regards to etiology, ethnic distribution, pathogenesis, precursor lesion, and location in the oesophagus. Historically squamous-cell carcinoma has been the predominant form of oesophageal carcinoma worldwide and still is in East Asia (Jain & Dhingra, 2017). However, there has been a shift in epidemiology in western countries and the incidence of adenocarcinoma now exceeds that of squamous-cell carcinoma (Pennathur *et al.*, 2013).

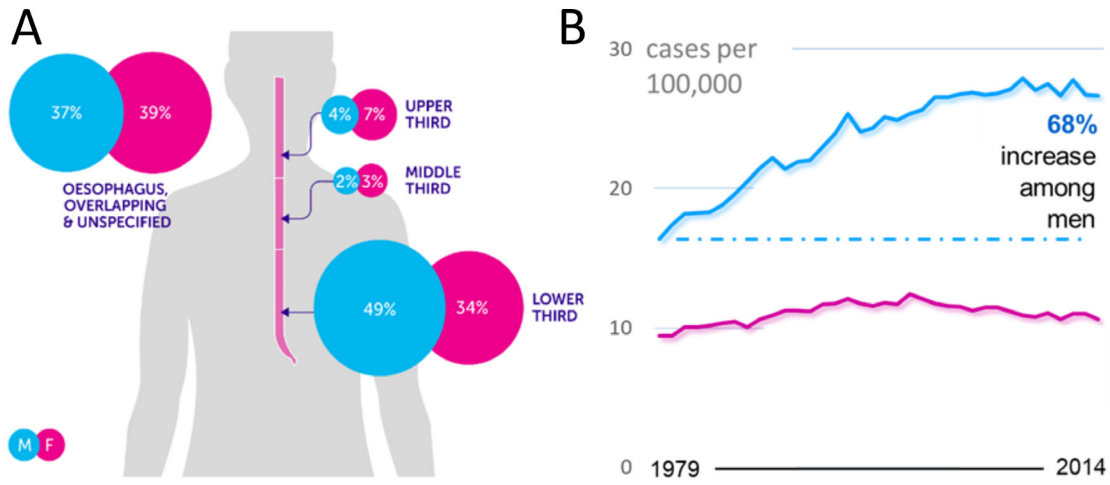



**Figure 1.1: Oesophageal cancer statistics.** A) Oesophageal adenocarcinoma occurrence by location in oesophagus for men and women. B) Oesophageal adenocarcinoma incidence in the UK split by gender between 1979 and 2014. Data from CRUK. Adapted from CRUK.

Forty years ago in the United States OSCC accounted for more than 90% of all oesophageal cancers. Now OAC is the leading cause of oesophageal cancer, representing 80% all of cases in the United States (Absi, Adelstein and Rice, 2010; Napier, Scheerer and Misra, 2014). Worldwide incidence of oesophageal adenocarcinoma has increased about three-fold since 1971 (Lepage *et al.*, 2008).

The oesophagus consists of four layers; the mucosa, the submucosa, muscularis and tunica adventitia. The inner most lining muscosa is made up of stratified squamous epithelium containing mucous glands. OSCC occurs in the proximal portion of the oesophagus while OAC occurs in the distal portion of the oesophagus and often at the gastro-oesophageal junction (Figure 1.1). The pathogenesis of OSCC is believed to be a multistep process, starting with a dysplastic precursor lesion where squamous epithelial cells exhibit nuclear atypia and abnormal maturation (Chen et al., 2017). However, the progression from dysplasia to OSCC is poorly understood. On the other hand, OAC develops via Barrett's oesophagus, a premalignant, nondysplastic disease state, through which a metaplastic change in stratified squamous epithelium occurs and is replaced with a columnar epithelium (Winters *et al.*, 1987; Hameeteman *et al.*, 1989; Guillem, 2005). This premalignant

disease state can progress through to high grade dysplasia and on to adenocarcinoma. Major risk factors for OAC include obesity and gastro-oesophageal reflux disease (GORD), while OSCC is linked to smoking and tobacco use (Jain & Dhingra, 2017). Symptoms often appear at the later stages of the disease, leading to a large proportion of late stage diagnoses. See Figure 1.2 for OAC milestones and discoveries.

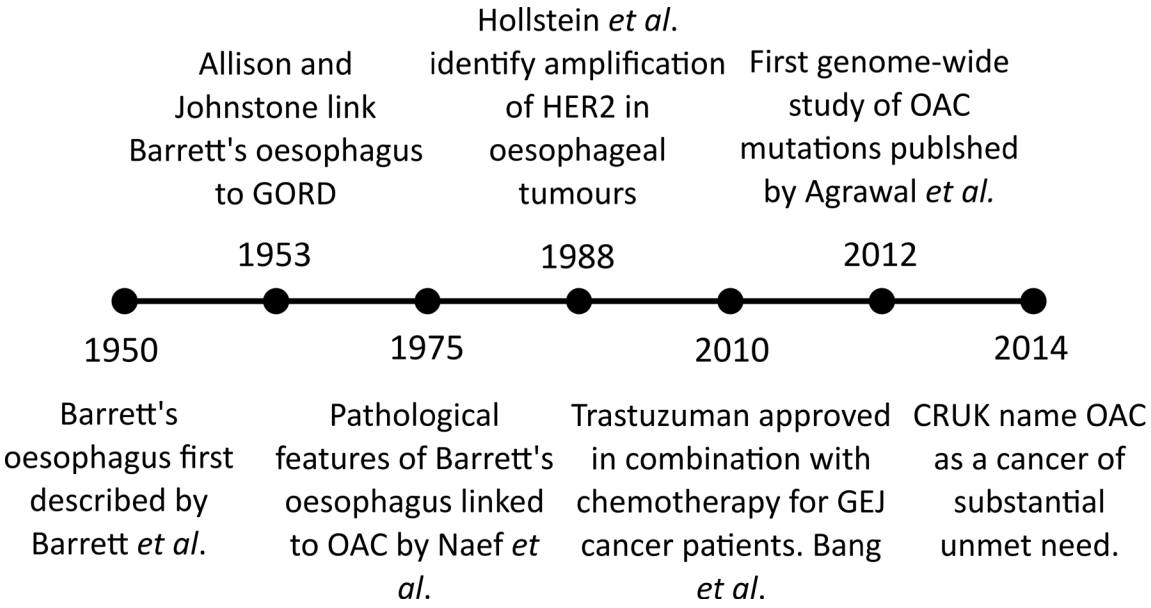



**Figure 1.2: Oesophageal adenocarcinoma milestones.** GORD - Gastro-esophageal reflux disease, OAC = Oesophageal adenocarcinoma, GEJ - Gastro-oesophageal junction, CRUK - Cancer Research UK.

For patients with organ-confined OAC the current standard-of-care is neoadjuvant treatment follow by surgical resection (Allum *et al.*, 2011). The optimal pre-operative treatment regime for OAC remains controversial with platinum-based neoadjuvant chemotherapy favoured in the UK and France and neoadjuvant chemoradiotherapy favoured in Ireland and the rest of mainland Europe (Messager *et al.*, 2016). Defining the optimal neoadjuvant regime is an area of active investigation (Reynolds *et al.*, 2017) as current treatments all carry a significant risk of systemic toxicity, histological response rates remain poor (Alderson *et al.*, 2017), and only a limited subgroup of patients experience any survival benefit over surgery alone (Noble *et al.*, 2017; O'Neill *et al.*, 2017).

The genomic landscape of OAC suggests it is a disease driven by structural variation rather than point mutations. Copy number alterations, structural rearrangements, and even genomic catastrophes including chromothripsis (30% of patients) are the predominant features (Secrier *et al.*, 2016). However, cancer genome sequencing studies have revealed several recurrently mutated genes in OAC, the most common of which is TP53, mutated in 81% of OAC cases, and associated with reduced overall survival (Fisher *et al.*, 2017). Second to TP53 is ARID1A, SMAD4 and CDKN2A, however these each account for less than 20% of OAC cases. Ordering of mutations in preinvasive disease stages revealed that only TP53 and SMAD4 mutations occur in a stage-specific manner, confined to high grade dysplasia and OAC respectively (Chung *et al.*, 2007; Weaver *et al.*, 2014).

The lack of driver mutations coupled with the fact that driver events are highly heterogeneous between cases, has made clinically meaningful subgroups and well validated therapeutic targets difficult to define. A study of the mutational signatures in OAC led to the identification of three etiological subgroups; C>A/T dominant, DNA damage response impaired, and mutagenic, and proposed subgroup specific treatment interventions (Secrier *et al.*, 2016). Whether there is any clinical relevance and therapeutic benefit to these subgroups has yet to be determined. More recent analysis has also proposed several genomic biomarkers, including CDK4 and CDK6 inhibitor sensitising events (Frankell *et al.*, 2019). However, a recent clinical trial has concluded that Palbociclib, a CDK4/6 inhibitor, has limited single-agent activity in gastroesophageal tumours, despite stratifying for predicted CDK4/6 inhibitor sensitive tumours (Karasic *et al.*, 2018, 2020). Further clinical trials with molecularly targeted therapies against receptor tyrosine kinases (RTKs) have been disappointing with small improvement in outcomes to date only observed in ERBB2 positive tumours (2.7 month increase in median overall survival) for Trastuzumab+chemotherapy combination relative to chemotherapy alone (Bang *et al.*, 2010). Alternative EGFR targeting therapies, Panitumumab or Gefitinib, either decreased overall survival or demonstrated no benefit in clinical studies (Bang *et al.*, 2010; Okines *et al.*, 2010; Dutton *et al.*, 2014; Hecht *et al.*, 2016). A potential explanation is the almost ubiquitous co-

amplification of alternative RTKs and downstream pathways leading to redundancy and drug resistance (Dulak *et al.*, 2012; Paterson *et al.*, 2013; Secrier *et al.*, 2016).

### 1.1.1 Previous screening studies in OAC

In order to try and identify new drugs and targets for OAC several small-scale screens have been carried out. The largest of which screened 215 FDA-approved and 163 epigenetic compounds and included two patient-derived oesophageal cancer cell lines as well as four established OAC cell lines (Lohse *et al.*, 2018). Here they discovered differing responses across the cell lines, highlighting the importance of cell panels in heterogeneous diseases. In an effort to better capture patient heterogeneity Li *et al.* carried out a screen of 24 anti-cancer compounds in nine patient-derived, ex-vivo organoids (Li et al. 2018). Again, they noted a range of sensitivities across the organoids, but saw increased sensitivity towards RTK inhibitors across the organoids, but unfortunately, as previously mentioned, these have shown limited success in clinical trials so far (Bang *et al.*, 2010).

siRNA screens have also been carried out in OAC to look for genetic vulnerabilities including 714 siRNAs targeting kinase and kinase related genes (Campbell *et al.*, 2016) in seven OAC cell lines as part of a pan-cancer study. RTK addiction, particularly ERBB2, and MYC amplification were identified in the OAC cell lines. They also demonstrated that integrating molecular profiling and protein-protein interaction data with siRNA screening data can predict drug sensitivity in genetically defined subsets of cancer. The information from this study was later combined with a screen using 80 compounds of approved and late stage cancer therapeutics to identify biomarker-defined therapeutic approaches for the treatment of OAC (Chong *et al.*, 2018). Here it was found that MYC and ERBB2 amplification led to increased sensitivity to BTK inhibitors. More recently an immune related signalling pathway linked to OAC cell survival was discovered in a large scale siRNA screen of 6022 siRNAs targeting the druggable genome in three OAC cell lines (Duggan *et al.*, 2018). This information was used to predict fostamatinib and its off-target inhibitory effects on the downstream immune

pathway as an effective treatment for OAC, which was then confirmed in a mouse model (Duggan *et al.*, 2018). A handful of OAC cell lines have also been included in two of the largest pan-cancer screening projects; The Cancer Dependency Map (DepMap) (Tsherniak *et al.*, 2017) and the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang *et al.*, 2013).

These small-scale studies of approved compounds in OAC may provide insights into therapeutic options for the treatment of OAC, as evidenced by the shared identification of RTK addiction and inhibition, a strategy currently being trialled for the treatment of OAC. However, they lack the breadth of large compound screens and novel chemical space to provide new insights into the disease itself or novel targeted therapeutics.

## 1.2 Drug Discovery

Drug discovery; the process by which new candidate drugs are discovered (Figure 1.3).

Historically, drugs were discovered through the identification of active ingredients from traditional remedies or by serendipitous discovery, as with penicillin (Abraham and Chain, 1940). Over time, and with advances in chemistry, this progressed to the screening of substances, whether crude extracts or purified chemicals, for biological activity without knowledge of the specific target, often carried out in whole organisms. This approach is known as classical pharmacology, forward pharmacology, or phenotypic drug discovery. However, the advent of molecular biology, and in particular, of genomic sciences led to a new concept in drug discovery- target based drug discovery (Drews, 2000).

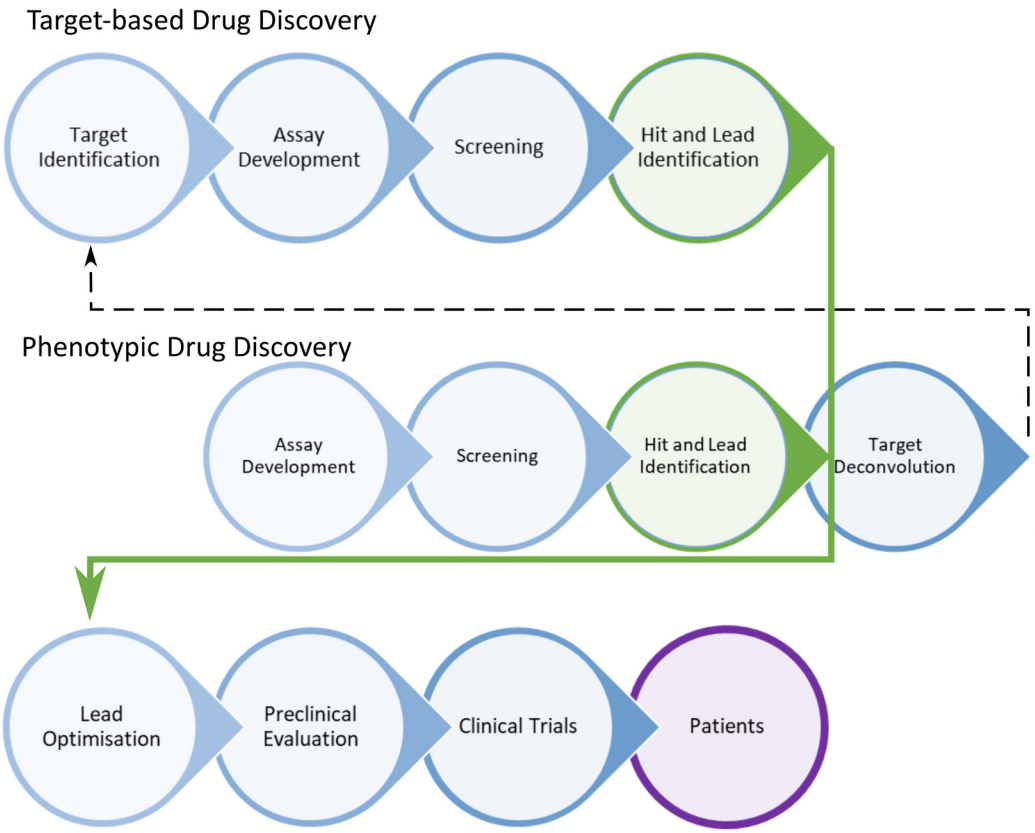



**Figure 1.3: Drug discovery workflow.** Target-based drug discovery starts with target identification whereas phenotypic drug discovery often finishes with this. Both target-based and phenotypic drug discovery pipelines lead to the identification of hit and lead molecules which then feed into lead optimization and progress through the rest of the drug discovery pipeline, with successful drugs eventually gaining approval and making it into the clinic. Phenotypic discoveries can further feed into target based approaches once target deconvolution occurs for the generation of further drugs acting on the newly identified target.

### 1.2.1 Target-based drug discovery

In the past three decades, much of the drug discovery industry has embraced a genomics-derived, target-based approach to the discovery of new classes of drugs to treat diseases, driven by advances in molecular biology and genomics (Moffat *et al.*, 2017). This target-based drug discovery approach, relies on the identification of a defined molecular target hypothesised (and hopefully validated) to play in important role in the disease. Compound libraries are then screened against the

target to look for those that cause the desired manipulation of target, most commonly inhibition.

With a molecular target in hand, target-based drug discovery efforts can utilize crystallography, biochemistry, binding kinetics, computational modelling, and mutational analysis to understand how a drug interacts with a target, enabling efficient structure–activity relationship (SAR) assessment, biomarker development, and the efficient generation of further drugs to act on the target. This is a major benefit of target- based drug discovery over the alternative, phenotypic drug discovery, and probably the reason that assessment of new molecular entities from the pharmaceutical industry over a 10 year period demonstrated target-based drug discovery to be the forerunner for follower drugs (Swinney and Anthony, 2011). Furthermore, between 1999 and 2013, of the 113 first-in-class drugs approved by the FDA, 70% were identified through target-based drug discovery (Eder, Sedrani and Wiesmann, 2014), attesting to the success of the strategy. However, while target-based drug discovery has many success stories, there are a few drawback to the approach. Firstly, while target-based drug discovery pulls ahead in overall new molecular entity approval rates, it falls behind phenotypic drug discovery for the identification of first-in-class molecules (Swinney and Anthony, 2011), something that is vital to the expansion of treatment options, the prevention of resistance, and the identification novel therapeutics. Another concern with target-based approaches is that recombinant systems can lead to oversimplification of complex disease aetiology, resulting in the pursual of projects and compounds that fail to translate into clinical results (Croston, 2017). This is reflected in the high attrition rates associated with target-based drug discovery (Waring *et al.*, 2015).

### 1.2.2 Phenotypic drug discovery

Phenotypic drug discovery is re-emerging as a valuable drug discovery approach (Swinney and Anthony, 2011; Moffat, Rudolph and Bailey, 2014; Moffat *et al.*, 2017). Unlike target-based drug discovery, it does not rely on the prior knowledge of a specific disease target, but instead uses biologically relevant assays with

quantifiable phenotypic endpoints to identify molecules with desirable disease altering properties. Phenotypic drug discovery encompasses a large variety of assays and endpoints, from univariate cell-line assays, to multivariate whole-organism assays.

While the advent of sequencing is often attributed to the uptake of target-based approaches, it is also likely the reason for the revival of phenotypic drug discovery. Next generation sequencing has revealed molecular heterogeneity in underlying disease mechanisms within and between patients in many disease types. Consequently, the “one disease-one target” assumption that target-based drug discovery makes grossly oversimplifies disease mechanisms, which are more likely driven by complex signalling networks within the biological interactome (Hopkins, 2008; Aguirre-Plans *et al.*, 2018). This highlights several favourable attributes of phenotypic drug discovery including the ability to; screen molecules in complex biological systems with intact signalling pathways designed to more closely model distinct disease states; expand novel drug target space; and identify hit molecules with any number of targets. Thus, phenotypic drug discovery may prove a beneficial strategy for complex, heterogeneous diseases were target biology is poorly understood and where modern target directed drug discovery strategies have made little impact on patient care, as exemplified in OAC.

## 1.3 High Content Phenotypic Profiling

A picture is worth a thousand words. An image is worth a thousand data points.

High content profiling, was introduced in 1997 (Giuliano *et al.*, 1997), and utilises high throughput microscopy and automated image analysis, to generate multivariate phenotypic datasets from images of biological systems ranging from standard 2D cell culture (Perlman *et al.*, 2004; Bickle, 2010), to 3D organoid cultures (Du *et al.*, 2020; Renner *et al.*, 2020), and even *in vivo* live imaging in whole organisms such as zebrafish (Geoffrey Burns *et al.*, 2005). The key development in the progression of the field was the development of the first high content platform, the ArrayScan, that converted semi-automated light microscopy into an automated

platform that did not require user intervention once a microplate scan was initiated (Lansing Taylor, 2010). See Figure 1.4 for a historical perspective of high content screening.

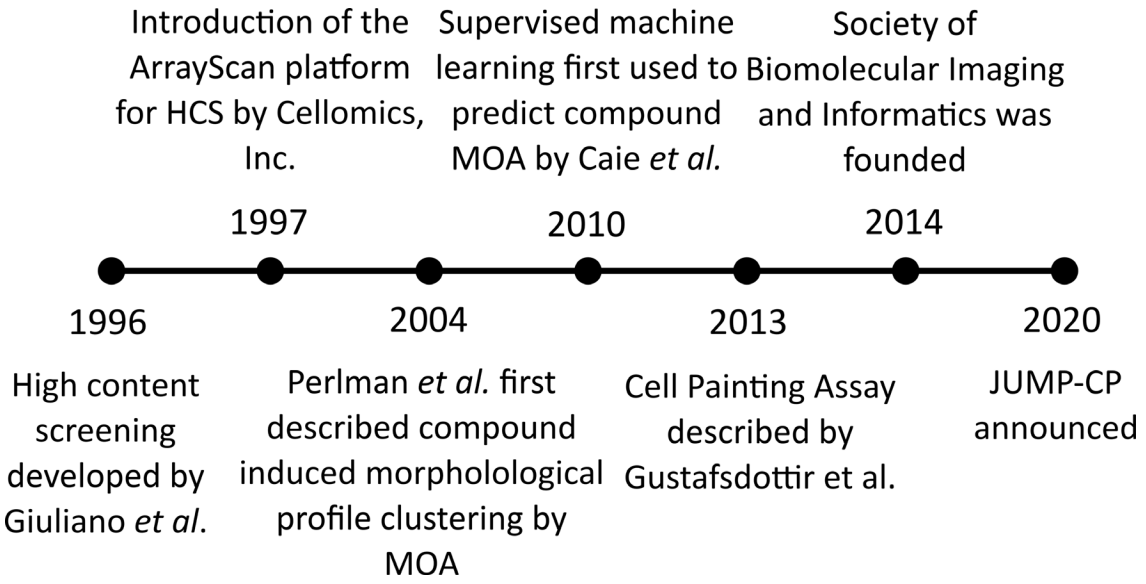



**Figure 1.4: High content screening milestones.** HCS - high content screeing, MOA - mechanism of action. JUMP-CP - a new high content imaging consortium (See section 1.2.7 JUMP-CP for further details).

Cell based high content profiling involves the capture of biological information relating to the cellular state that is contained in multiple cellular markers including those associated with morphology. Upon perturbation; chemical, genetic, or otherwise, cells undergo detectable changes in a subset of profiled features, allowing quantitative data to be extracted from microscopy images and a phenotypic fingerprint to be assigned to each perturbation (Perlman *et al.*, 2004; Young *et al.*, 2008; Caie *et al.*, 2010; Ljosa *et al.*, 2013). This is because manipulation of cell function and gene expression via perturbations, impacts on a cells morphological phenotype (Kamath and Ahringer, 2003; Kiger *et al.*, 2003; Perlman *et al.*, 2004). This allows the systematic screening of perturbations and the quantification of induced morphological changes for the selection of hits based on a desired phenotypic endpoint (Giuliano *et al.*, 1997; Kamath and Ahringer, 2003; Perlman *et al.*, 2004; Bickle, 2010; Rose *et al.*, 2018). Since these morphological changes are a result of the pathways and/or networks affected by the

perturbations, high content screening is capable of detecting changes that arise from modulation at the pathway/network level making it suitable for studying complex diseases, multi-target drugs, and drug combinations.

### 1.3.1 Cell painting

In order to get a complete picture of cellular morphology in a high content assay, broad morphological information is required. To address this need, in 2013 the Cell Painting assay was developed as an unbiased, cell based, morphological profiling assay capable of capturing subtle changes in morphology. It combines six fluorescent dyes, imaged in five channels, to visualise seven cellular components and organelles; nuclei, F-actin, endoplasmic reticulum, mitochondria, golgi, plasma membrane, and nucleoli (Gustafsdottir *et al.*, 2013; Bray *et al.*, 2016) (Figure 1.5).

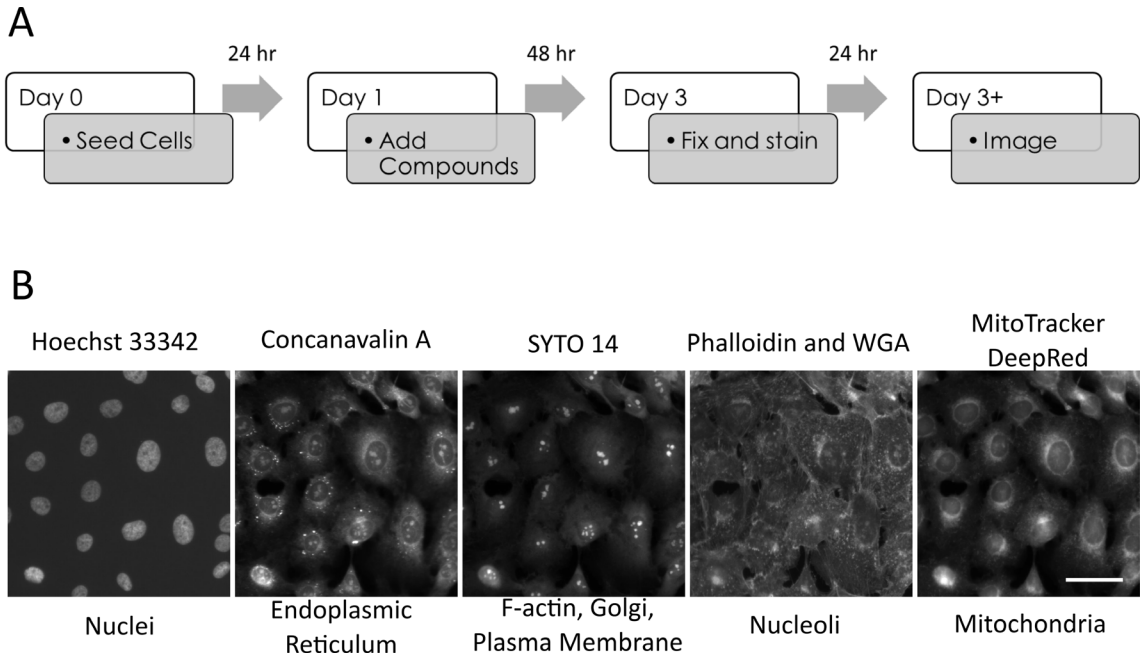



**Figure 1.5: Cell Painting assay.** A) Assay workflow. B) The five channels imaged in the Cell Painting assay with dyes and cellular structures labelled. Scale bar is 50 µm. See Table 2.4 for additional details about the stains and channels imaged.

Quantitative data are then extracted from the microscopy images to identify the phenotypic impact of chemical perturbations. Due to the broad morphological information contained in the images, Cell Painting has been exploited for many research applications, including mechanism of action prediction.

### 1.3.2 Image analysis

In high content profiling, image analysis is the process by which raw images taken on a high throughput microscope are converted into a dataset of numerical feature measurements used to describe the morphology of the cell or organism in the assay.

The gold standard approach to extracting numerical features from cell based screening is through object identification, image segmentation, and extraction of single-cell morphological measurements. The process usually starts with the identification and segmentation of nuclei from the background. There are a number of well-established algorithms to do this, and most use intensity thresholding to binarise an image into “object” and “background” based on a histogram of the pixel intensities (Otsu, 1979; Padmanabhan, Eddy and Crowley, 2010). The position of the centre of each nuclei is then used as a “seed” location to detect cellular boundaries in the image containing a cytoplasmic marker. This is done through expansion out from the seed locations up to the cell plasma membranes using a change in cytoplasmic marker intensity to define boundaries and edges.

After cellular and any further subcellular objects have been identified and segmented, morphological characteristics are then measured for each object. These measurements can relate to a wide variety of characteristics depending on the aims of the assay, although they can be loosely grouped into four classes:

Shape – Calculated from the segmented object properties, shape features such as area, perimeter, and eccentricity provide robust, interpretable measurements relating to the shape of an object, be it the cell, the nucleus, or other subcellular object.

Intensity – Measures of pixel intensity values within the object boundaries. They can be calculated for multiple channels and include measurements such as average intensity, integrated intensity, and radial distribution of intensity.

Texture – Patterns of intensities within object boundaries. They can be used to quantify morphologies such as puncta and speckles, as well as fibres and lines across any given imaging channel or dye.

Spatial context – Relationships between objects in an image. These include simple object count numbers giving the cell count in an image; number of neighbouring objects, such as cells or nuclei; percentage of cell boundary in contact with neighbouring cells; and even co-localisation across two or more of the channel dyes.

These image features are then aggregated into an $n$ x $m$ dataset, where $n$ is the number of objects and $m$ is the number of features. Common practise within the field of high content profiling is the aggregation of the single cell level data to an image, micro-titre well, or treatment level, often using the median value (Caicedo *et al.*, 2017). This then gives a treatment (or other aggregation level) a phenotypic "fingerprint" (Figure 1.6).

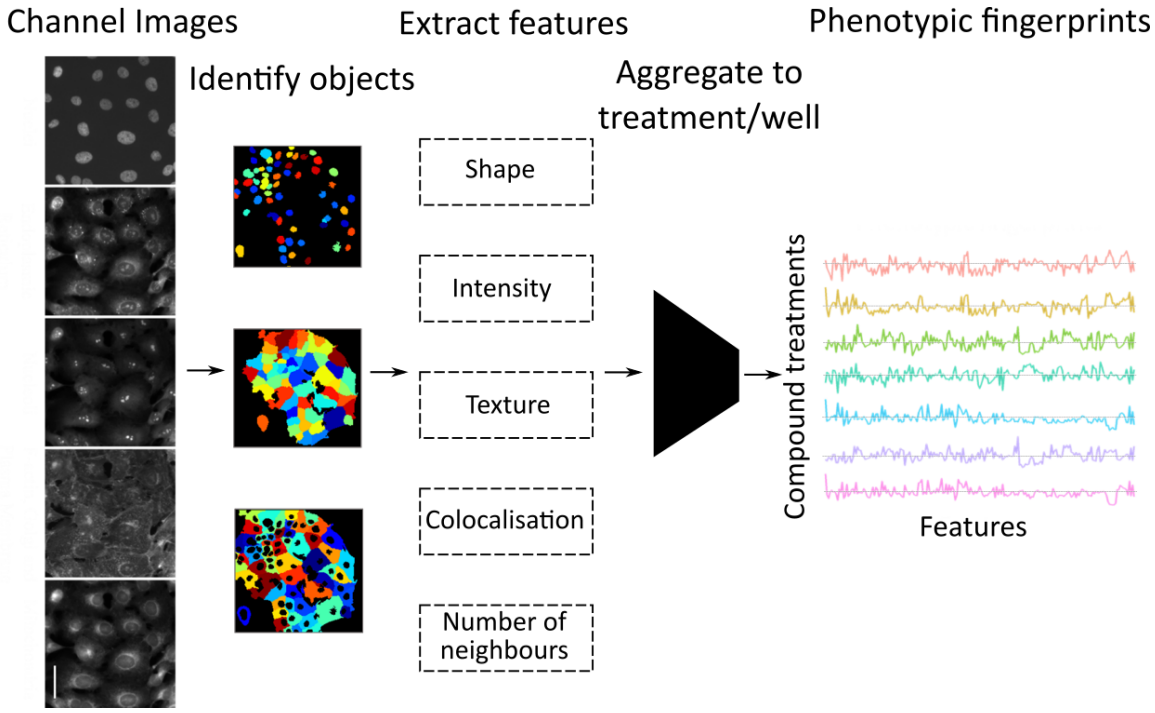



**Figure 1.6: Image analysis pipeline.**

### 1.3.3 Phenotypic fingerprints and mechanism of action studies

High content profiling assays have been exploited for many research applications but one area of significant research is the elucidation of compound mechanism of action from the primary assay readouts.

Since the morphological changes associated with a compound treatment are a consequence of its effects on particular cellular pathways, compounds acting on a given pathway elicit more similar morphological responses than those acting on different pathways. This means that the morphological information in the images can be used to identify phenotype altering compounds and elucidate mechanism of action simultaneously. This was first demonstrated in 2004 when Perlman *et al.* showed that compound induced morphologies could be grouped by mechanism of action using similarity measures and hierarchical clustering (Perlman *et al.*, 2004). Since then, numerous research groups have implemented unsupervised clustering methods to group treatments based on phenotypic response (Carpenter *et al.*, 2006; Young *et al.*, 2008; Gustafsdottir *et al.*, 2013; Ljosa *et al.*, 2013). More recently,

mechanism of action has been inferred by similarity comparison to reference compound sets using multivariate statistics and a variety of supervised machine learning techniques. One of the earliest studies using reference libraries set out to compare phenotypic profiles and draw mechanistic inferences from a number of reference compounds using *k*-nearest neighbour proximity labelling to predict the mechanism of hits (Caie *et al.*, 2010). More recently this has been achieved through use of trained classifiers (Warchal, Dawson and Carragher, 2019; Dawson *et al.*, 2020; Hughes *et al.*, 2020).

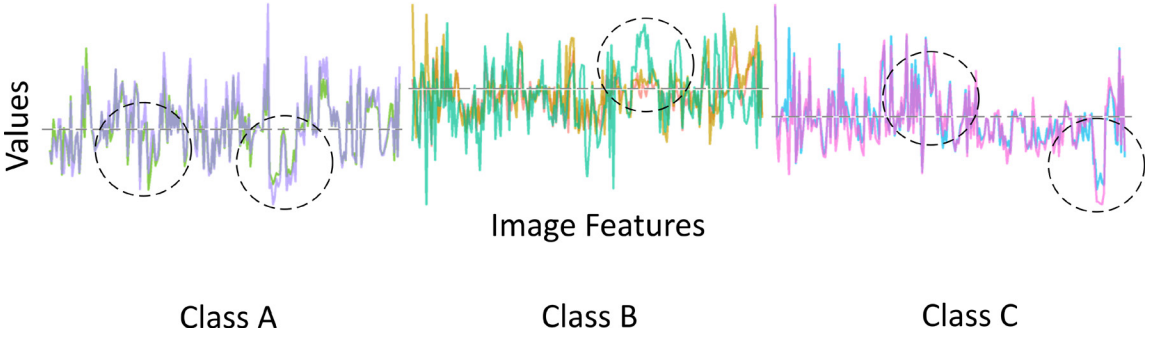



**Figure 1.7: Compound class specific fingerprints.** Overlays of image extracted features for multiple compounds belonging to one of three classes. Each class shows reproducible and unique aspects to the fingerprints (highlighted in ellipses) allowing distinction of classes by morphological image features.

Lastly, deep learning is an area of growing interest in the field of biological image analysis and is being applied to a number of problems including classification of small-molecule mechanism of action. Several methods have been developed to predict compound mechanism of action including supervised networks trained on segmented or whole images (Kraus, Ba and Frey, 2016; Godinez *et al.*, 2017), self-supervised methods using pseudo-labels (Ando, McLean and Berndl, 2017; Godinez, Hossain and Zhang, 2018), and even unsupervised methods using just image intensity values, making segmentation and feature extraction unnecessary steps (Janssens *et al.*, 2020).

### 1.3.4 JUMP-CP

Very recently a new initiative was announced to create an academic-industry cell imaging consortium; The Joint Undertaking in Morphological Profiling with Cell Painting (JUMP-CP), which aims to create the largest Cell Painting dataset of cells imaged under more than 140,000 different conditions including both small molecule and genetic perturbations, acting as a morphological atlas to match diseases, drugs and genes (https://jump-cellpainting.broadinstitute.org/). This resource will provide a huge reference dataset to advance the field of cell-based phenotypic screening and enhance applications in predicting mechanism of action, cellular toxicities, and drug repurposing opportunities.

## 1.4 Phenotype to Genotype

With the development of genome sequencing came the promise of functional genomics. While we are still a long way from understanding the function of many genes, substantial progress has been made, owing to the study of single genes in model organisms such as Drosophila (Benfey and Mitchell-Olds, 2008). However, reversing this process to go from phenotype to genotype is far more complex. This is because phenotypes are most often the result of a complex interactions. This holds true when trying to unravel genes involved in compound induced phenotypes for target identification owing to the interaction of compounds with multiple targets (Lu *et al.*, 2012), and for identification of predictors of drug response, due to the interaction of mutations and expression status across multiple genes and biological pathways (Nguyen, Dang and Ballester, 2017).

Pharmacogenomics is the study of how genes affect drug response and is the key to personalised medicine. It is thought that its history may stretch back more than 2000 years to when Pythagoras noted that ingestion of the fava bean caused potentially fatal reactions only in a subset of people (Nebert, 1999; Pirmohamed, 2001). Although the underlying reason for this was not known at the time, this has since been identified as haemolytic anaemia caused by a glucose-6-phosphate dehydrogenase deficiency (Carson *et al.*, 1956). Others propose that Snyder's study

concerning the 'phenylthiourea nontaster' phenotype in 1932 (Snyder, 1932) represents the dawn of modern pharmacogenomics. However, it wasn't until the early 1990s, with the advances in DNA technology, that the term pharmacogenomics was introduced, encompassing the use of genomics and pharmacology to develop rational means to optimise drug therapy. For example when a gene variant is associated with a particular drug response in a patient, there is the potential for making clinical decisions based on genetics by choosing which drug to give a patient.

### 1.4.1 Biomarkers

A biomarker is a naturally occurring molecule, gene or characteristic by which a normal biological process, a pathological process or a biological response to a therapeutic intervention can be evaluated (Atkinson *et al.*, 2001). There are three main types of biomarkers: Diagnostic and prognostic biomarkers, used to confirm presence of disease and measure likelihood of a clinical event such as recurrence and progression; efficacy biomarkers, which measure response to a treatment; and predictive biomarkers, used to predict treatment response.

Clinical response to anticancer therapies are often restricted to a subset of patients. The identification of genetic biomarkers which predict patient response to therapy have had a major impact upon clinical trial design and patient selection for the treatment of cancers with molecular targeted therapies (McDermott, 2015). The objective of personalized medicine is to identify those subgroups of patients which are most likely to respond to therapy and thus increase clinical trial success rates and improve patient outcomes. Examples of drug sensitizing mutations used in cancer personalized medicine strategies include the use of BRAF mutant status in melanoma for Vemurafenib (Chapman *et al.*, 2011), Debrafenib (Hauschild *et al.*, 2012), and Trametinib treatment (Flaherty *et al.*, 2012; Robert *et al.*, 2015); EGFR mutation in non-small-cell lung cancer for Afatinib treatment (Chih-Hsin Yang *et al.*, 2013; Sequist *et al.*, 2013); HER2 amplification in breast cancer for Trastuzumab treatment (Verma *et al.*, 2012); and BRCA1/2 in ovarian cancer for Olaparib treatment (Kaufman *et al.*, 2015).

Biomarker identification presents a bottleneck in the drug discovery process for phenotypic drug discovery. This is because unless screening in isogenic cell pairs with defined genetic backgrounds, the target(s), pathway(s), and genetic alterations influencing drug efficacy are for the most part unknown. While it is not a prerequisite for the progression of candidate drugs, there is a preference to carry out target deconvolution and identify the targets responsible for a compounds phenotypic activity (Moffat *et al.*, 2017; Vincent *et al.*, 2020). This can aid further understanding of disease mechanisms, the ability to integrate *in silico* and target based approaches for the development of second-in-class and follower molecules, and the identification of biomarkers.

While high content profiling data has been used to infer compound mechanism, it is more difficult to infer at the level of compound target. This was demonstrated in a recent study showing compounds with different annotated cellular targets can induce a shared phenotypic response (Schneidewind *et al.*, 2020). Thus, specifically linking phenotype to genotype for both target and biomarker identification requires integration with other profiling platforms and technologies.

### 1.4.2 Functional genomic approaches

A common approach to identifying the genomic causes of a phenotypic response is a functional genomic screen to identify targets that when knocked down or out convey alterations in sensitivity to the compound of interest. Until recently RNA interference (RNAi) was the predominant method for genome-wide loss-of-function screening, but its utility is limited by the lack of complete protein depletion and confounding off-target effects (Echeverri *et al.*, 2006; Jackson *et al.*, 2006). The clustered regularly interspaced short palindrome repeats (CRISPR) - based gene-editing system (Jinek *et al.*, 2012; Cong *et al.*, 2013; Mali *et al.*, 2013) combined with pooled or arrayed RNA-guide libraries (Shalem *et al.*, 2014; Wang *et al.*, 2014) has enabled the evaluation of highly specific and penetrant gene knockout and loss-of-function effects on cell fitness genome-wide in hundreds of cancer cell lines. These methods have provided rich functional resources to

explore cancer vulnerabilities and to identify candidate drug targets, mechanisms, and biomarkers.

Functional genomic approaches to biomarker predictions, such as CRISPR screens, assume sensitivity to drugs depends on a single gene, and indeed single gene-drug associations have been identified for a number of common anti-cancer therapies (see 1.3.1 Biomarkers). However, this is less common than at first thought, with single gene-drug associations only rarely able to explain the range of drug sensitivities (Garnett *et al.*, 2012; Gonçalves *et al.*, 2020). A second problem with CRISPR studies is the inability to study the effects of gene/protein expression as a continuous variable as it is in biological systems, rather than as a more binary on-off system.

### 1.4.3 Pharmacological screening in cell panels

Pharmacological screens have been used to profile the activity of hundreds of compounds across large, annotated collections of cancer cell lines with the aim of identifying molecular markers of drug sensitivity to guide clinical development. These allow the assessment of proteomic, transcriptomic, metabolomic, and genomic data to provide a powerful biomarker discovery platform to guide rational cancer therapeutic strategies.

Recent large scale initiatives in cancer cell line panel screening have produced a number of resources to help connect phenotype to genotype through the identification of genetic dependencies and small molecule sensitivities. The Cancer Dependency Map Consortium is an extension of the Broad Institutes DepMap project, an initiative to systematically catalogue and identify biomarkers of genetic vulnerabilities and drug sensitivities in hundreds of cancer cell lines. DepMap combines the Cancer Cell Line Encyclopedia (CCLE) with Project Achilles, a catalogue of genetic vulnerabilities using genome-wide RNAi and CRISPR loss-of-function to identify genes essential for cancers, and a comprehensive cell line drug sensitivity resource (Tsherniak *et al.*, 2017). Another project has led to the Genomics of Drug Sensitivity in Cancer (GDSC), a resource for therapeutic biomarker discovery in cancer cells, combining large scale drug screening and

genomic analyses (Yang *et al.*, 2013). These databases have been used to aid phenotypic drug discovery and elucidate mechanism of action (Rees *et al.*, 2016; Gonçalves *et al.*, 2020) and identify genomic markers of drug sensitivity (Garnett *et al.*, 2012). For example, the sensitivity of a small-molecule compound ML239, originally identified in a phenotypic screen to exhibit selectivity for breast cancer stem cells, was profiled across a large cell panel and correlation with basal gene expression and follow up studies indicated that fatty acid desaturase 2 (FADS2) activity is required for ML239 sensitivity (Rees *et al.*, 2016).

### 1.4.4 Pathway analysis

The problem remains that compounds and genes are often distilled down to single compound-target interactions to link phenotypes with genotypes and such perturbations are not viewed within the context of integrated pathway signalling networks leading to lack of insight for complex diseases and compounds with multiple targets. This is often true for both target and biomarker identification and can lead to exclusion of patients from clinical trials and therapies that they would actually benefit from. For example, 15 years ago, *Farmer et al*. discovered that BRCA1 and BRCA2 dysfunction profoundly sensitises cells to Poly(ADP-ribose) polymerase (PARP) inhibitors (Farmer *et al.*, 2005). In 2014, Olaparib was approved by the EMA and FDA for the treatment of patients with ovarian cancer with mutations in the BRCA genes (Kaufman *et al.*, 2015). However, the same phenotype; DNA repair deficiency, can be brought about by many genomic events (Lee, Ledermann and Kohn, 2014), highlighting the need to identify biomarkers at the pathway level.

Leveraging knowledge contained in pathway databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), Reactome (Vastrik *et al.*, 2007; Croft *et al.*, 2014), BioCarta (Nishimura, 2001), a pathway models genes and gene products as nodes in a network, and edges symbolise interactions among them based on prior knowledge and predictions from experimental work and computational analysis. Pathway analysis approaches

often overlay gene expression data on these databases to build networks and identify pathways which are significantly impacted in a given condition.

The rationale behind pathway based methods was to tackle the emergence of complex phenotypes, as seen in predicting tumour response to therapy, that are often caused by a myriad of biological interactions and signalling events (Hernández-Lemus and Martínez-García, 2021).

### 1.4.5 Compound induced gene expression signatures

Genome-wide transcriptomic and proteomic changes in response to small-molecule treatment and can provide insight into the cellular pathways and processes deregulated by treatment. Gene expression profiling can also be used in a compound exposure setting to generate compound induced gene expression signatures. Comparison of signatures can be used to discover new connections among compounds as well as to identify targets and mechanisms of action in a more complex setting than single gene contributions. Systematic profiling of gene expression signatures following perturbation was first realised 20 years ago, when Hughes *et al.* showed that expression data from yeast deletion mutants and drug treatments could yield insights into gene functions and drug targets (Hughes *et al.*, 2000). Since then, the development of the Connectivity Map project has enabled a systematic comparative analysis of gene expression profiles (Lamb *et al.*, 2006; Subramanian *et al.*, 2017) across large panels of pharmacologically perturbed samples to support similarity profiling of gene expression patterns and infer compound mechanism of action (Lamb *et al.*, 2006) and vulnerability biomarkers to guide clinical interventions (Milioli *et al.*, 2015).

The Connectivity Map concept was originally based on data provided from gene expression microarrays. However, advances in gene expression profiling, such as L1000 (Subramanian *et al.*, 2017) and NanoString (Geiss *et al.*, 2008), allow profiling at a fraction of the cost, and in small samples. These technologies profile just a subset of genes, with L1000 profiling just over 1000 landmark and control genes before inferring global gene expression computationally (Subramanian *et al.*, 2017), and NanoString which profiles targeted panels of genes across pathways of

interest combined with pathway and network analysis can infer the status of pathways, upstream regulators, and downstream targets. Further, high throughput, multiplexed expression profiling with DRUG-Seq (Ye *et al.*, 2018) and MIX-seq (McFarland *et al.*, 2020) allow perturbations and cell type multiplexing at scale and at single cell resolution. These technologies, combined with other omic datasets, including metabolomics, promise to extend the application of the connectivity map concept across larger compound libraries and more complex biological samples for the identification of drug similarities, mechanisms, and biomarkers.

## 1.5 Aims of the Project

In the chapters that follow, I describe the development and validation of a high content profiling assay with integrated machine learning and transcriptomic pathway analysis to identify drug repurposing opportunities in OAC and provide patient stratification hypotheses to support clinical translation.

Chapter 1 aims to predict drug mechanism of action across heterogeneous panels of cell lines using cellular morphological changes. It then explores the use of this to develop a multiparametric assay quality control measure to assess cell lines and develop a high content screen for OAC.

Chapter 2 aims to identify compounds and target classes selective for OAC and suitable for repurposing or new target-based drug discovery pipelines.

Chapter 3 aims to identify the anti-cancer mechanism of action for three of the most selective and potent compound hits identified in the previous chapter.

Chapter 4 aims to leverage the benefits of cell panels and integrated transcriptomic data to explore genes and pathways responsible for the sensitivity and selectivity of the top three hits, as part of a personalised medicine strategy.

# 2| *Methods*

The methods detailed here relate only to work directly performed by myself or in direct collaboration with others. Some work pertinent to the results, though not performed by myself, is therefore not included unless strictly necessary. The thesis as a whole makes use of both novel and previously published datasets, and only those datasets produced *de novo* for the purposes of this thesis are described in detail, but access and links to published datasets are included.

## 2.1 General Methods

These methods are used throughout the work in this thesis and are listed here to reduce repetition. Specific methods relating to each given chapter are described subsequently.

### 2.1.1 Cell culture

#### Cell line authentication

Cell line authentication (not carried out for the EPC2-hTERT line, as there is no reference sequence) was confirmed by short tandem repeat (STR) genotyping (Cell Line Authentication, Public Health England). Cells were trypsinised and re-suspended in media (5 x $10^4$ cells/mL) and blotted (50 µL) on the provided cards. Results confirmed all cell lines tested.

The cell lines were confirmed to be mycoplasma negative using the Venor™GeM Mycoplasma Detection PCR kit (Sigma; #MP0025) and routine testing was implemented to ensure no subsequent infections occurred.

### Cell subculture

Oesophageal adenocarcinoma (OAC) lines were grown in Roswell Park Memorial Institute (RPMI) (Life Technologies; #11875101) supplemented with FBS (10 %, Life Technologies; #16140071) and L-glutamine (2 mM, Life Technologies; #A2916801) and incubated under standard tissue culture conditions (37 °C and 5 % $CO_2$). The Barrett's oesophagus line CP-A and the oesophageal epithelial line EPC2-hTERT were grown in KSFM (Life Technologies; #17005075) supplemented with human recombinant epidermal growth factor (5g/L) and bovine pituitary extract (50 mg/L).

For subculture of OAC lines, growth media was aspirated and cells were washed with phosphate buffered saline (PBS) (1X, 10 ml, pH 7.4), detached with trypsin (0.25 %, 1 mL, 5 minutes, 37 °C), and neutralised with fresh growth media. For subculture of CP-A and EPC2-hTERT lines, soy-bean trypsin inhibitor (250 mg/L, 5 mL) was added to neutralise the trypsin and centrifuged (5 minutes, 250 x g). The cell pellet was then re-suspended in fresh media.

### Cell plating in 384-well plates (Reference Library, Primary Screen, Validation)

Cells were seeded (50 µL per well) (Table 2.1) into 384-well, CELLSTAR® Cell Culture Microplates (Greiner, #781091), and incubated under standard tissue culture conditions for 24 hrs before the addition of compounds.

**Table 2.1: Seeding densities.** Cells per well.

| Cell Line | Seeding density |
|---|---|
| **JH-EsoAd1** | 1500 |
| **CP-A** | 800 |

| | |
|---|---|
| **EPC2-hTERT** | 1500 |
| **FLO-1** | 1500 |
| **MFD-1** | 1500 |
| **OAC-P4C** | 1500 |
| **OE33** | 1500 |
| **SK-GT-4** | 1000 |

### 2.1.2 Compound management

Compounds were stored at -20 °C in 100 % DMSO at stock concentrations. These were transferred to compound source plates, v-bottom 384-well (Corning, #3363), at 1,000-fold assay concentration. DMSO and paclitaxel (300 nM) were included on plates as controls. All compounds and liquid handling was carried out using a Biomek FX liquid handler (Beckman)[i].

#### Mechanism of action reference library

Compounds were screening in dose responses. Compound source plates were made from compound stocks (Table 2.2) by serial dilution from 10 to 0.01 mM (1,000-fold assay concentrations) in semi-log concentrations.

**Table 2.2: Mechanism of action reference compounds.** Compound name, mechanism of action, sub-class and supplier and catalogue number.

| Compound Name | Mechanism of Action | Sub-Class | Supplier; Catalogue number |
|---|---|---|---|
| Cytochalasin B | Actin disrupting | Actin disruptor | Sigma; C8273 |
| Cytochalasin D | Actin disrupting | Actin disruptor | Sigma; C6762 |
| Latrunculin | Actin disrupting | Actin stabiliser | Sigma; L5288 |
| Camptothecin | DNA damaging | Topoisomerase-1 inhibitor | Selleckchem; S1288 |
| SN38 | DNA damaging | Topoisomerase-1 inhibitor | Selleckchem; S4908 |

[i] With thanks to Ashraff Makda, for programming the robot for all liquid compound handing

| | | | |
|---|---|---|---|
| Dasatinib | Kinase inhibitor | Src- EMT | Selleckchem; S1021 |
| Saracatinib | Kinase inhibitor | Src-EMT | Selleckchem; S1006 |
| Epothilone B | Microtubule disrupting | Microtubule stabiliser | Selleckchem; S1364 |
| Paclitaxel | Microtubule disrupting | Microtubule stabiliser | Sigma; T7402 |
| Colchicine | Microtubule disrupting | Microtubule destabiliser | Sigma; C9754 |
| Nocodazole | Microtubule disrupting | Microtubule destabiliser | Sigma; M1404 |
| Monastrol | Microtubule disrupting | Eg5 kinesin inhibitor | Sigma; M8515 |
| ARQ621 | Microtubule disrupting | Eg5 kinesin inhibitor | Selleckchem; S7355 |
| Barasertib | Microtubule disrupting | Aurora kinase B inhibitor | Selleckchem; S1147 |
| ZM447439 | Microtubule disrupting | Aurora kinase B inhibitor | Selleckchem; 1103 |
| MG132 | Protein degradation | Proteasome | Selleckchem; S2619 |
| Lactacystin | Protein degradation | Proteasome | Tocris; 2267 |
| ALLN | Protein degradation | Cysteine/calpain | Sigma; A6165 |
| ALLM | Protein degradation | Cysteine/calpain | Sigma; A6060 |
| Cycloheximide | Protein synthesis | Protein synthesis | Sigma; 01810 |
| Emetine | Protein synthesis | Protein synthesis | Sigma; E2375 |
| Lovastatin | Statin | Statin | Sigma; PHR1285 |
| Simvastatin | Statin | Statin | Sigma; PHR1438 |
| SAHA | HDAC inhibitor | HDAC inhibitor | Sigma; SML0061 |
| Panobinostat | HDAC inhibitor | HDAC inhibitor | Selleckchem; S1030 |
| Trichostatin A | HDAC inhibitor | HDAC inhibitor | Selleckchem; S1045 |
| Romidepsin | HDAC inhibitor | HDAC inhibitor | Selleckchem; S3020 |
| Entinostat | HDAC inhibitor | HDAC inhibitor | Selleckchem; S1053 |
| Quisinostat | HDAC inhibitor | HDAC inhibitor | Selleckchem; S1096 |
| Ricolinostat | HDAC inhibitor | HDAC inhibitor | Selleckchem; S8001 |
| Tubastatin A | HDAC inhibitor | HDAC inhibitor | Selleckchem; S8049 |
| Droxinostat | HDAC inhibitor | HDAC inhibitor | Selleckchem; S1422 |
| PCI34051 | HDAC inhibitor | HDAC inhibitor | Selleckchem; S2021 |
| TMP195 | HDAC inhibitor | HDAC inhibitor | Selleckchem; S8502 |
| LMK235 | HDAC inhibitor | HDAC inhibitor | Selleckchem; S7569 |
| CUDC-907 | HDAC inhibitor | HDAC inhibitor | Selleckchem; S2759 |
| Belinostat | HDAC inhibitor | HDAC inhibitor | Selleckchem; S1085 |

### Primary screen (Prestwick, LOPAC, BioAscent, and bespoke libraries)

Primary screen source plates came printed in DMSO at 1,000-fold assay concentration (Table 2.3).

**Table 2.3: Library concentrations.** CRUK therapeutics discovery laboratories Library (CRT).

| Library | Number of Plates | Source Plate Concentration (µM) | Assay Concentration (µM) |
|---|---|---|---|
| Prestwick | 4 | 1,000 | 1 |
| LOPAC | 5 | 3,000 | 3 |
| Bespoke | 1 | 1,000 | 1 |
| BioAscent | 10 | 10,000 | 10 |
| CRT | 44 | 10,000- 12,000 | 10-12 |

### Compound dosing

Compound preparation and dosing was carried out 24 hrs after cell seeding. All source plates were diluted in media by printing the compounds (1.5 µL) into intermediate plates (Greiner, #781128) containing cell culture media (74.5 µL), giving a 1:50 dilution. The intermediate plates were then printed (2.5 µL) into the assay plates containing cells and media (50 µL) giving a 1:20 dilution. This gave an overall dilution of 1:1000 from source to assay plate.

### Cell Painting

The Cell Painting protocol (Gustafsdottir *et al.*, 2013; Bray *et al.*, 2016) was adapted for ease of use in a high throughput setting and optimized for our cell lines. Specifically, cells were fixed first and all of the Cell Painting reagents were added after to help with timings when running large batches of plates at a time. Dye concentrations were also optimised specifically for our cell panel. The protocol was run as follows:

After 48 hrs incubation in the presence of the compounds, cells were fixed by the addition of an equal volume of paraformaldehyde (4 %, 50 µL) (#28908, Thermo) to the existing media, incubated at room temperature (20 minutes) and washed twice in PBS. Cells were then permeabilised in Triton-X100 (0.1%, 50 uL) and incubated at room temperature (20 minutes) followed by two more washes with PBS.

A 48 hr time point was chosen as this has been shown previously in our group to be optimal for morphological screening studies. Morphological changes occur temporally with compound incubation but are likely slower than transcriptomic changes and therefore 24 hrs may not capture end state morphological changes induced. Incubation periods longer than 48 hrs cause suboptimal cell densities for cell segmentation during image analysis and mean that a start to finish assay run would be longer than a week, leading to further screening difficulties. While temporal studies of morphological changes are of great interest it is not feasible at the primary screening stage and therefore a single time point of 48 hrs was taken forward.

The staining solution (Table 2.4) was prepared in bovine serum albumin solution (1 %). Staining solution was added to each well (25 µL) and incubated in the dark at room temperature (30 minutes), followed by three washes with PBS and no final aspiration. Plates were sealed with a foil seal.

**Table 2.4: Cell Painting reagents.** Concentrations, excitation/emission wavelengths of the filters used for imaging, and suppliers. ex: excitation, em: emission

| Stain | Structure | Wavelength (ex/em [nm]) | Channel | Concentration | Cat No; Supplier |
|---|---|---|---|---|---|
| Hoescht 33342 | Nuclei | 387/447 | DAPI | 4 µg/mL | #H1399; Mol. Probes |
| SYTO 14 | Nucleoli | 531/593 | CY3 | 3 µM | #S7576; Invitrogen |
| Phalloidin 594 | F-actin | 562/624 | TxRED | 0.14X | #ab176757; Abcam |
| Wheat germ agglutinin Alexa Fluor 594 | Golgi and Plasma Membrane | 562/624 | TxRED | 1 µg/mL | #W11262; Invitrogen |
| Concanavalin A Alexa Fluor 488 | Endoplas-mic Reticulum | 462/520 | FITC | 20 µg/mL | #C11252; Invitrogen |
| MitoTracker DeepRed | Mitochond-ria | 628/692 | CY5 | 600 nM | #M22426; Invitrogen |

### 2.1.3 Image analysis

#### Image acquisition

Plates were imaged on an ImageXpress micro XLS (Molecular Devices, USA) equipped with a robotic plate loader (Scara4, PAA, UK). Four fields of view were captured per well using a 20x objective and five filters (Table 2.4).

#### CellProfiler 2D image analysis

CellProfiler v3.0.0 (Carpenter *et al.*, 2006) image analysis software was used to extract 733 features per cell per image. First the pipeline identified the nuclei from the DAPI channel and used these as seeds to aid a segmentation algorithm to identify the cell boundaries, and finally these two masks were subtracted to give the cytoplasm. These three masks marking the cellular boundaries were then used

to measure morphological features including size, shape, texture, and intensity per object across all five image channels.

### Image pre-processing

The cell level data was aggregated to image level by taking the median for each measured feature per image. Low quality images and image artefacts were then identified and removed using image quality metrics extracted by CellProfiler. Images with 20 cells or less were also removed from the dataset. For the remaining images, features were normalised on a plate-by-plate basis by dividing each feature by the median DMSO response for that feature. Features with NA values were removed, as were features with zero or near zero variance, using the 'findCorrelation' and 'nearZero' functions in the R package Caret. All remaining features were scaled and centred globally by dividing by the standard deviation of each feature and subtracting the feature mean respectively. The pair-wise correlations were calculated for all remaining features, and highly correlated features (>0.95) were removed. Finally, the image data was aggregated to the well (compound) level and this was used in the analysis.

## 2.1.4 Data analysis

All analysis was conducted in R http://www.r-project.org using software packages available via CRAN http://cran.r-project.org and Bioconductor http://www.bioconductor.org.

### Dimension reduction

High dimensional datasets, such as those seen in high content screening or in transcriptomic analyses, present a challenge for interpretation and visualisation. Dimension reduction techniques, such as principal component analysis (PCA), allow for the simplification of high dimensional multivariate data to low dimensional space using the combination of variables explaining the greatest degree of variation across the data. This allows for the variation across multiple variables to be visualized in a 2D plot capturing the inherent structure of the data.

In this work principal component analysis was implemented using the built-in R function, prcomp, to visualise high dimensional datasets.

### Hierarchical clustering

Hierarchical clustering reorders variables into clusters based on similarities, to reveal shared patterns in the data. Each observation starts in its own cluster and clusters are merged iteratively by pairwise similarity until all observations belong to one single cluster. A cut-off can then be defined to determine the number of clusters in the data.

In this thesis, the functions 'hclust' and 'dist' in the R stats package were implemented. Euclidean distance was used as a measure of similarity and linkage between clusters was calculated using Ward's criterion, which uses the increase in variance for the cluster being merged (Ward, 1963). Clustering can be represented as a dendrogram, a diverging tree diagram that places variables from the same cluster adjacently. Similarity between individual variables or aggregated clusters are encoded by dendrogram height, with more similar variables being connected by smaller connections.

### Random forest

Random forests is a supervised machine learning algorithm based on decision trees. A decision tree predicts an items class or group based on the observations or features of the item. In this tree structure, the leaves represent the class labels and the branches represent the features that lead to those class labels (Figure 2.1). Much like a forest is comprised of trees, a random forest is comprised of decision trees. Random forests create decision trees based on randomly selected data samples, takes predictions from each tree and selects the most likely outcome or class by means of voting (Figure 2.1).

See specific chapters for details relating to the implementation of random forests in this work.

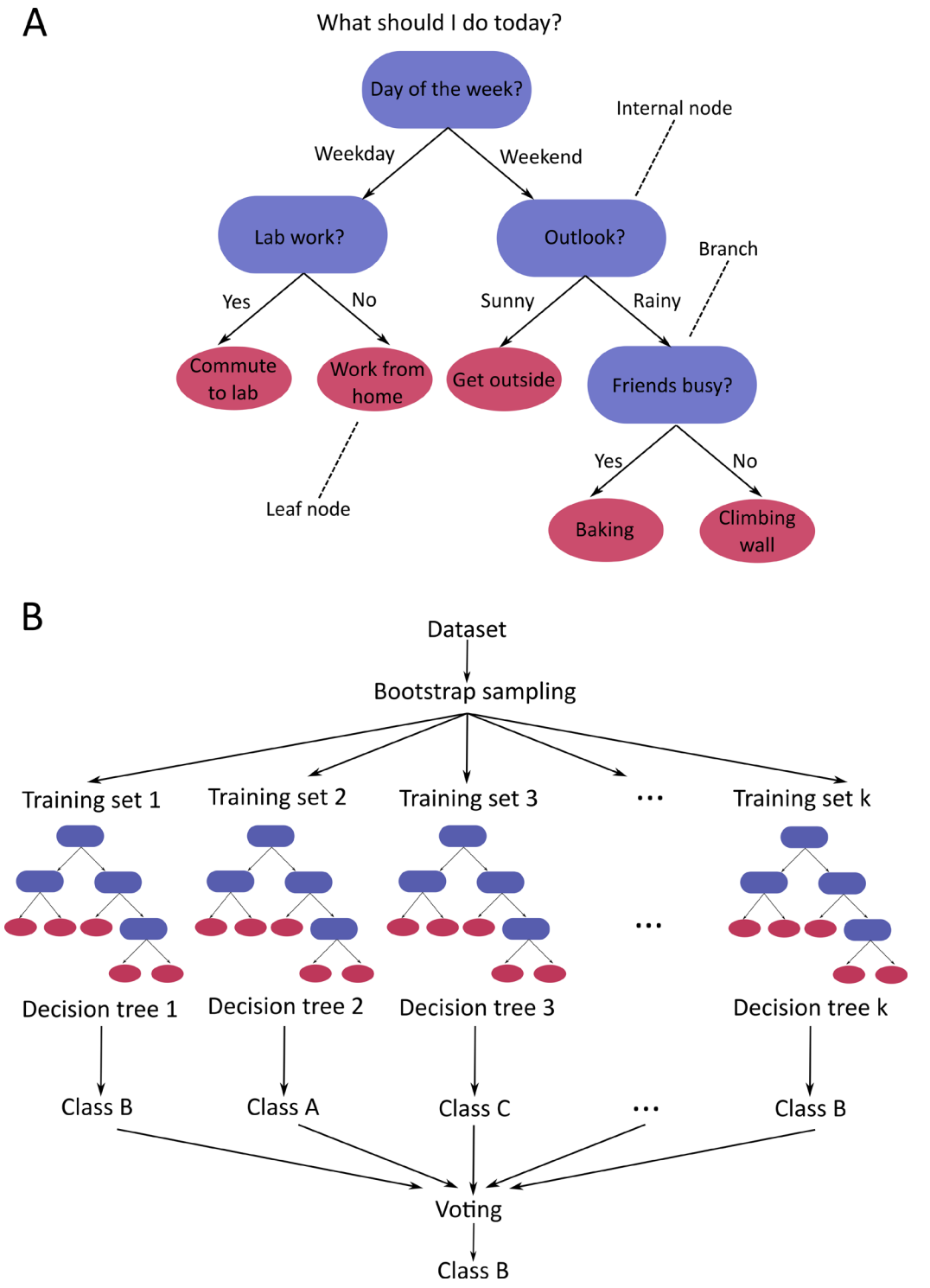



**Figure 2.1: Random forests.** A) Decision tree example. B) Random Forests example.

## NanoString transcriptomic analysis

The NanoString nCounter platform provides a simple and cost-effective solution for the multiplexed gene expression analysis. Comparative studies have demonstrated that the nCounter gene expression platform is more sensitive than microarrays and similar in sensitivity and accuracy to real-time PCR (Geiss *et al.*, 2008). However, it's multiplexed capabilities and scalability, requiring just nanograms of RNA, make it ideally suited to studing many samples, conditions or treatments at the same time.

All NanoString nCounter analyses were carried out on the Human PanCancer Pathways (Catalogue number XT-CSO-PATH1-1) and Metabolic Pathways (Catalogue number XT-CSO-HMP-12) panels, covering 1449 genes.

Cells were seeded at 8 x $10^4$ cells in 6-well plates and incubated under standard tissue culture conditions for 24 hrs. For basal analysis media was then removed and plates were washed twice with ice cold PBS before being snap frozen at -80. For treatment induced analysis, media was replaced with fresh media containing compound treatments (DMSO (0.1%), disulfiram (600 nM), or elesclomol (200 nM)) before further incubation for 6 hrs, and then samples were processed the same as for basal samples.

For RNA extraction cells were scraped and lysed using QIAshredders (#79654, QIAGEN), RNA was extracted by means of the Qiagen RNeasy Mini kit (#74104, QIAGEN) (with β-mercaptoethanol) according to manufacturer instructions, and included a DNase digestion step (#79254, QIAGEN).

Of the purified RNA, 100 ng were used as input for amplification-free RNA quantification by the NanoString nCounter Analysis System with the Human PanCancer Pathways and Metabolic Pathways panels [ii] . Raw counts were normalized to the internal positive controls and housekeeping genes, using the nSolver 4.0 software.

[ii] With thanks to Alison Munro, HTPU Services Facility Manager for running the samples.

### Differential gene expression analysis

Differential gene expression analysis allows class-specific differences in gene expression to be quantified, generating a measure of statistical significance in conjunction with a magnitude of fold change. Within this thesis, nSolver 4.0 software and the NanoStringDiff algorithm (Wang *et al.*, 2016) were used to determine differentially expressed genes between groups. See chapter specific methods for group details. P-values were adjusted using the Benjamini-Yekutieli approach (Benjamini and Yekutieli, 2001).

### Pathway analysis

Analysing gene expression patterns at the pathway level allows changes in individual genes to be summarised as a collective effect. This allows the functional significance of multiple genes contributing to a given biological pathway to be combined, it simplifies long gene lists to more manageable collections of pathways, and brings biological meaning to analyses.

Ingenuity pathway analysis (IPA) implements the causal analytics tools 'Upstream Regulator Analysis', 'Mechanistic Networks', 'Causal Network Analysis' and 'Downstream Effects Analysis' to infer and score regulator networks upstream of gene expression data, and predict downstream effects on biological functions and diseases (Krämer *et al.*, 2014).

IPA analyses were implemented using the NanoString Differential Expression data and 0.3 log expression ratio cut-offs for individual genes. The background reference gene set was defined as the total gene list covered by the NanoString panels to prevent biases. All other settings were kept standard.

## 2.1.5 Published datasets

### DepMap

The area under the dose response curve (AUC) dataset is available at https://depmap.org/portal/download/. Dataset version is the PRISM Repurposing

Secondary Screen 19Q4, containing 373 cell lines screened against both disulfiram and elesclomol.

#### GDSC

The genomics of drug sensitivity in cancer data is available from https://www.cancerrxgene.org/downloads/bulk_download.

GDSC1000 downloads:

The Robust Multichip Average (RMA) processed dataset is available at http://www.cancerrxgene.org/gdsc1000//Data/preprocessed/Cell_line_RMA_proc_basalExp.txt.zip.

The AUC dataset is available at https://www.cancerrxgene.org/gdsc1000/GDSC1000_WebResources//Drug_screening_data.html

#### Connectivity Map scores

The Connectivity Map can be accessed through a portal at https://clue.io/.

The connectivity score (tau) is a standardized measure ranging from -100 to 100. For each perturbagen in the list of query results, the score corresponds to the fraction of reference gene sets with a greater similarity to the perturbagen than the current query. A score of 95 indicates that only 5% of reference gene sets showed stronger connectivity than the current query to the perturbagen in question. In general, a tau of +90 or higher, and of -90 or lower, is considered as strong and could be considered as hypotheses for further study.

### 2.1.6 Seahorse

The Agilent Seahorse XF Cell Mito Stress Test measures key parameters of mitochondrial function by directly measuring the oxygen consumption rate (OCR) of cells on the Seahorse XFe Extracellular Flux Analyzer. It is a plate-based live cell assay that allows us to monitor OCR in real time. Compounds are added sequentially to measure different aspects of mitochondrial respiration including; basal respiration, ATP-linked respiration, maximal respiratory capacity, spare respiratory capacity, proton leak and non-mitochondrial respiration (Figure 2.2).

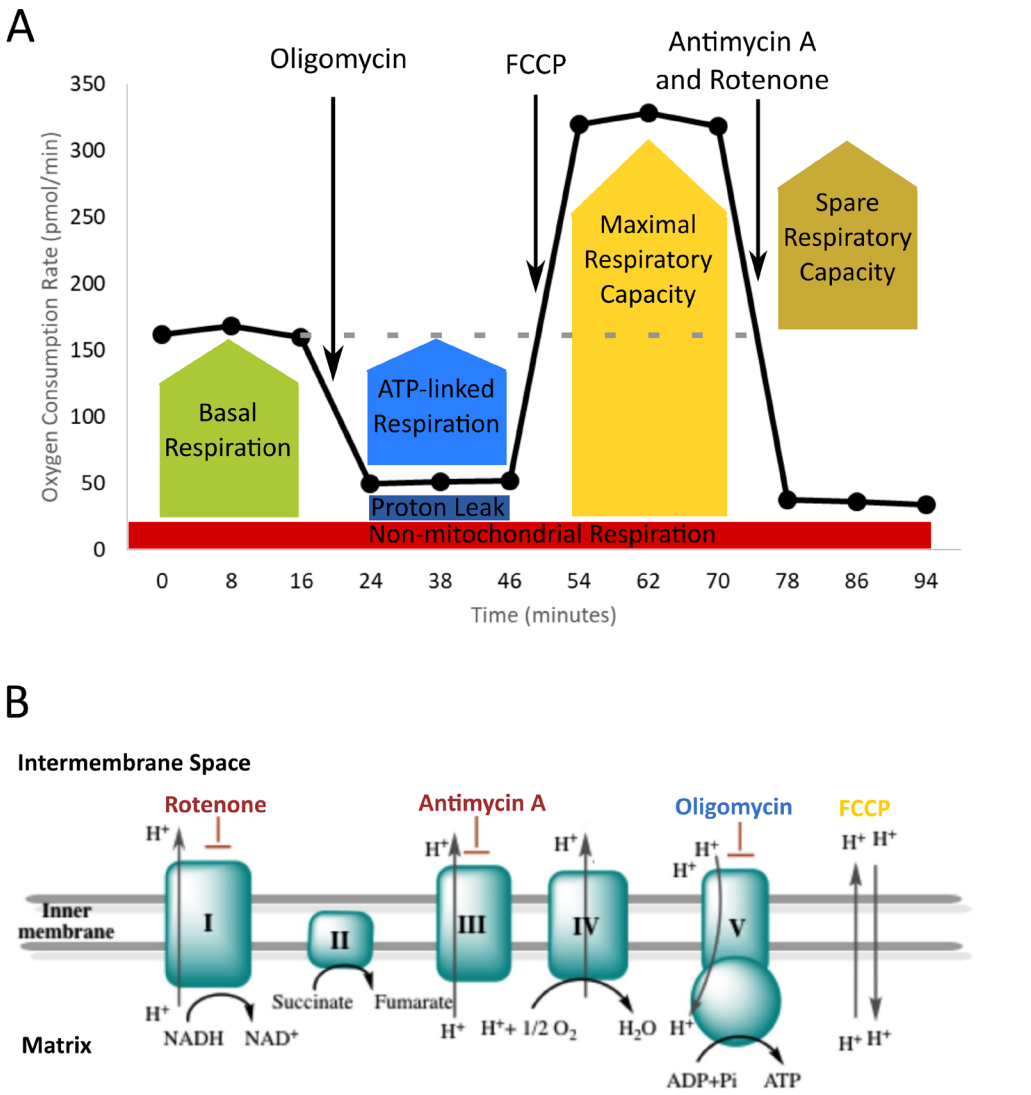



**Figure 2.2: Seahorse assay.** A) Oxygen consumption trace for Agilent Cell Mito Stress Test, compounds added and outcomes measured. B) The electron transport chain complexes and where the seahorse compounds act in the chain. Energy obtained through the transfer of electrons down the transport chain is used to pump hydrogen ions (proton) across the mitochondrial inner membrane creating an electrochemical proton gradient. This allows ATP synthase (complex V) to use the flow of hydrogen ions through the enzyme back to the matrix to generate ATP. FCCP is a mitochondrial uncoupler and allows the flow of hydrogen ions independent of ATP synthase, reducing the gradient without the production of ATP.

Cells were seeded (Table 2.5) in Seahorse 24-well plates (100 µL per well) and left for 30 minutes at room temperature before incubation under standard tissue culture conditions. After cells had adhered (5 hrs) a further 150 µL media was added to each well, and plates were incubated overnight.

**Table 2.5: Seahorse assay conditions.** FCCP- Trifluoromethoxy carbonylcyanide phenylhydrazone

| Cell Line | Seeding Density (per well) | Assay conditions (µM) | | |
|---|---|---|---|---|
| | | Oligomycin | FCCP | Antimycin/ Rotenone |
| OAC-P4C | 10,000 | 2 | 0.6 | 3 |
| SK-GT-4 | 8,000 | 1 | 1.2 | 2 |
| FLO-1 | 8,000 | 1 | 1.2 | 3 |
| JH-EsoAd1 | 10,000 | 1 | 2.4 | 1 |

After 24 hrs incubation, assay media was made from Seahorse XF RPMI assay media kit (Agilent; #103681-100), with glucose (10 mM), pyruvate (1 mM) and glutamine (2 mM). For treatment induced OCR experiments drugs (DMSO (0.1%), or Elesclomol treatments (50 and 200 nM)) were then added and further incubated for 6 hrs. Then all but 40 µL media was aspirated from cell plates and washed with assay media (500 µL) twice before adding assay media to a final volume of 525 µL. For basal experiments, the media removal and replacement with assay media was carried out immediately after the 24 hr incubation. All plates were then incubated at 37 ºC in $CO_2$ free incubator for 30 minutes before being loaded into Seahorse FX for analysis.

Three basal OCR measurements were taken in 8-min intervals, followed by sequential injections of oligomycin, FCCP, and antimycin A and rotenone, taking three measurements at 8-min intervals following each treatment. Basal respiration was determined as the difference between basal OCR and OCR after antimycin A treatment, and maximal respiration as the difference between OCR after FCCP and antimycin A treatments (Figure 2.2).

For normalisation hoescht staining was used to get a cell count. Plates were fixed in formaldehyde (4 % final concentration) and incubated at room temperature for 20 minutes. Plates were then washed twice in PBS. Staining solution; Hoescht 33342 (#H1399; Mol. Probes) 2 µg/ml, Triton X-100 (#X100-500ML; Sigma) 0.1% in PBS, was added (100 µL per well), incubated for 40 minutes, and washed out with PBS before imaging on ImageXpress micro XLS (Molecular Devices, USA).

Cell count was summed over 4 images per well and multiplied by well coverage factor to get total cells per well. OCR rates were divided by the cell count and multiplied by 10,000 to get rates per 10,000 cells[iii].

See specific chapter methods for cell lines and treatments.

### 2.1.7 ICP-MS

Inductively coupled plasma mass spectrometry (ICP-MS) is a type of mass spectrometry that uses inductively coupled plasma to ionize the sample. It is known and used for its ability to detect metals in liquid samples at very low concentrations, and can detect different isotopes of the same element. Compared to atomic absorption spectroscopy, ICP-MS has greater speed, precision, and sensitivity (Wilschefski and Baxter, 2019).

5 x $10^6$ cells were seeded in a T175 flask and incubated in 25ml RPMI overnight. Media was then removed and replaced with compound treatments (DMSO (0.1%), disulfiram (600 nM), or elesclomol (200 nM)) before further incubation for 6 hrs before cells were collected. Media was removed and cells were washed in PBS, typsinised and counted. For each sample 2 x $10^6$ cells were pelleted in an Eppendorf, the supernatant was discarded and samples were frozen at -80 °C.

For analysis, samples were thawed and concentrated nitric acid was added (100 µL per well) and mixed. Samples were then vortexed and sonicated and left overnight at room temperature. Samples were made up to 1 mL using Di water

[iii] With thanks to Dr. Rod Cater, Molecular Metabolism Group, for guidance on the assay and lengthy discussions regarding optimisation and interpretation of results.

and then further diluted fivefold and analysed in the ICP Facility, School of Chemistry, University of Edinburgh[iv].

## 2.2 Chapter 3: Morphology Predicts Mechanism

### 2.2.1 Unsupervised machine learning: Clustering

All OAC cell lines were assessed for high content Cell Painting suitability and cell lines that passed were taken forward for mechanism of action prediction using a library of 37 reference compounds (Table 2.2).

#### Dimension reduction

Principal component analysis was implemented on the globally normalised median per image feature data and multiple concentrations of each compound were included (0.1, 3, and 10 μM).

#### Hierarchical clustering

Data was subset for globally normalised median per image feature data for the FLO-1 cell line treated with the library of reference compounds at 10 μM. A correlation distance matrix was calculated between the rows (treatments) of the dataset, and used to cluster the data. Clusters were set to the number of classes in the library - eight. DMSO treatments were also included for reference but were not included in the initial clustering.

Clustering of the subtle phenotypes was carried out as above except the library was subset for Protein degradation, Protein synthesis, and Statin classes and clusters were fixed to three.

### 2.2.2 Supervised machine learning: Classification

#### Random forest classifier

The random forest classifier was implemented using R's Random Forest package with default parameters except for the following specified parameters; data was

[iv] With thanks to Dr. Lorna Eades, ICP Facility Manager for digesting and running the samples.

stratified by class and the sample size was set to the smallest class size for balance. Two different analyses were run, first, mechanism of action prediction was run for each cell line individually, here three concentrations (0.1, 3, 10 μM) of each compound were pooled and treated as a single class. Second, the classifier was trained on five of the OAC cell lines and tested on the remaining one, here only the top two concentrations of compound treatments (3 and 10 μM) were pooled and treated as one class.

To measure model quality F1 scores were calculated and expressed as the mean F1 score (macro-F1 score) per cell line as a percentage. An F1-score is the weighted average of precision and recall and therefore takes both into account when assessing a model.

Precision is the fraction of correct predictions for a given class and is defined as:

$$Precision = \frac{true\ positives}{true\ positives + false\ positives}$$

$$= \frac{true\ positives}{total\ predicted\ positives}$$

Recall is the fraction of instances of a class that were correctly predicted and is defined as:

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

$$= \frac{true\ positives}{total\ actual\ positives}$$

F1 score is calculated as:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

F1 scores were calculated for each class within each classifier using the 'correlationMatrix' function in the R package Caret. The macro-F1 score for each classifier was then calculated as the mean F1 score across all classes.

# 2.3 Chapter 4: High Content Profiling in Oesophageal Adenocarcinoma.

Cell lines were maintained in 3 T175 flask each for screening purposes.

Five compound library source plates (Table 2.3) were prepared a week and each intermediate plate was printed into eight assay plates (one per cell line), making a total of 40 plates a week.

## 2.3.1 Hit identification

### Multiparametric analysis

Principal component analysis was calculated on the normalized, well median, CellProfiler feature data in R. Principal components were limited to the first 15 principal components, which accounted for approximately 90% of the variation across each cell line. The Mahalanobis distance threshold was then applied as a metric of compound activity. The Mahalanobis distance measures the distance of each point from the data distribution (in this case the DMSO controls). The data distribution takes into account the mean and the spread of the data points using the covariance matrix as a normalization factor (Mahalanobis, 1936). It therefore addresses problems of both scale and correlation of the variables and is particularly beneficial for large multivariate datasets. This leads to elliptic decision boundaries rather than circular as is the case for Euclidean.

Hits were defined as having a Mahalanobis distance of 1500 or greater in the OAC cell lines and a difference of at least 1000 with one or both of the control lines. This comparison was made between each OAC and tissue-matched cell line and taken forward if it was identified as a hit in two or more OAC cell lines.

### Univariate Analysis

Cell numbers were obtained from the nuclei channel object counts from CellProfiler and these were converted to z-scores using the DMSO controls on a plate by plate basis.

$$Z = \frac{x - \mu}{\sigma}$$

Where $x$ is the observed value, μ is the mean of the DMSO controls and σ is the standard deviation of the DMSO.

Hits were defined as having a z-score of – 3 or more in the OAC lines and a difference of at least 2 in one or both of the tissue-matched control cell lines e.g. for a hit with a z-score of -3 in an OAC line, the z-score in the tissue-matched control cell lines would have to be greater than or equal to -1. This comparison was made between each OAC line and the tissue-matched control cell lines to define hits and then selected if they were active is at least two OAC lines across the panel

## 2.3.2 Hierarchical clustering

Z-scores and Mahalanobis scores were centred and scaled for each compound across the panel of cell lines. Hierarchical clustering was then carried out as stated in general methods above.

## 2.3.3 Validation

### Validation screen

Stock compounds were supplied at 10 mM and were diluted in 100 % DMSO to create eight point dose responses from 10-0.003 mM in semi-log concentrations in source plates. Each plate was screened in triplicate.

### Univariate dose responses

Cell numbers from the nuclei counts were normalized to the DMSO controls on a plate by plate basis and curves were fitted in GraphPad Prism 6. Active compounds were defined as those displaying activity at 3 μM or less. Selective compounds had a window of selectivity for at least two OAC cell lines compared to the controls as

defined by eye from the curves. Assessment by $IC_{50}$ was not possible due to the nature of curve fitting over multiple cell lines and compounds, e.g. some compounds were active at all validation concentrations. $IC_{50}$ values were determined at a later stage through further adjustment of concentrations and repeat dose responses.

#### Multiparametric dose responses

Activity was assessed in the same manner as the initial screen, by PCA and Mahalanobis distance, across the dose response range. Validated hit dose responses were then visually confirmed by eye.

### 2.3.4 Machine learning

Using the individual classifiers trained in chapter 1, the feature data for the antimetabolite compounds; Methotrexate, Pemetrexed and Raltitrexed, was fed into the classifiers and classes were predicted for each concentration in each cell line. DNA damaging agent class probabilities were then extracted across the extended dose responses from 10- 0.0001 μM.

## 2.4 Chapter 5: Ammonium pyrrolidinedithiocarbamate, Disulfiram, and Elesclomol: Towards a unified theory of drug mechanism

### 2.4.1 Patient derived organoids

Patient derived organoids were screened by Xiaodun Li (University of Cambridge), according to the published protocol (Li *et al.*, 2018). Dose response curves were fitted in GraphPad Prism 6.

### 2.4.2 Differential gene expression analysis

Differential gene expression analysis was carried out between the control (DMSO) and treatment (200 nM elesclomol) samples for the two sensitive cell lines (OAC-P4C and SK-GT-4) pooled together. N=3 per cell line. Difference in cell lines was taken into account as a confounder in the analysis.

### 2.4.3 Copper chelator bathocuproinedisulfonic acid (BCS)

The dose responses were carried out as stated for the validation dose responses except disulfiram and elesclomol dose responses were carried out as 10 point doubling dilutions from 1 μM in the presence or absence of 200 μM BCS pre-incubation for 15 minutes.

### 2.4.4 N-acetyl-L-cysteine dose response

Disulfiram and elesclomol dose responses were carried out as before except in the presence and absence of a 15 min pre-incubation with 10 mM N-acetyl-L-cysteine (NAC) (A7250; Sigma) or 1/5 mM Trolox (238813; Sigma). NAC was made up in media and 0.1M NaOH was used to neutralise the media and bring the pH back to 7.4. This was then syringe filtered before being added to the cells

### 2.4.5 $H_2O_2$ dose response

Work carried out by Dr. Richard Elliott. Cells were plated in 384-well plates as specified in section 2.1.1 Cell culture. After 24 hrs incubation, $H_2O_2$ (H1009; Sigma) was diluted in media to give an 18 point dose response starting at 1 mM and incubated for 48 hrs. Viability was assessed by Alamar Blue.

## 2.5 Chapter 6: Pharmacogenomics of Disulfiram and Elesclomol

### 2.5.1 Differential gene expression analysis

Treatment induced differential gene expression:

Analysis was carried out between the control (DMSO) and treatment (200 nM elesclomol or 600 nM disulfiram) samples for the two tissue-matched control cell lines (CP-A and EPC2-hTERT) pooled together. N=1 per cell line. Difference in cell lines was taken into account as a confounder in the analysis.

Basal gene expression:

OAC versus tissue-matched control differential gene expression analysis was carried out between the tissue-matched control cell lines (CP-A and EPC2-hTERT) and the OAC cell lines (OAC-P4C, SK-GT-4, JH-EsoAd1 and FLO-1). The tissue-matched control cell lines were used as the base line. N= 1

Sensitive versus resistant OAC differential expression analysis was carried out between the sensitive cell lines (OAC-P4C, SK-GT-4, OE19, and ESO26) and the resistant cell lines (FLO-1, JH-EsoAd1, and ESO51). N=3.

### 2.5.2 Calculation of OAC $IC_{50}$ and NanoString gene expression analysis

The Pearson correlation between $IC_{50}$ and basal gene expression was calculated for each gene in the NanoString panels (1449 genes) across nine OAC cell lines (OAC-P4C, SK-GT-4, ESO26, OE19, OE33, ESO51, MFD-1, JH-EsoAd1, FLO-1). N= 3.

### 2.5.3 Calculation of pan cancer AUC and gene expression correlation

The Pearson correlation between the GDSC AUC's (916 cell lines) and each genes expression profile was calculated (RMA normalised expression values for 17,737 genes). P values were computed and corrected using the Benjamini–Hochberg correction. See 2.1 General Methods for data downloads.

### 2.5.4 Gene set enrichment analysis

Enrichment analysis was conducted using GSEA (Subramanian *et al.*, 2005), using the H and C2 gene sets from the molecular signatures database (MSigDB; see below for details). Default settings were used except Pearson was used as the metric for gene ranking. H-Hallmark gene set collection includes 50 gene sets (Liberzon *et al.*, 2015). C2-Canonical pathways KEGG collection includes 186 gene sets (Kanehisa and Goto, 2000)

# *3| Morphology Predicts Mechanism*

Cellular morphological changes can be used to predict drug mechanism of action across heterogeneous panels of cell lines.

## 3.1 Introduction

Cellular function and cellular morphology are highly interconnected. Manipulation of cell function and gene expression via perturbations; chemical, genetic, or otherwise, impacts on a cells morphological phenotype (Kamath and Ahringer, 2003; Kiger *et al.*, 2003; Perlman *et al.*, 2004). Upon perturbation, cells undergo detectable changes in a subset of profiled features, allowing quantitative data to be extracted from microscopy images and a phenotypic fingerprint to be assigned to a perturbation (Perlman *et al.*, 2004; Young *et al.*, 2008; Caie *et al.*, 2010; Ljosa *et al.*, 2013). High content profiling combines systematic perturbations with high content imaging and image informatics allowing the quantification of induced morphological changes and insights into biological processes (Giuliano *et al.*, 1997; Kamath and Ahringer, 2003; Perlman *et al.*, 2004; Bickle, 2010; Rose *et al.*, 2018).

Since the morphological changes associated with a compound treatment are a consequence of its effects on particular cellular processes, we might expect that compounds acting on a given target or process elicit more similar morphological responses than those acting on different processes. Perlman *et al.* first demonstrated this in 2004, when they showed that compound induced

morphologies could be grouped by mechanism of action using similarity measures and hierarchical clustering (Perlman *et al.*, 2004). Since then, numerous research groups have implemented unsupervised clustering methods to group treatments based on phenotypic response (Carpenter *et al.*, 2006; Young *et al.*, 2008; Gustafsdottir *et al.*, 2013; Ljosa *et al.*, 2013). This has further evolved to allow mechanism of action inference through similarity comparison to reference compound sets using multivariate statistics and a variety of supervised machine learning techniques (Warchal, Dawson and Carragher, 2019; Hughes *et al.*, 2020).

This chapter presents a comprehensive study of compound induced cell morphology in the context of a heterogeneous panel of eight oesophageal cell lines.

Firstly, I wanted to assess the performance of the Cell Painting assay (Figure 1.5) and OAC cell lines in a high content setting, an important step for future large scale screening studies (Chapter 4). Standard assay quality control metrics such as Z'Factor are unsuitable for multiparametric assays, particularly when there is no specific phenotypic outcome desired or there is a lack of positive controls (Sui and Wu, 2007; Birmingham *et al.*, 2009; Kummel *et al.*, 2010). Instead I propose using multivariate statistics and clustering to benchmark cell lines and assay conditions on their ability to discriminate distinct mechanistic classes of compounds and accurately predict their mechanism of action as assay quality control metrics. To achieve this I profiled a small number of reference compounds and built a library of phenotypic fingerprints for each mechanistic class and then interrogated it using a variety of supervised and unsupervised machine learning techniques.

Secondly, I sought to determine whether prior learning could be applied to new cell lines. The hypothesis being that compounds induce core morphological changes present across cell lines. However this idea is complicated by underlying cellular heterogeneity, both morphological and genetic, and their differential response to chemical treatments, potentially confounding mechanistic predictions across distinct cell lines. This concept is important because it would allow the classification of a compound treatment for a previously unseen cell line,

presenting an opportunity to develop a highly generalizable classifier, a powerful tool for drug discovery.

## 3.2 Results

### 3.2.1 Building a high content panel of OAC cell lines

Cell panel screening provides drug response data across genetically diverse cell lines and can be used both within and across disease areas. This reduces false positives, provides information on sensitivity and resistance, and allows patient stratification hypotheses and genetic biomarkers to be identified early on in the discovery pipeline. It is also an effective strategy for heterogeneous diseases such as OAC, where one model cell line does not represent all primary tumours and it therefore avoids narrowly relevant mechanisms of action (Vincent *et al.*, 2020). Tissue matched controls can also be built in to cell panels allowing assessment of disease selectivity.

In order to develop a cell panel for high content profiling in OAC that recapitulates the genetic heterogeneity of OAC, as determined by recent next-generation sequencing studies in patient samples (Secrier *et al.*, 2016), I assessed a broad panel of OAC cell lines expressing the genetic hallmarks of the disease (Weaver *et al.*, 2014; Contino *et al.*, 2016); JH-EsoAD1, FLO-1, MFD-1, OE33, OACM5.1, OAC-P4C, SK-GT-4, ESO51, ESO26 and OE19, and two tissue-matched control lines; a Barrett's oesophagus line CP-A, and a normal oesophageal squamous line EPC2-hTERT. I evaluated each cell line against a list of criteria (Table 3.1) that indicate high performance for high content screening, ensuring image quality and information content, high-throughput capabilities of the cell lines, and image segmentation accuracy for downstream analysis pipelines. Importantly, I also confirmed cell line identities using Short Tandem Repeat (STR) profiling for those that passed all other criteria.

**Table 3.1: High content cell panel criteria.**

| Cell Lines | Myco-plasma | Adhere-nce* | Prolifera-tion* | Cell Painting Assay | Image Segmen-tation | MOA Prediction | STR Pro-filing |
|---|---|---|---|---|---|---|---|
| CP-A | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| EPC2-hTERT | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | NA |
| ESO51 | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | - |
| ESO26 | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | - |
| FLO-1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| JH-EsoAD1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| MFD-1 | ✓** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| OACM5.1 | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | - |
| OAC-P4C | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| OE19 | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | - |
| OE33 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SK-GT-4 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*In 384-well plates; ** cell line contaminated at source but treated and subsequently tested negative; ✗ failed quality control metric; 'NA' reference profile unavailable; ' — ' not profiled as deemed unsuitable for primary cell panel. MOA = mechanism of action, STR = short tandem repeat

Non-adherent cells, cells that clump and cluster, and cells with a very rounded morphology all present focus and segmentation problems. This leads to a vastly reduced information content in the images, making them unsuitable for high content imaging. For these reasons, four cell lines (ESO51, ESO26, OACM5.1, and OE19) were removed from the cell panel early on due to adhesion quality or undesirable cellular morphology that could not be readily adapted for high content imaging. ESO51 and OACM5.1 are both suspension cell lines and while

OE19 and ESO26 are adherent, OE19 forms rounded multi-cell balls and ESO26 forms clouds of adherent cells, both leading to very out of focus images and presenting image segmentation issues. I therefore took forward the following eight cell lines for high content assay development including image segmentation and machine learning analysis (CP-A, EPC2-hTERT, FLO-1, JH-EsoAD1, MFD-1, OAC-P4C, OE33, and SK-GT-4) (Figure 3.1).

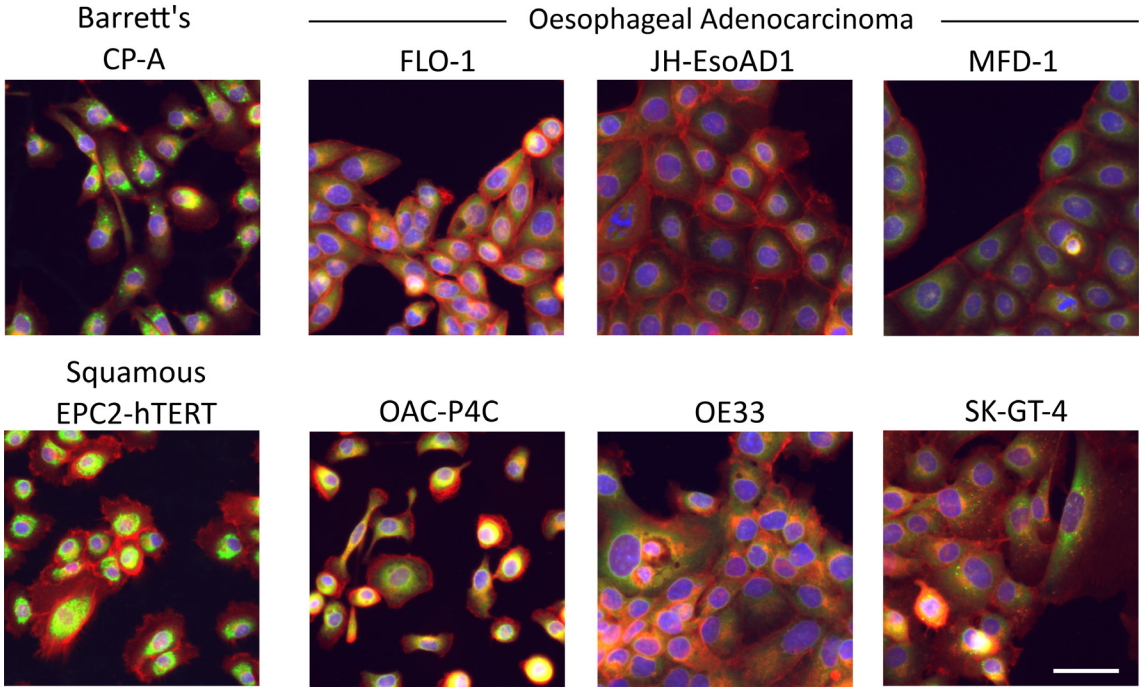



**Figure 3.1: Oesophageal adenocarcinoma cell panel.** A) Colour combined representative control (DMSO) images of the eight cell lines in the cell panel: Hoescht 33342-nuclei (blue); Phalloidin 594, Wheat germ agglutinin and Alexa Fluor 594- F-actin, golgi and plasma membrane (red); Concanavalin A Alexa Fluor 488- Endoplasmic reticulum (green). Scale bar is 50 μm. See Table 2.4 for additional details about the stains and channels imaged.

### 3.2.2 Predicting mechanism of action from phenotypic information

Benchmarking high content assays and cell panels presents a challenge since univariate statistics are inadequate for such high-dimensional datasets. Furthermore, desired phenotypic outcomes and positive controls are often

unknown (Sui and Wu, 2007; Birmingham *et al.*, 2009; Kummel *et al.*, 2010). Therefore I used a small, well-annotated library of reference compounds (Table 2.2) with a known mechanism of action to benchmark the cell lines on their ability to discriminate mechanistic classes of compounds. Clustering and classification performance was used to assess whether the cell lines were amenable to the Cell Painting assay, further validate image segmentation was accurate, and ensure that the phenotypic information extracted was sufficient to allow distinction between mechanisms of action.

### Unsupervised Machine Learning: Clustering

In order to visually investigate patterns in the multivariate dataset, I performed data reduction using principal component analysis (PCA). PCA is a linear transformation that converts a number of correlated variables into a smaller number of uncorrelated variables, or principal components, in such a way as to preserve the variance of the higher-dimensional data and so that most of the variability in the dataset can be accounted for in the first few components. Importantly, component values are not a physical constant, but rather are relative to the spread of the input data.

Firstly, I analysed the normalised image-level data of CellProfiler extracted features for the library of reference compounds at multiple concentrations (0.1 µM, 3 µM, 10 µM) and plotted the first two components coloured by mechanistic class (Figure 3.2). These plots can be thought of as representing the phenotypic space of each cell line treated with the library of reference compounds.

I predicted that for a given compound the magnitude of response would be dependent on the concentration, and that different classes, affecting different cellular processes and therefore different cellular attributes, would exhibit distinct distributions within the phenotypic space. As expected, at the lowest concentration, compounds cluster with DMSO controls. This central cluster of DMSO and inactive compounds represents the basal cellular phenotypic profile of each cell line, when the compounds have no effect on cellular morphology. Higher concentrations led the majority of compounds to have increasing effects on

cellular morphology and therefore move further away from the basal phenotype. PCA shows that the compound classes have distinct density distributions, clustering by class. It is also noteworthy that the magnitude of response for a given concentration varied between compound classes with some, such as the statins, producing a weaker response than others.

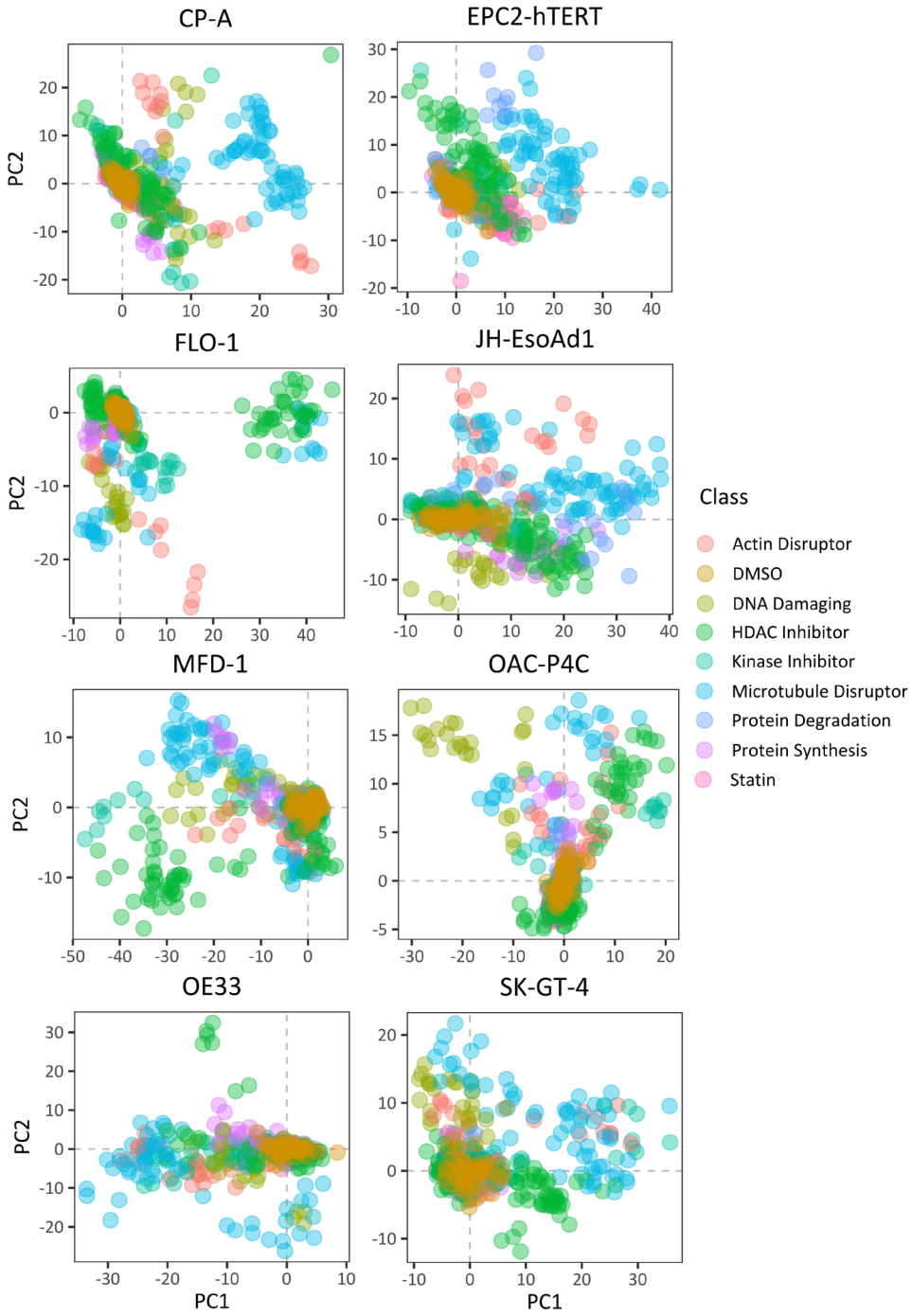



**Figure 3.2: Reference library compounds unsupervised machine learning**. Principal component analysis plots for the reference library of compounds across eight oesophageal cell lines. Points represent a single compound and concentration and are coloured by mechanistic class. Each compound is present at multiple concentrations (0.1, 3 and 10 μM).

Hierarchical clustering shows compounds cluster by mechanistic class (Figure 3.3). When split into eight clusters (the number of compound classes in the reference library) most clusters represent just one class of compounds. Cluster eight is the largest and contains the DMSO treatments. This suggests the compounds in cluster eight do not elicit a strong phenotypic response and therefore reflect the basal cellular phenotype. This may be compound, concentration or mechanism dependent. Overall, the hierarchical clustering agrees with the PCA analysis; morphology is a strong predictor of compound mechanism of action.

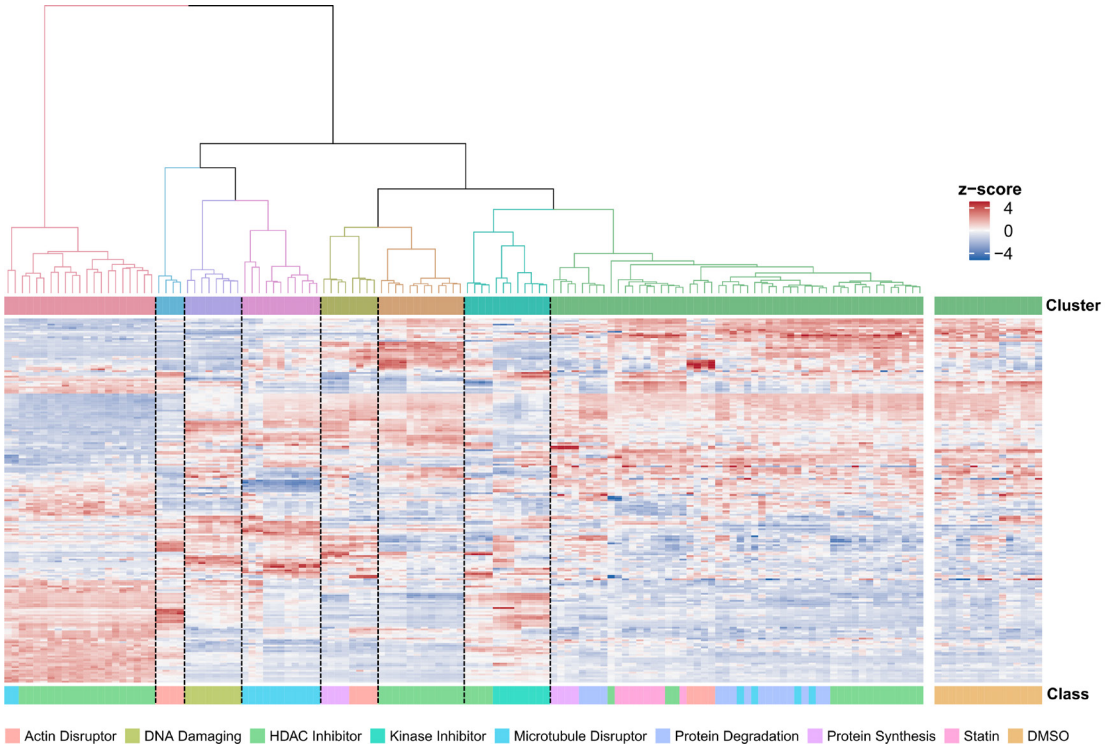



**Figure 3.3: Hierarchical clustering heatmap for reference library compounds.** Compound treatments at 10 µM (columns) clustered by phenotypic features (rows) for FLO-1 cell line. DMSO treatments shown for reference (right hand heatmap). Compounds split into eight clusters shown by dendrogram colouring and cluster annotation at the top, and mechanistic class is shown along the bottom for each compound. Dendrogram built using Euclidean distance and Ward's method.

Given that component values for PCA are not a physical constant, but rather are relative to the spread of the input data, I took a subset of the data containing the compounds pertaining to weaker phenotypic classes; Protein Degradation, Protein Synthesis, and Statins, and re-analysed the data to see if subclusters could be identified (Figure 3.4). Here I found that while these classes are indistinguishable in the first analysis, they can be separated out when focusing in on these weaker phenotypes. There are also cell line dependent effects since these compounds are more distinct in the FLO-1 than the MFD-1 cells. When I performed hierarchical clustering I found that, when split into three clusters, cluster 1 represents protein synthesis, cluster two represents statins, and cluster 3 contains the protein degradation compounds along with some of the protein synthesis compound treatments.

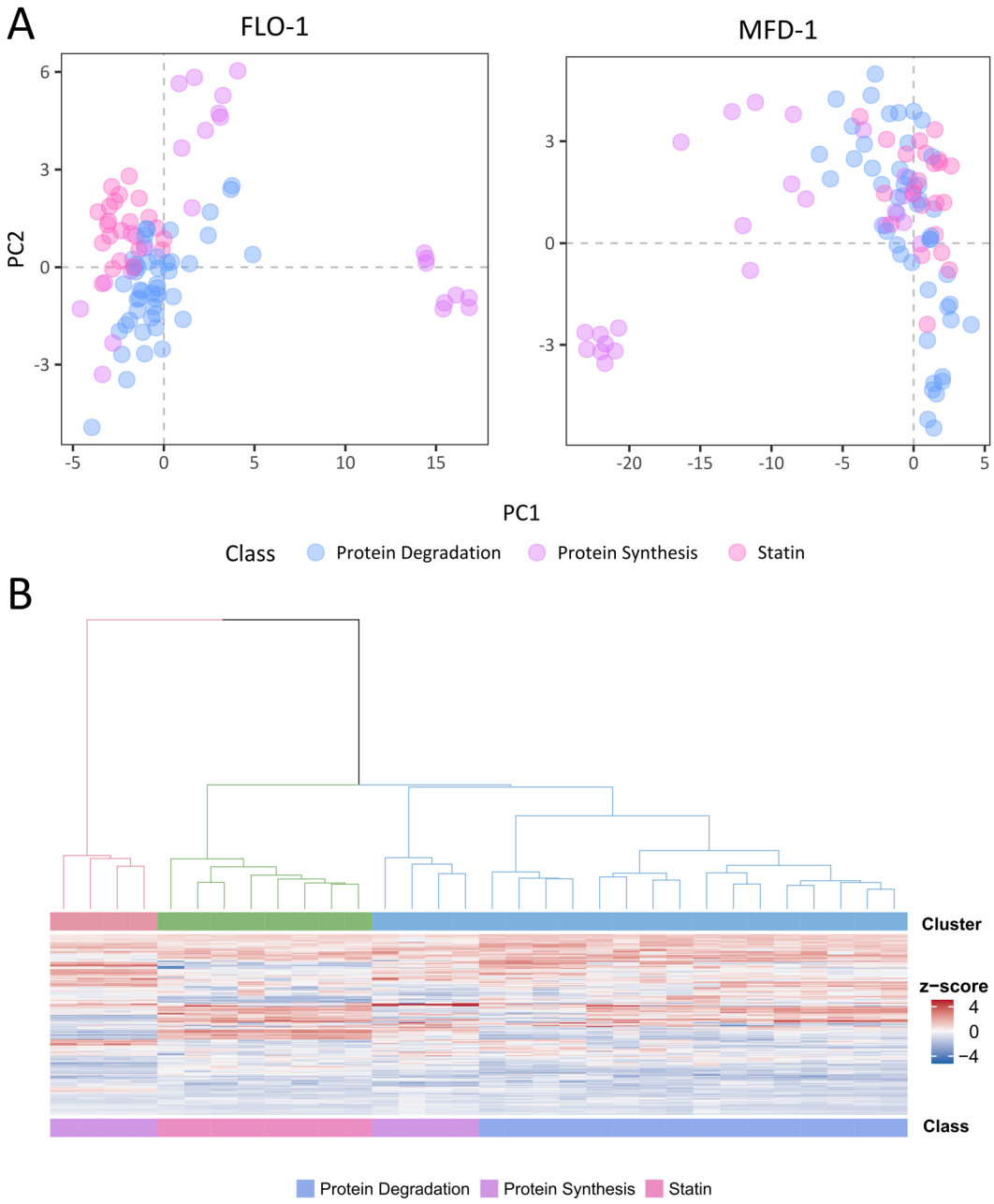



**Figure 3.4: Unsupervised clustering of subtle phenotypic effects**. A) Principal Component plots for the OAC lines FLO-1 and MFD-1 when the reference library of compounds is subset for Protein Degradation, Protein Synthesis and Statins. B) Hierarchical clustering heatmap for the FLO-1 cell line subset for the same compounds as in A at 10 μM. Dendrogram built using Euclidean distance and Ward's method.

## Supervised Machine Learning: Classification

I next optimised a random forest classifier on the same CellProfiler extracted feature dataset to quantitate mechanism of action prediction. The extracted features from three concentrations of each compound were pooled and used to train the classifier. Here I chose 0.1, 3 and 10 μM, since using a broad range of concentrations means compound potency does not need to be optimised individually across each cell line.

When trained and tested on each cell line individually overall prediction accuracy ranged from 75-85 % (Figure 3.5, Table 3.2), demonstrating the suitability of all eight cell lines to the Cell Painting assay.

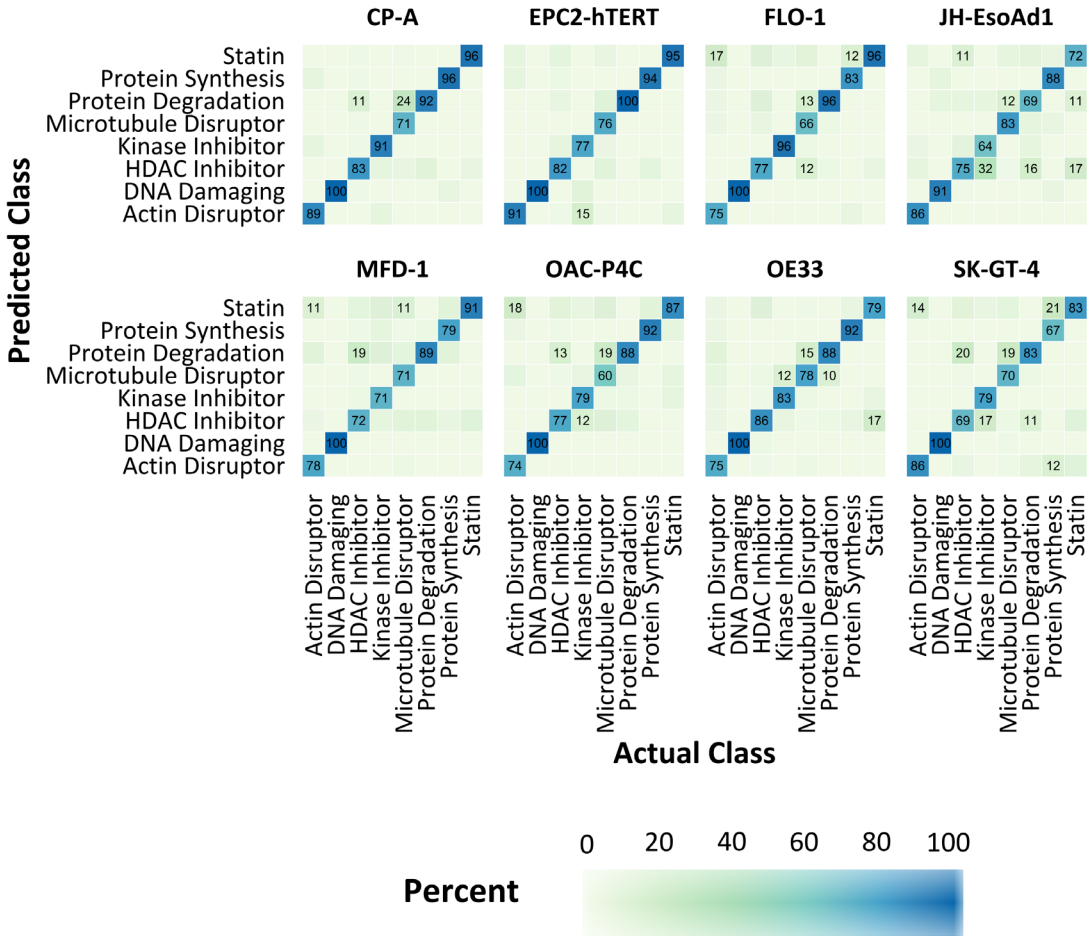



**Figure 3.5: Reference library supervised machine learning.** Confusion matrices of random forest prediction accuracies when trained and tested on each individual cell line. Diagonal values show class sensitivities. Off-class predictions greater than 10% are also displayed.

In the final section of this study, I wanted to demonstrate the transferability of the classifier to new cell lines; implemented leave-one-out cross-validation, training the classifier on five of the six OAC cell lines and testing it on the remaining one. As expected it performed less well overall, suffering a decrease in accuracy for all cell lines. However, the overall accuracy for each cell line ranged from 62-76% (Figure 3.6, Table 3.2), indicating the ability of the classifier to be transferred to new cell lines despite no prior training on them. This demonstrates that the

underlying phenotypic response is consistent between cell lines despite differences in basal morphology and strength of response to any given compound. Thus there is potential for application of the classifier across a broader panel of cell lines without the need to train each cell line individually.

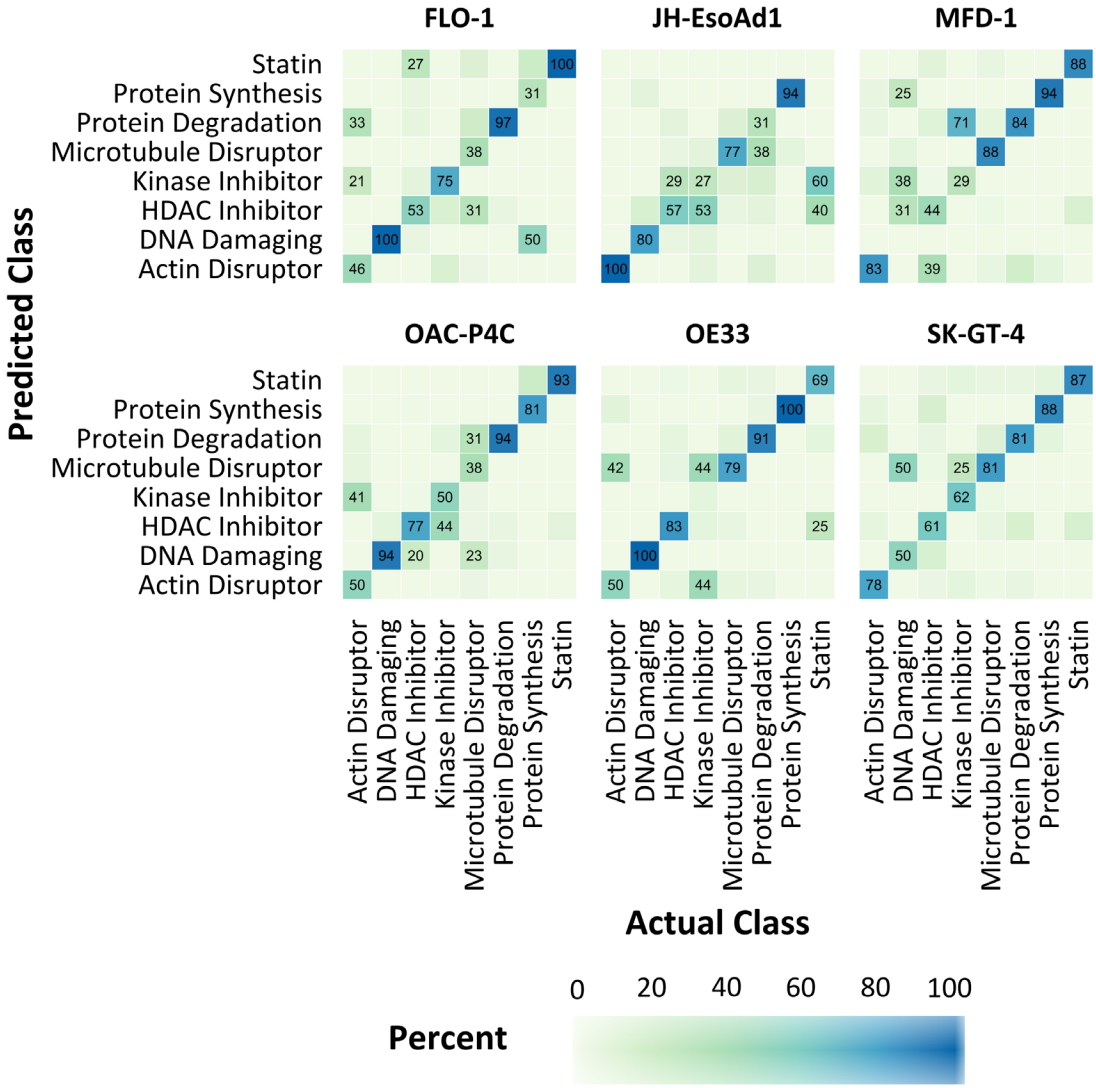



**Figure 3.6: Leave one cell line out-reference library supervised machine learning**. Confusion matrices of random forest prediction accuracies when trained on five of the oesophageal adenocarcinoma cell lines and tested on the remaining one (leave one out). Diagonal values show class sensitivities. Off-class predictions greater than 10% are also displayed numerically.

**Table 3.2: Random forest classifier accuracy for individually trained and leave-one-out classifiers**. Macro-F1-score across all compound classes expressed as a percentage.

| | CP-A | EPC2-hTERT | FLO-1 | JH-EsoAD1 | MFD-1 | OAC-P4C | OE33 | SK-GT-4 |
|---|---|---|---|---|---|---|---|---|
| Individual | 86.94 | 83.51 | 83.00 | 76.10 | 78.10 | 79.35 | 83.51 | 76.57 |
| LOO | - | - | 61.22 | 59.81 | 61.66 | 68.28 | 75.90 | 71.11 |

## 3.3 Discussion

This study of compound induced cell morphology answers a number of important questions. Firstly, I wanted to assess the suitability of the cell lines for the Cell Painting assay and a large-scale screen. From a technical standpoint, I began by asking whether the assay is able to capture sufficient information to distinguish distinct compound treatments in each cell line. While it has been known for a while that compounds induce quantifiable morphological changes (Perlman *et al.*, 2004), and that compounds with a shared mechanism of action produce similar morphological changes (Caie *et al.*, 2010; Gustafsdottir *et al.*, 2013; Ljosa *et al.*, 2013), alongside others, I have shown that this is applicable at scale across diverse compound classes using a reference library of known compounds (Warchal, Dawson and Carragher, 2019; Janssens *et al.*, 2020; Hughes *et al.*, 2020). Using unsupervised machine-learning methods this work demonstrates that the Cell Painting assay, cell lines and downstream analysis pipeline produce quantifiable morphological changes in response to compound treatments and that the strength of this response varies depending on compound class, concentration, and to a lesser extent, cell line. Further this work shows for the first time in oesophageal cancer cells, as has been seen in other assays and cell lines, compounds cluster morphologically by mechanistic class and importantly that classes can be differentiated based on induced cellular morphological changes alone.

In order to quantitatively assess the suitability of the cell lines I tested the accuracy of a random forest classifier in predicting compound mechanism of action. The supervised machine learning carried out in this section confirms that all eight cell lines that had passed the previous screening criteria (Table 3.1) performed equally well and met our criteria for high content profiling. I therefore included all eight cell lines in a large-scale compound study in chapter 2.

There are several points of interest regarding compound classification using high content data. Firstly it is important to note that phenotypic changes associated with a compound treatment occur at the pathway or network level and therefore phenotypic clusters are based on a shared mode of action rather than a shared target (Schneidewind *et al.*, 2020). This has two important implications; firstly, within a given mechanistic cluster there could be multiple compound targets so further mechanistic studies are required to narrow down the exact target/targets associated with a compound; secondly, compound classes can be subdivided many times over and exploring sub-divisions of clusters at different levels and the ability to correctly separate labelled sub-classes, e.g. the microtubule destabiliser subclass within the microtubule disruptor class, may allow more specific labelling of compounds.

Comparing the methods used here with those previously described in the literature, I find that while specific methods used are diverse, the performance is generally comparable prior to more rigorous optimisation of both the data included and methods applied (Ljosa *et al.*, 2013; Twarog *et al.*, 2016; Rose *et al.*, 2018; Warchal, Dawson and Carragher, 2019; Janssens *et al.*, 2020). Variations to improve accuracy include utilising single cell data rather than well level medians (Ljosa *et al.*, 2013), removing phenotypically inactive concentrations of compounds (Twarog *et al.*, 2016), and increasing the number of features included in the dataset (Rose *et al.*, 2018). Interestingly, regarding the number of features used, it has recently be shown that comparable levels of accuracy can be achieved using only the nuclear features (accuracy 0.68), increasing slightly with the inclusion of additional stains (Rose *et al.*, 2018), suggesting nuclear morphology reflects the

general state of the cell. It further suggests a high level of redundancy in phenotypic features.

Lastly, this work investigated the transfer of a classifier between six heterogeneous OAC cell lines. Here I was able to generate a classifier capable of accurately predicting compound mechanism of action for a previously unseen cell line, validating the hypothesis that compounds induce underlying core morphological responses that can be detected despite cell line variations in genetics, basal morphology, and strength of response. Testing this across different cancers and diseases would be very interesting and could lead to a universal classifier for compound mechanism of action studies.

To apply the methods developed here to mechanism of action deconvolution for high content screening hits in the future several questions remain to be answered. Firstly, in the work discussed here, compounds were considered as individual treatments, in isolation from their concentration information. Furthermore only a single fixed time point was studied. However, we know that compounds do not cause a single phenotype but rather they show dose dependent phenotypic trajectories, moving from phenotypically inactive to phenotypically active over a series of increasing concentrations (Perlman *et al.*, 2004; Twarog *et al.*, 2016). While compounds of different classes may produce similar phenotypes at fixed concentrations, it is likely that their phenotypic trajectories differ across dose responses. In the future, integrating the concentration information to look at phenotypic trajectories over dose ranges may provide more accurate predictions of compound mechanisms and distinction of sub-classes based on intermediate phenotypes. It may also allow the identification of off-target effects. While untested, my hypothesis is that compounds acting on a single target will more or less cause phenotypic changes in a single phenotypic direction, increasing in strength with increasing concentration. Whereas, compounds that act on multiple targets with varying efficacy will display a more varied phenotypic trajectory across dose ranges.

Although very little work has been carried out to study temporal phenotypic changes induced by compounds, it is likely that this could give additional information as compounds may inducing differing intermediate phenotypes, again increasing our ability to distinguish compound classes. In proteomic and transcriptomic studies, temporal and concentration dependent changes are often taken into account when studying compound induced effects. For example, it was recently demonstrated that using fewer cell lines and two time points led to the identification of more pharmacologically perturbed gene expression changes than using more cell lines at a single time point (De Wolf *et al.*, 2018).

A key limitation in this work is the breadth and depth of the reference library. A classifier is only as good as the ground truth set from which it is built. There are limited compounds in several of the classes, potentially limiting the ability of the classifier to predict subclasses within each mechanistic class. However, there is a trade-off between using a very well annotated set of clean compounds and using a larger number of more weakly annotated compounds. Regarding the breadth of the library, increasing the number of classes would improve the utility of the classifier as a tool to aid phenotypic drug discovery. The recently announced Joint Undertaking in Morphological Profiling with Cell Painting (JUMP-CP), which aims to create the largest Cell Painting dataset including 140,000 different treatment conditions including both small molecule and genetic perturbations (https://jump-cellpainting.broadinstitute.org/), could provide a huge reference dataset to advance the field of cell-based phenotypic screening and enhance applications in predicting mechanism of action.

In summary, I have shown that mechanistic compound classes can be differentiated based on induced cellular morphological changes and that underlying core changes can be detected across oesophageal cell lines despite cell line variations in basal morphology and genetics. In the following chapter I build on this work and use cell morphology to identify active compounds from a large-scale screen, group them based on response, and using this reference library and classifier, define processes altered by the compounds to elucidate mechanism of action.

# *4| High Content Profiling in Oesophageal Adenocarcinoma*

## 4.1 Introduction

OAC is a highly heterogeneous disease, dominated by large scale genomic rearrangements and copy number alterations (Secrier *et al.*, 2016). This has made the identification of clinically meaningful subgroups and well validated therapeutic targets difficult to define. Clinical trials with new molecular targeted agents have predominantly been directed towards epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2) receptors (Bang *et al.*, 2010; Okines *et al.*, 2010; Dutton *et al.*, 2014; Hecht *et al.*, 2016) but thus far have shown little benefit. A potential explanation is the almost ubiquitous co-amplification of alternative receptor tyrosine kinases (RTKs) and downstream pathways leading to redundancy and drug resistance (Dulak *et al.*, 2012; Paterson *et al.*, 2013; Secrier *et al.*, 2016). Failure of target-based drug discovery to generate effective treatments contributes to the classification of OAC as a complex heterogeneous disease of unmet therapeutic need.

Phenotypic drug discovery, represents a discovery strategy which applies screening technologies across physiological models of disease to identify novel therapeutic targets or therapeutic agents without prior knowledge of the drug target. Phenotypic drug discovery has therefore been proposed as an alternative

and complementary approach to contemporary target-directed drug discovery strategies for addressing complex diseases of unmet need where target biology is poorly understood (Swinney and Anthony, 2011; Moffat, Rudolph and Bailey, 2014). OAC has previously been studied in a number of small scale compound phenotypic screens (Li *et al.*, 2018; Lohse *et al.*, 2018). Of note, both studies highlighted differing responses across the OAC cell lines, demonstrating the heterogeneity of the disease and the importance of cell panel screening. Although, these small scale studies of approved compounds may lead to a better standard of care therapy but are unlikely to provide new insights into the disease itself or novel targeted therapeutics.

This chapter describes a comprehensive small-molecule phenotypic screen, using a high content Cell Painting assay to measure multiple phenotypic endpoints across a panel of six OAC cell lines and two tissue matched normal controls. In total, this assay phenotypically profiled 19,555 small molecules, consisting of commercially available, pharmacologically active, target-annotated, and off-patent drugs (2947 compounds) representing repurposing opportunities; and diverse chemical and lead-like compounds (16,608 compounds) representing novel chemical space. This phenotypic screen encompassed eight cell lines, 512x384 well plates, 3.9 million images and 36 Terabytes of imaging data.

High content profiling offers a number of advantages over both target-based and single endpoint phenotypic methods; it is suited to the study of complex diseases with unmet therapeutic need where target biology is poorly understood; it is more robust to false positives and negatives (morphological changes occur at a wider range of concentrations than cell death); it also facilitates characterization of mechanism-of-action in parallel to hit identification, streamlining the discovery process. It was for these reasons that we chose a high content screen to identify new therapeutics and novel drug targets in OAC.

We began our analysis with the subset of the 19,555 compounds, representing repurposing libraries:

- Prestwick Chemical Library® of 1280 FDA/EMA approved drugs designed to represent broad pharmacological diversity of all approved small molecule drug classes.
- Library of Pharmacologically Active Compounds (LOPAC) 1280 small molecules compounds representing approved drugs, safe drugs (passed Phase 1) and those in clinical development
- Bespoke libraries focussed on target and class, including SCREEN-WELL® Protease (53 compounds); Kinase (80 compounds), Epigenetic (43compounds) compound libraries from Enzo Life Sciences and 116 epigenetic target focused compounds provided by the Target Discovery Institute and Structural Genomics Consortium, University of Oxford.

In this work I set out to identify compounds and target classes that are selective for OAC and suitable for repurposing or new target-based drug discovery pipelines. Included in our phenotypic profiling process we wanted to determine whether we can use a high-throughput morphological analysis of single compound concentration treatments to predict compounds that go on to validate as selectively killing cancer cells at the dose response stage. Secondly, following on from the previous chapter we wanted to evaluate whether we can use the phenotypic dose response information to define molecular processes and pathways altered by hit compounds and elucidate mechanism of action. This is made possible using the repurposing and target-annotated libraries, representing compounds with predefined mechanisms, as we can use this knowledge to validate our predictions made through similarity clustering and machine learning. This in turn would allow us to apply our phenotypic profiling assay with confidence to larger diverse chemical libraries in future work.

## 4.2 Results

### 4.2.1 Primary hit identification

To identify active compounds, I employed two parallel analyses of the Cell Painting assay data for primary hit selection; one, a multivariate analysis based on broad morphological changes, and the other, a univariate survival analysis using nuclei counts. Using a panel of cell lines allows us make comparisons between the OAC and tissue-matched control lines and prioritise compounds that are selectively active across multiple OAC cell lines.

Firstly, I used the Cell Painting data to identify compounds that induced strong morphological changes selectively in the OAC cells. To do this I applied principal component analysis (PCA) to reduce the dimensions of the data, and then calculated the Mahalanobis distance (Mahalanobis, 1936) for each compound to the DMSO controls. This multiparametric analysis led to the identification of 62 compounds (a 2 % hit rate) that were selectively active in two or more of the OAC lines over the tissue-matched control cells (Table A.1). See Methods for selection criteria details. Visual inspection by PCA showed enrichment of pharmacological classes within the phenotypic clusters produced, suggesting hits clustered mechanistically (Figure 4.1).

Toxic compounds that killed a cell line could not be assessed in the phenotypic analysis due to a lack of cells and were instead identified in a survival analysis based on cell number. This analysis identified 27 compounds (0.9 % hit rate) that were selective for at least two of the OAC cell lines when compared to the tissue-matched controls (Table A.1). Hits were enriched for several therapeutic classes suggesting mechanistic pathways that may be selective for OAC cell survival. These classes include antimetabolites and histone deacetylase (HDAC) inhibitors.

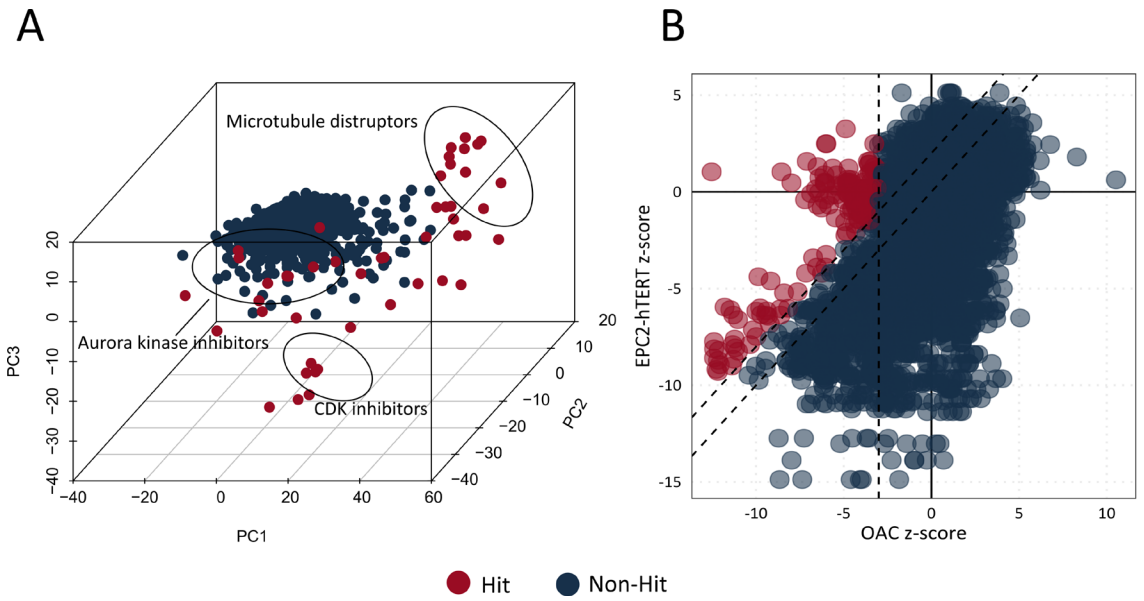



**Figure 4.1: Hit analysis.** A) Multivariate Phenotypic hits: First three principal components for exemplar JH-EsoAD1 cell line. Mechanistic clusters are highlighted. B) Univariate Nuclei Count hits: OAC z-scores vs exemplar EPC2-hTERT comparison. Hits are highlighted in red.

Hierarchical clustering of both the Mahalanobis metric (multivariate analysis) and z-scores (univariate analysis) enabled global trends in the cell lines responses to the annotated compound libraries (Prestwick, LOPAC and bespoke) to be assessed (Figure **4.2**). Clustering of both analyses showed that the tissue-matched control cell lines; EPC2-hTERT and CP-A, can be differentiated from the OAC cell lines based on global screening data, providing confidence that our high content assay can identify compounds with selectivity for OAC over the tissue-matched control lines.

Overall, analysis of the subset of the repurposing libraries led to the identification of 74 selective hits that were prioritised for validation.

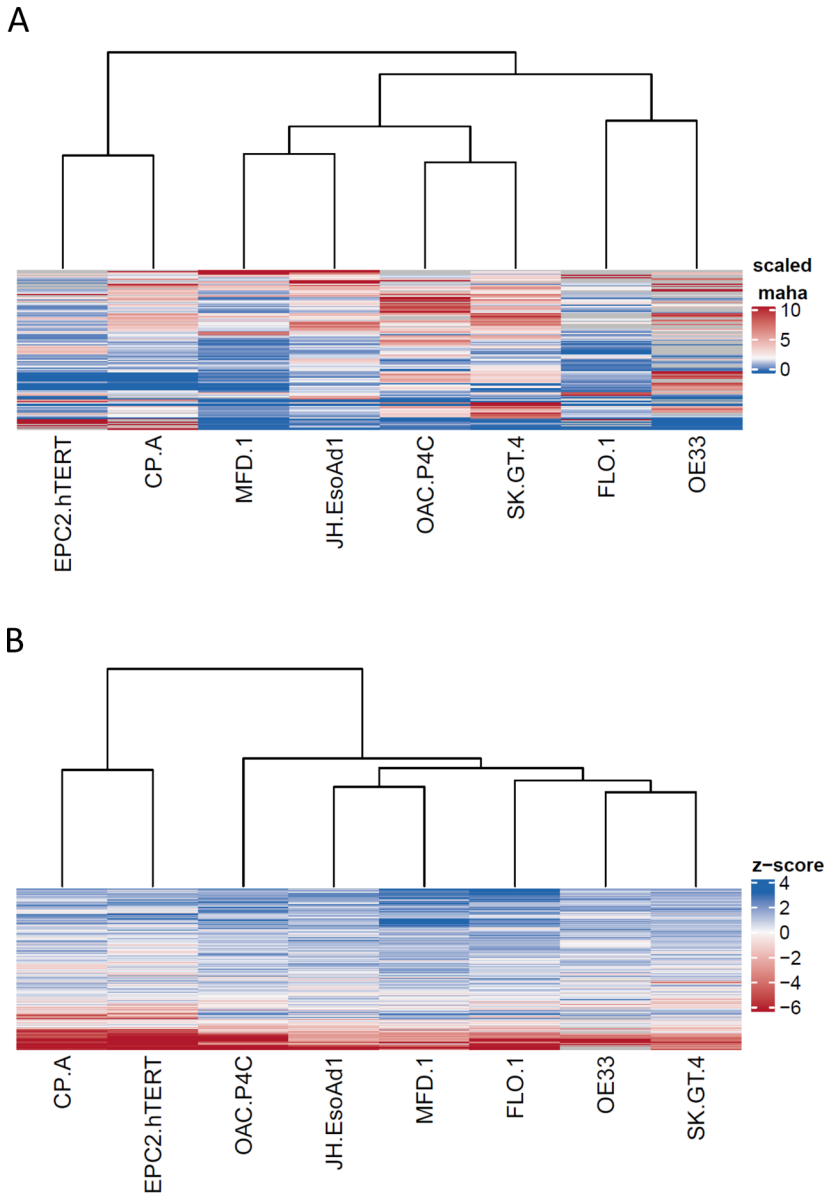



**Figure 4.2: Hierarchical clustering heatmap for the cell panel across the repurposing libraries**. A) Multivariate phenotypic clustering of Mahalanobis distances. Cell lines (columns) clustered by compound induced Mahalanobis distances (rows). B) Univariate nuclei count clustering of z-scores (negative z-score = active compound). Cell lines (columns) clustered by compound z-scores (rows). Dendrograms built using Euclidean distance and Ward's method.

### 4.2.2 Dose response validation

Concentration response relationships describe the magnitude of response as a function of dose for a given exposure time. The shape of the concentration response curve reflects the biological activity and potency of a compound. Dose responses are commonly used to validate screening hits and rank compound activity. In cancer cell line compound screening univariate dose responses usually follow cell survival as a function of compound concentration in order to determine the concentration at which 50% of the cells survive compared to a DMSO control ($IC_{50}$). Here I also wanted to develop a multivariate dose response, to reflect the multivariate primary endpoint, in order to validate morphological changes over a range of concentrations from maximal to minimal effect. Hits were profiled in 10 point semi-log dose responses against the cell panel using the Cell Painting assay to get a univariate (cell survival) and multivariate (morphological) concentration response.

72 of the 74 compound hits were available for validation.

#### Univariate Dose responses

Testing the 72 compounds for dose dependent cell survival, we found that 47 of the compounds reduced cell survival in a dose dependent manner (Table A.2) of which 11 showed selectivity for the OAC cells over the tissue-matched control lines (Figure 4.3**,** Table 4.1). Since only 27 hits were identified in the primary univariate analysis a large number of these 47 validated hits come from the multivariate primary analysis, demonstrating the ability of phenotypic profiling to identify compounds that go on to validate as affecting cell survival. Classes that were identified repeatedly include Aurora kinase inhibitors, HDAC inhibitors, Thymidylate synthase inhibitors, Topoisomerase II inhibitors and microtubule disruptors. Moreover, three of the 11 compounds that selectively affected OAC survival were also initially identified in the multivariate primary analysis not the cell survival analysis, further demonstrating the utility of phenotypic screening, capable of detecting compounds that would otherwise go undetected at the single concentration stage.

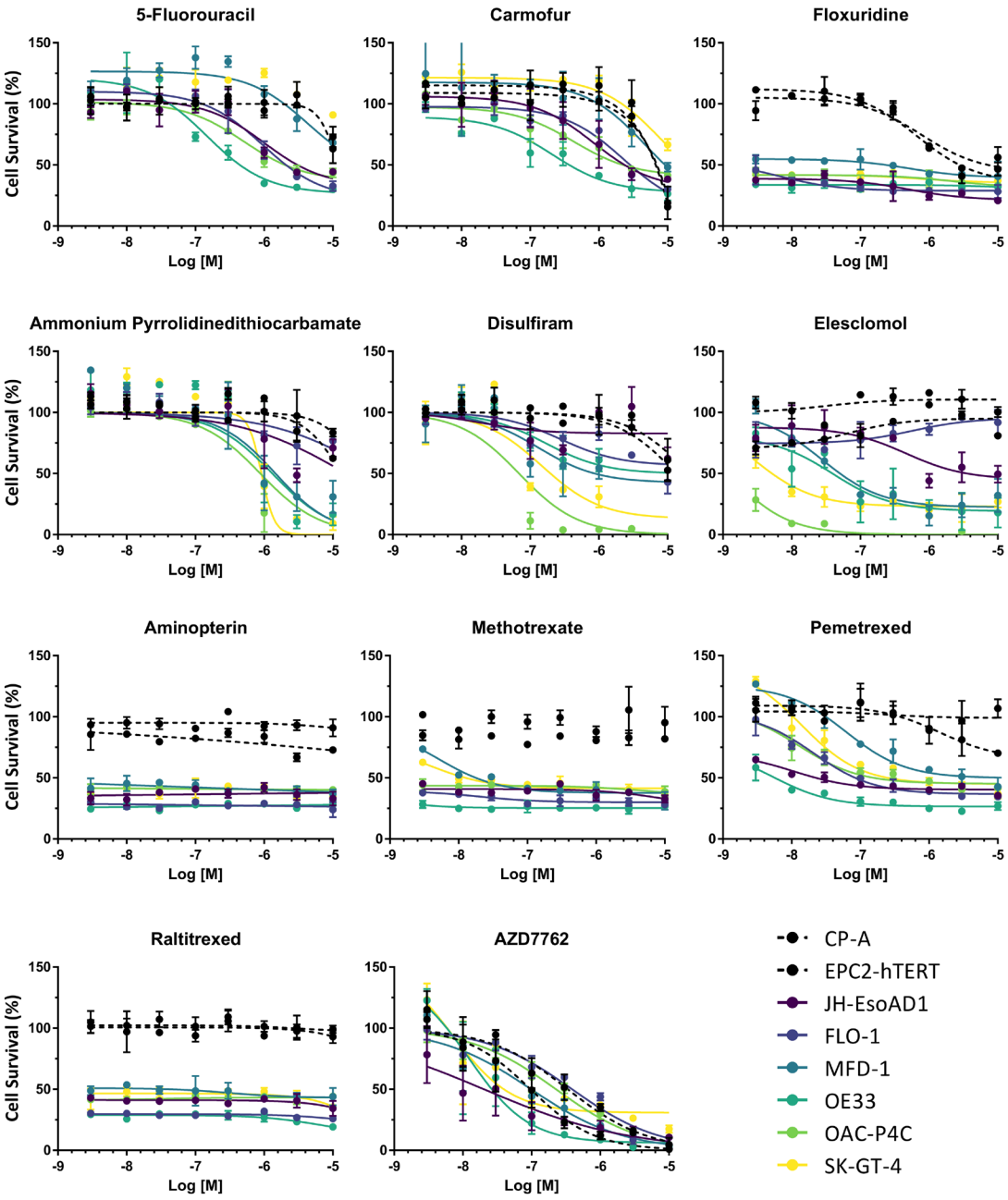



**Figure 4.3: Univariate dose responses for 11 OAC selective compounds.** Cell counts normalised to DMSO. Tissue matched control cell lines are in black and curves are dashed. Curves fitted in GraphPad using the sigmoidal dose response.

**Table 4.1: OAC selective compound hits.**

| Compound | Mechanism |
|---|---|
| 5-Fluorouracil | Thymidylate synthase inhibitor |
| Carmofur | 5-FU prodrug/acid ceramidase inhibitor |
| Floxuridine | 5-FU prodrug |
| Methotrexate | Dihydrofolate reductase inhibitor |
| Pemetrexed | Dihydrofolate reductase inhibitor |
| Raltitrexed | Dihydrofolate reductase inhibitor |
| Aminopterin | Dihydrofolate reductase inhibitor |
| Ammonium pyrrolidinedithiocarbamate | NF-kB inhibitor |
| Disulfiram | Mitochondrial aldehyde dehydrogenase inhibitor |
| Elesclomol | Hsp70 inducer |
| AZD7762 | Checkpoint kinase inhibitor |

### Multivariate Dose Reponses

Having calculated the univariate dose responses, it was then important to validate the multivariate analysis by way of multivariate dose responses, to quantify dose dependent phenotypic changes. Here 51 of the 72 total compounds (71%) caused dose dependent phenotypic changes (Table A.2), of which, 46 (64%) were selective for the OAC cells over the tissue-matched controls.

In summary, this work has led to the identification of 72 compounds of which 48 are selectively active in two or more of the OAC cell lines compared with the tissue-matched controls (Table 4.2 and Table A.2). Further analysis and experiments in appropriate preclinical models could identify which of these compounds show promise as repurposing opportunities. Three of the most potent and selective hits (Elesclomol, Disulfiram, Ammonium pyrrolidinedithiocarbamate) identified in both the univariate and multivariate analyses were put forward for further study in chapter 3.

**Table 4.2: Screen and validation summary.** See Appendices for tables of compounds identified in each analysis

| Total hits | **Total Validated** | Validated Phenotypic | Validated Survival | **Total Selective** | Selective Phenotypic | Selective Survival |
|---|---|---|---|---|---|---|
| 72 | **51 (71 %)** | 51 (71 %) | 47 (65 %) | **48 (67 %)** | 46 (64 %) | 11 (15 %) |

### 4.2.3 Validation of machine learning and class prediction

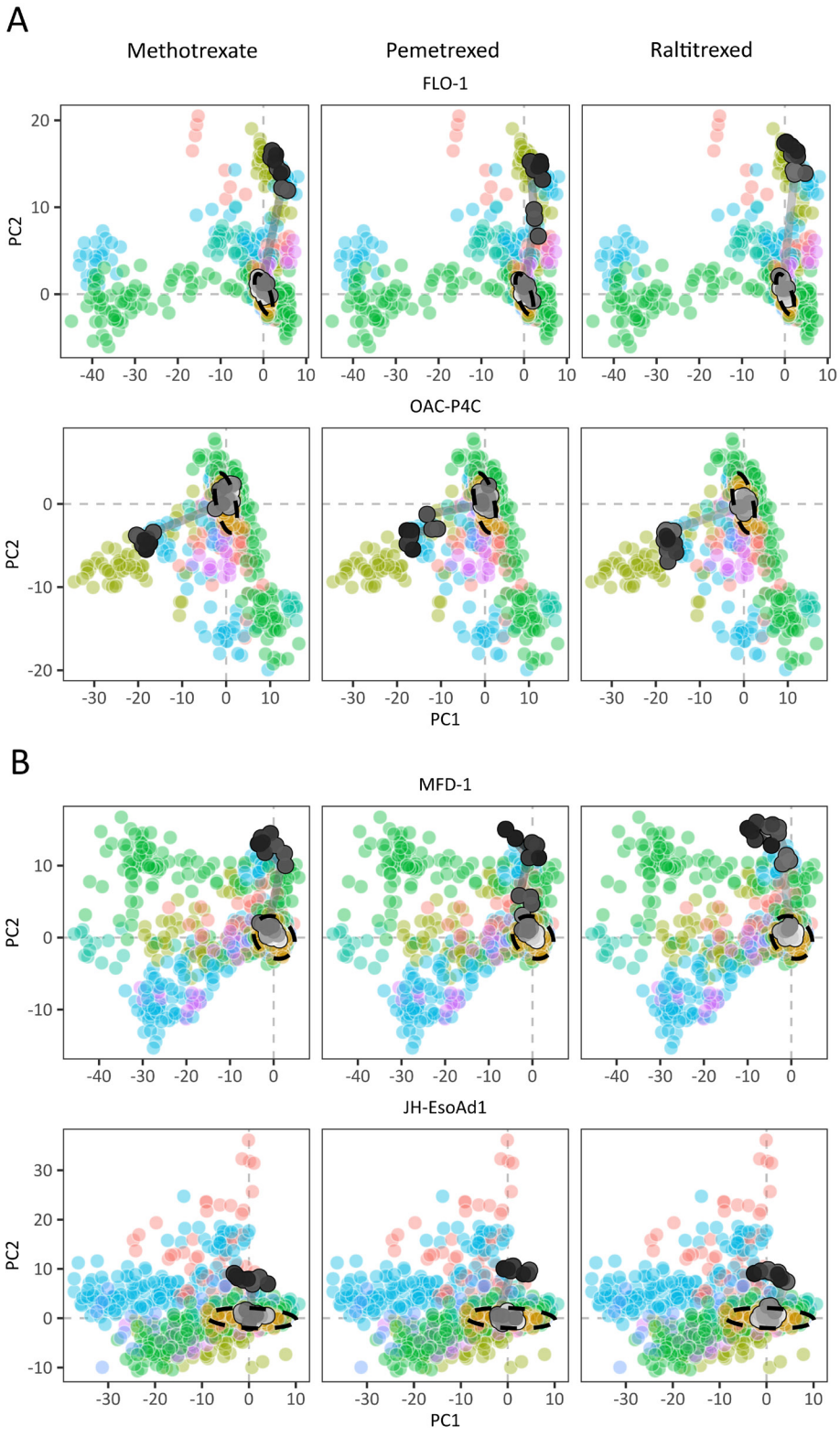
#### Phenotypic trajectories and clustering

Using these multivariate dose responses, I wanted to answer a number of questions. Firstly, I wanted to validate the hypothesis that compounds show dose dependent phenotypic trajectories from the DMSO, moving to cluster by their mechanism of action at active concentrations. In order to visually interrogate the multivariate dose responses, I implemented PCA. Since PCA depends on the variance of the data input, I overlaid the dose response data for each compound on top of the reference library of compounds, as this gives a known phenotypic space for each cell line. In general, I found compounds displayed dose dependent phenotypic trajectories starting at the DMSO controls, representing the basal phenotype, and increasing in activity towards their maximal phenotypic effect, validating my hypothesis (Figure 4.4).

Secondly, I hypothesised that compound hits representing one of the mechanistic classes contained in the reference library should show phenotypic dose response trajectories akin to the compounds of that class in the reference library. Moving from inactive at low concentrations, resembling a DMSO phenotype, towards a class specific phenotype at higher concentrations clustering with the representative class. To confirm this I considered three antimetabolites identified in our screen as selective; Methotrexate, Raltitrexed, and Pemetrexed that are known DNA damaging agents.

Interestingly, while half of the OAC cell lines responded as predicted, showing strong dose dependent phenotypic changes, moving from phenotypically inactive (clustering with DMSO controls) to clustering with the DNA damaging agents at active concentrations (Figure 4.4**A**), intriguingly the JH-EsoAD1 and MFD-1 lines did not. The dose responses for all three compounds in these cell lines follows a trajectory that ultimately clusters away from the DNA damaging agents, with a subset of the microtubule disruptors- Aurora kinase B inhibitors. Further inspection of the cell images confirms a visual similarity between the Aurora kinase B inhibitor Barasertib and the three antimetabolites in the JH-EsoAD1 cell line (Figure 4.5**)**. However, the reason for this differential response in the JH-EsoAD1 and MFD-1 cells is unknown. It is clear that the reference DNA damaging agents are relatively toxic to the JH-EsoAD1 cell line even at 0.03 μM (figure 4.5) and it is therefore possible that the overall phenotype detected for this class in the reference library is one of cell death and is therefore not class specific. It may be that at lower concentrations the reference DNA damaging agents produce a more class specific phenotype, and that this may better match the antimetabolites phenotype. These findings highlight the importance of concentration optimisation if such phenotypic profiling tools are to be used to predict the mechanism of hits in high content screens in the future.

All three compounds also showed little or no effect in the control lines EPC2-hTERT and CP-A, clustering closely with the DMSO controls at all concentrations tested (Figure 4.4C). This reflects the selectivity towards OAC seen in the univariate dose responses.


A
Methotrexate
Pemetrexed
Raltitrexed
FLO-1
OAC-P4C
PC2
PC1
B
MFD-1
JH-EsoAd1
PC2
PC1

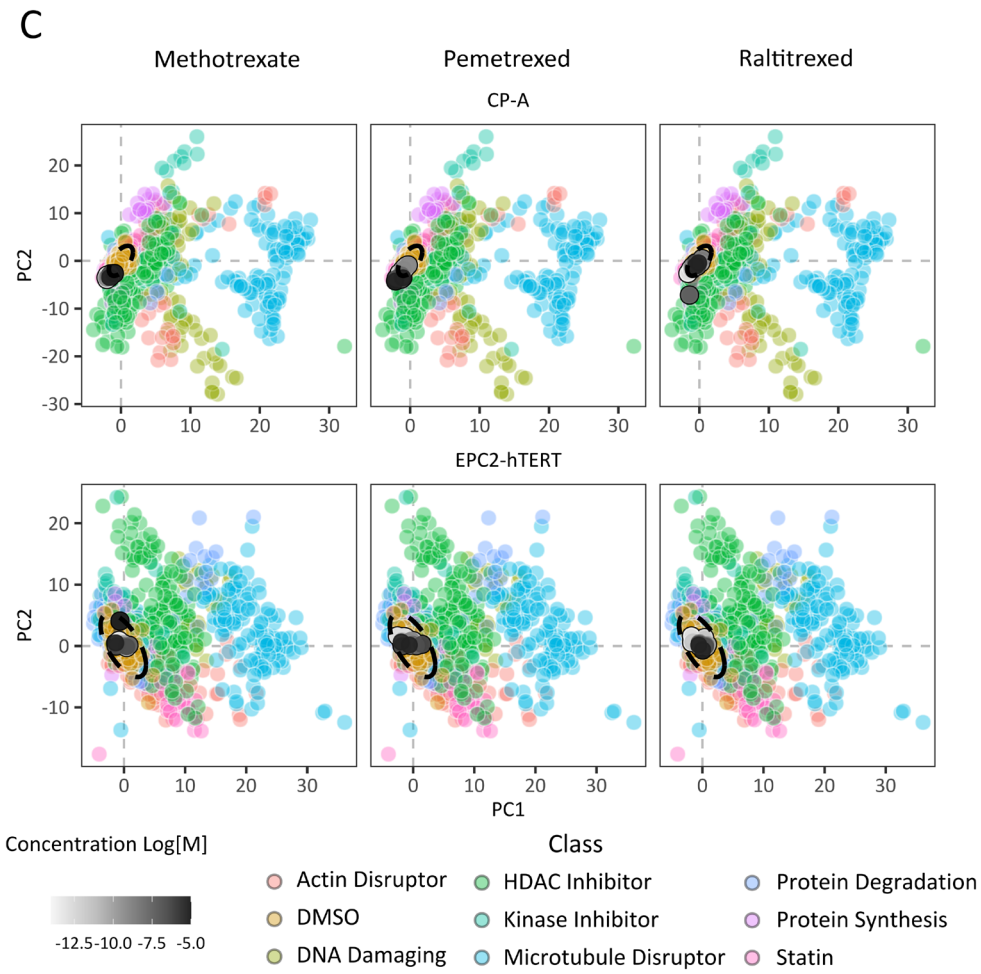



**Figure 4.4: Multivariate phenotypic dose responses for antimetabolite hit compounds.** Phenotypic dose responses for Methotrexate, Pemetrexed and Raltitrexed in A) Exemplar OAC cell lines with expected trajectories. B) Cell lines with unexpected trajectories. C) Tissue matched control cell lines. Library of reference compounds coloured by mechanistic class. Compound dose response overlay coloured by concentration from 10- 0.0001 µM.

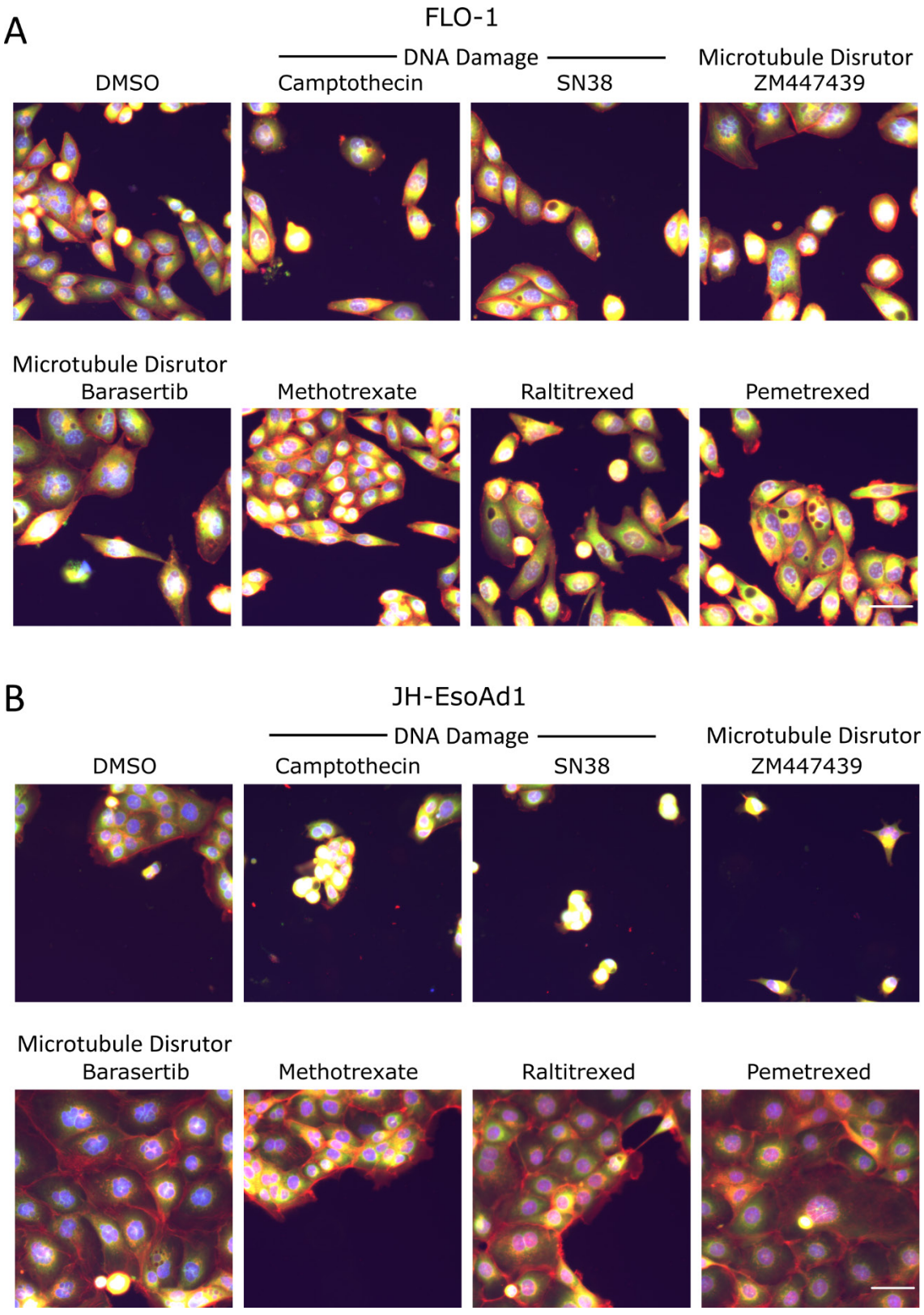

A
FLO-1
DNA Damage
DMSO
Camptothecin
SN38
Microtubule Disrutor
ZM447439
Microtubule Disrutor
Barasertib
Methotrexate
Raltitrexed
Pemetrexed
B
JH-EsoAd1
DNA Damage
DMSO
Camptothecin
SN38
Microtubule Disrutor
ZM447439
Microtubule Disrutor
Barasertib
Methotrexate
Raltitrexed
Pemetrexed

**Figure 4.5: Colour combined images of reference classes and antimetabolites.** Colour combined exemplar images for DMSO, DNA damage (0.03 μM) and microtubule disruptor (10 μM) reference compounds and the three antimetabolites (0.3 μM) for A) FLO-1 and B) JH-EsoAD1 cell lines. Hoescht 33342- nuclei (blue); Phalloidin 594, Wheat germ agglutinin and Alexa Fluor 594- F-actin, golgi and plasma membrane (red); Concanavalin A Alexa Fluor 488- Endoplasmic reticulum (green). Scale bars 50 μm.

### Supervised machine learning

Using the same compounds above, I applied the pre-trained machine learning classifier (chapter 1) to predict the mechanism of action across the dose response data. Here I predicted that compounds would be classified as DMSO at very low concentrations, and then as the concentration increases they would be classified as DNA damaging agents with increasing probability. From the clustering and images described above, I also expected that the classifier would fail to correctly classify the compounds in both the JH-EsoAD1 and MFD-1 cell lines.

Results from the pre-trained machine learning model for each of the DNA damaging compounds are consistent with the unsupervised PCA data. Class probabilities predict them to belong to the DNA damage class for all but the MFD-1 and JH-EsoAD1 lines **(**Figure 4.6**).** As predicted, probabilities increase in a dose dependent manner indicating cellular phenotypic activity follows an on-target dose response relationship.

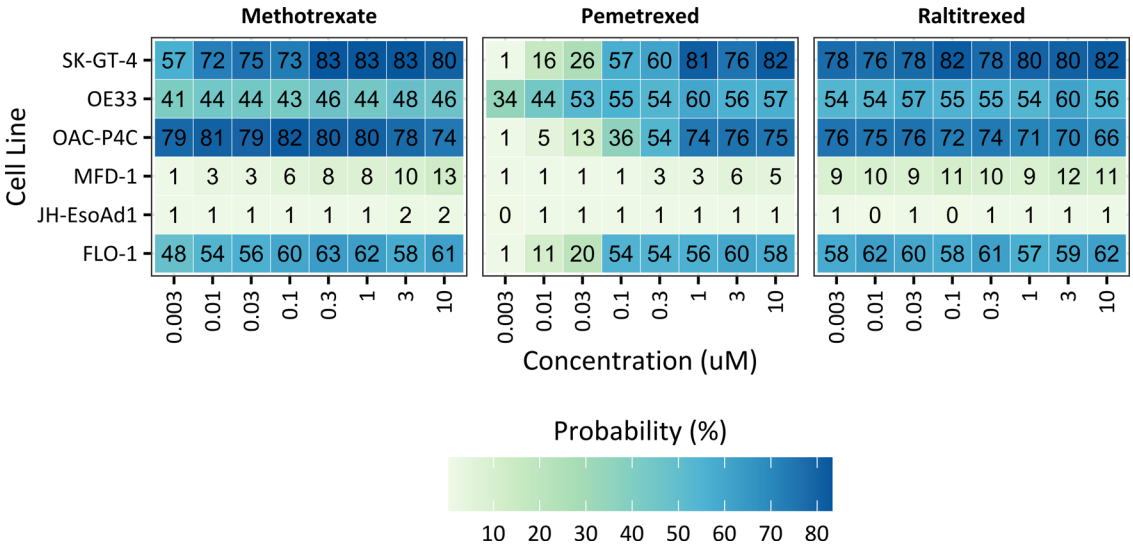



**Figure 4.6: Machine learning class probabilities for the antimetabolites.** Probability that Methotrexate, Pemetrexed and Raltitrexed belong to the DNA damaging class across a dose response in each OAC cell line. Probability expressed as a percentage.

Taken together, using a class of compounds identified in the screen with known activity has further confirmed the ability of the Cell Painting assay to accurately predict mechanism of action for validated hit compounds. However it also highlights how compounds acting via different mechanisms or targets within the same broad class can show differential responses across cell panels, possibly dependent upon the underlying genetic background of the cell line.

## 4.3 Discussion

Conventional target-directed drug discovery strategies remain to make any impact on the discovery and translation of effective new treatments for oesophageal cancer patients. One of the key challenges in oesophageal cancer is the heterogeneity of the genetic landscape with few mutations in oncogenic drivers, thereby confounding the identification of clear drug-target hypotheses and modern personalized medicine strategies. This study sought to adapt and evaluate the utility of an advanced high content phenotypic screening method as an

empirical strategy to identify pharmacological classes which target OAC, and subsequently to identify their mechanism of action.

Here I have shown that combining high content screening and image informatics with machine-learning can prove effective in the identification and mechanistic characterisation of hit compounds with selective activity upon OAC cell phenotypes.

This chapter has focused on the annotated libraries (Prestwick, LOPAC and bespoke target-annotated libraries) with assay validation and repurposing opportunities in mind. Using the repurposing libraries I have validated the assay through its ability to detect the current standard of care therapies in OAC- 5-Fluorouracil (5-FU) and Epirubicin (Ellence) were identified in the initial screen and both showed dose dependent survival changes. 5-FU, along with two of its prodrugs (Carmofur and Floxuridine) accounted for three of the 11 selective hits validated at the dose response stage, whereas Epirubicin did not validate. This result recapitulates what has been seen in the clinic, with a recent move away from ECF (epirubicin, cisplatin, and 5-fluorouracil) regime to FLOT (5-fluorouracil, leucovorin, oxaliplatin, and Docetaxel) for OAC (Bose *et al.*, 2017). As well as identifying the standard of care compounds, a number of novel repurposing opportunities have also been identified in these libraries and will be followed up in more detail in the remaining chapters.

It is clear from the work presented here that the majority of strong, compound induced phenotypic changes are associated with cell death and consequently we have identified selectively toxic compounds from the approved libraries. It is unsurprising therefore that most of the top hits are chemotherapeutic agents. However, compounds that selectively induce more subtle phenotypes in OAC should be of equal interest for future drug development since these are still selectively targeting OAC specific biology. For example, the selective phenotypic hits that did not induce cell death at validation (Table 4.2 and appendices Table A.2). Integration of this phenotypic information with pre- and post-treatment proteomic and transcriptomic data would allow the identification of the pathways

and networks involved and lead to a better understanding of OAC specific signalling pathways. These in turn could lead to the development of new therapies and biomarkers for OAC.

Beyond the repurposing libraries, and although not the focus of this chapter, there remains a wealth of information in the diverse chemical libraries provided by Cancer Therapeutics Discovery labs and purchased from BioAscsent consisting of 16,608 compounds. Applying the same criteria used here for hit selection and hit validation to the diverse chemical library screening data may elucidate new compounds with novel mechanisms of OAC selectivity.

Using the antimetabolites as an exemplar class identified in the screen, I have demonstrated the power of combining high content profiling and machine learning with a reference library in order to elucidate the mechanism of phenotypic hits. However, why this class of compounds display a very distinct phenotype in the MFD-1 and JH-EsoAD1 cell lines remains to be explored. It appears that these cell lines are significantly more sensitive to the DNA damaging agents in the reference library than the other OAC cell lines. Thus, using lower concentrations may be necessary to obtain a class specific phenotype, and perhaps then the antimetabolites will better cluster with the reference DNA damaging agents. It is possible that there is a genetic difference underlying the differential response displayed by these cells, highlighting the benefits of cell panels in heterogeneous diseases where compounds may display varying activities, mechanisms, and efficacies. Although not tested in this work it is possible that a classifier trained on all OAC cell lines combined would better capture broad class specific phenotypes, negate the need to optimise compound concentrations and be better able to correctly predict the mechanistic class for the antimetabolites.

While we have demonstrated the utility of combining high content profiling and machine learning in order to elucidate compound mechanism its full potential will be realised when studying hits from the diverse chemical libraries, where there is no prior knowledge of targets or mechanisms. Here, high content profiling could enable similarity matching of phenotypic fingerprints to guide target

deconvolution. For example, if the phenotypic fingerprint of an unknown chemical hit matches that of known kinase inhibitors then kinome profiling could be carried out to confirm the mechanism and further identify specific targets.

# 5| *Ammonium pyrrolidinedithiocarbamate, Disulfiram, and Elesclomol: Towards a unified theory of drug mechanism*

An exploration in compound mechanism-of-action.

## 5.1 Introduction

One of the benefits of a phenotypic screen is that it is target agnostic and therefore requires limited understanding of the molecular mechanisms of disease and the underlying target biology. This also creates its Achilles' heel since it means that the mechanism of action of any hit molecule is unknown. Target deconvolution and detailed study of drug action are therefore necessary for the progression of hits and identification of new disease specific pathways.

This chapter follows three of the most selective hits identified in the phenotypic screen; elesclomol, disulfiram, and ammonium pyrrolidinedithiocarbamate. While there is data relating to the activity of all three compounds across a range of diseases and organisms, none have previously been studied in relation to OAC, and therefore disease specific mechanisms are yet to be uncovered.

Structurally, all three compounds contain thiocarbonyl groups. Ammonium pyrrolidinedithiocarbamate is part of the family of dithiocarbamates, while disulfiram, a thiuram disulphide, breaks down in acidic or Cu(II) (copper ion carrying a double positive charge) rich environments to produce a dithiocarbamate; diethyldithiocarbamate (DETC) (Kragh, 2008). DETC and ammonium pyrrolidinedithiocarbamate, like other dithiocarbamates, are known metal chelators and form stable complexes with transition elements, but are most stable in a Cu(II) chelate (Dalecki *et al.*, 2015). Although elesclomol is structurally unrelated to the dithiocarbamates, it is still a potent metal chelator, particularly Cu(II), due to its two thiocarbonyl groups (Wu *et al.*, 2011; Nagai *et al.*, 2012).

Ammonium pyrrolidinedithiocarbamate is primarily thought of as a nuclear factor kB (NF-kB) inhibitor (Schreck *et al.*, 1992; Ziegler-Heitbrock *et al.*, 1993), but it has also been shown to increase NF-kB activity, and exert anti-apoptotic and pro-apoptotic properties depending on the cellular context (Erl, Weber and Hansson, 2000). It is also a known antioxidant.

By chance, whilst study the copper chelating ability of disulfiram, it was discovered through concomitant intake of disulfiram and alcohol, that this produced a number of unpleasant reactions, leading to its introduction in 1948 as an important drug in the treatment of alcoholism (Hald and Jacobsen, 1948; Martensen-Larsen, 1948). The mechanism for which was later discovered to be via inhibition of aldehyde dehydrogenase (ALDH) (Jacobsen, 1952). Since then it has been linked to numerous biological activities including antitumour activity reported to be via a number of cancer-associated pathways including induction of reactive oxygen species (ROS), and inhibition of the proteasome (Lövborg *et al.*, 2006).

Elesclomol is an investigational drug in clinical trials for several cancers and thought to induce reactive oxygen species (ROS) in cancer cells. It was developed by Synta Pharmaceuticals and GlaxoSmithKline as a chemotherapy adjuvant to induce HSP70 via oxidative stress due to build-up of reactive oxygen species (Chen *et al.*, 2013). Several genetic screens have been carried out for elesclomol in a

variety of disease models and organisms looking for a cellular protein target. A chemo proteomic screen in yeast concluded that elesclomol does not have a protein target, since heterozygous deletion strains showed only modest changes in sensitivity (Blackman *et al.*, 2012). Conversely, a recent CRISPR screen proposed FDX1, a mitochondrial reductase involved in iron-sulfur (Fe-S) cluster formation, to be the target of elesclomol since knockdown conferred some level of resistance (Tsvetkov *et al.*, 2019). However, this study was actually carried out using elesclomol analogues, not elesclomol itself.

In summary, all three compounds have been linked to multiple and sometimes conflicting targets and modes of action. This could be for several reasons, firstly, most studies to date have been carried out at excessively high concentrations, often in the micromolar range, possibly leading to off-target and non-specific interactions, perhaps contributing to the many numerous targets proposed in the literature. Secondly, no one study has systematically explored all potential mechanisms in a holistic approach across a consistent panel of sensitive and insensitive cell lines. Lastly, since the mechanism of all three compounds may involve copper, a highly redox active metal, it is possible that these compounds have numerous targets and activities. Systematically assessing the activity of these compounds in OAC using a panel of cell lines displaying distinct sensitivities to the compounds could reveal which of these mechanisms are relevant to their anticancer activity.

In this chapter I begin to characterise elesclomol, disulfiram, and ammonium pyrrolidinedithiocarbamate and provide evidence of a universal mechanism of action shared by all three compounds in OAC.

## 5.2 Results and discussion

### 5.2.1 Validation of copper chelating compounds in OAC

Across both the univariate and multivariate analyses three of the most potent and selective hits identified from the repurposing libraries were elesclomol, disulfiram, and ammonium pyrrolidinedithiocarbamate. These compounds

showed significant selectivity for the OAC cells over the tissue-matched control cell lines (CP-A and EPC2-hTERT) (Table 5.1, Figure 5.1). Regarding survival, elesclomol was the most potent with $IC_{50}$ values in the nanomolar range, then disulfiram and finally ammonium pyrrolidinedithiocarbamate, though there is marked variation in sensitivity across the OAC cell panel.

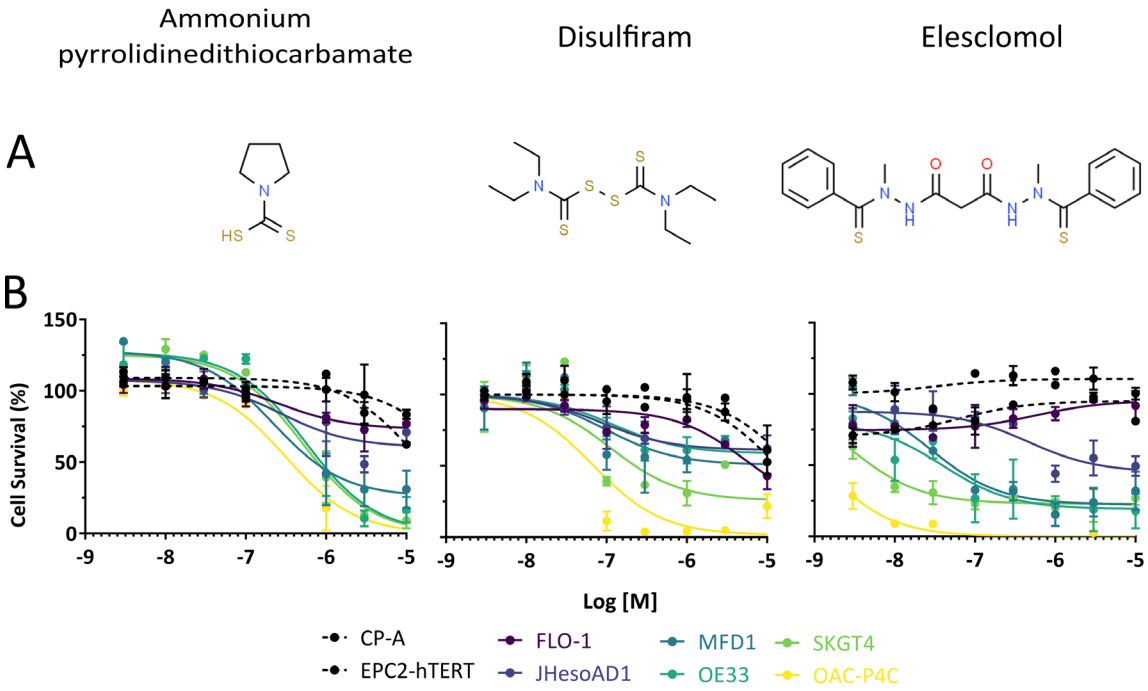



**Figure 5.1: Dose responses for ammonium pyrrolidinedithiocarbamate, disulfiram and elesclomol.** A) Chemical structures B) Univariate validation dose responses across the primary screen cell panel.

**Table 5.1: Compound $IC_{50}$ values across panel of cell lines.**

| | $IC_{50}$ **(nM)** | | |
|---|---|---|---|
| Cell Line | Elesclomol | Disulfiram | Ammonium pyrrolidinedithiocarbamate |
| CP-A | >10,000 | >10,000 | >10,000 |
| EPC2-hTERT | >10,000 | >10,000 | >10,000 |
| FLO-1 | >10,000 | >10,000 | >10,000 |

| | | | |
|---|---|---|---|
| JH-EsoAd1 | 3582 | >10,000 | >10,000 |
| MFD-1 | 51 | 868 | 1463 |
| OE33 | 39 | >10,000 | 1317 |
| SK-GT-4 | 5 | 197 | 915 |
| OAC-P4C | 1 | 69 | 880 |

Using the phenotypic information from the concentration responses to build multivariate dose responses, we can gain a deeper understanding of compound activity, mechanism of action, cellular response and selectivity over tissue-matched control cells. The phenotypic dose response data shows that both disulfiram and ammonium pyrrolidinedithiocarbamate move from phenotypically inactive at low concentrations, clustering with the DMSO, to phenotypically active at higher concentrations (Figure 5.2). However, they does not cluster with any of the classes of compounds in the reference library, suggesting a distinct mechanism of action. Elesclomol is phenotypically active at all concentrations tested, clustering in a similar location of multidimensional phenotypic space to disulfiram and ammonium pyrrolidinedithiocarbamate. Interestingly, the sensitivity profile of the cell panel is the same across both the univariate and multivariate dose responses with OAC-P4C showing the strongest responses and FLO-1 showing the least, to the extent that there were no morphological effects in the more resistant OAC cell lines FLO-1 and JH-EsoAd1.

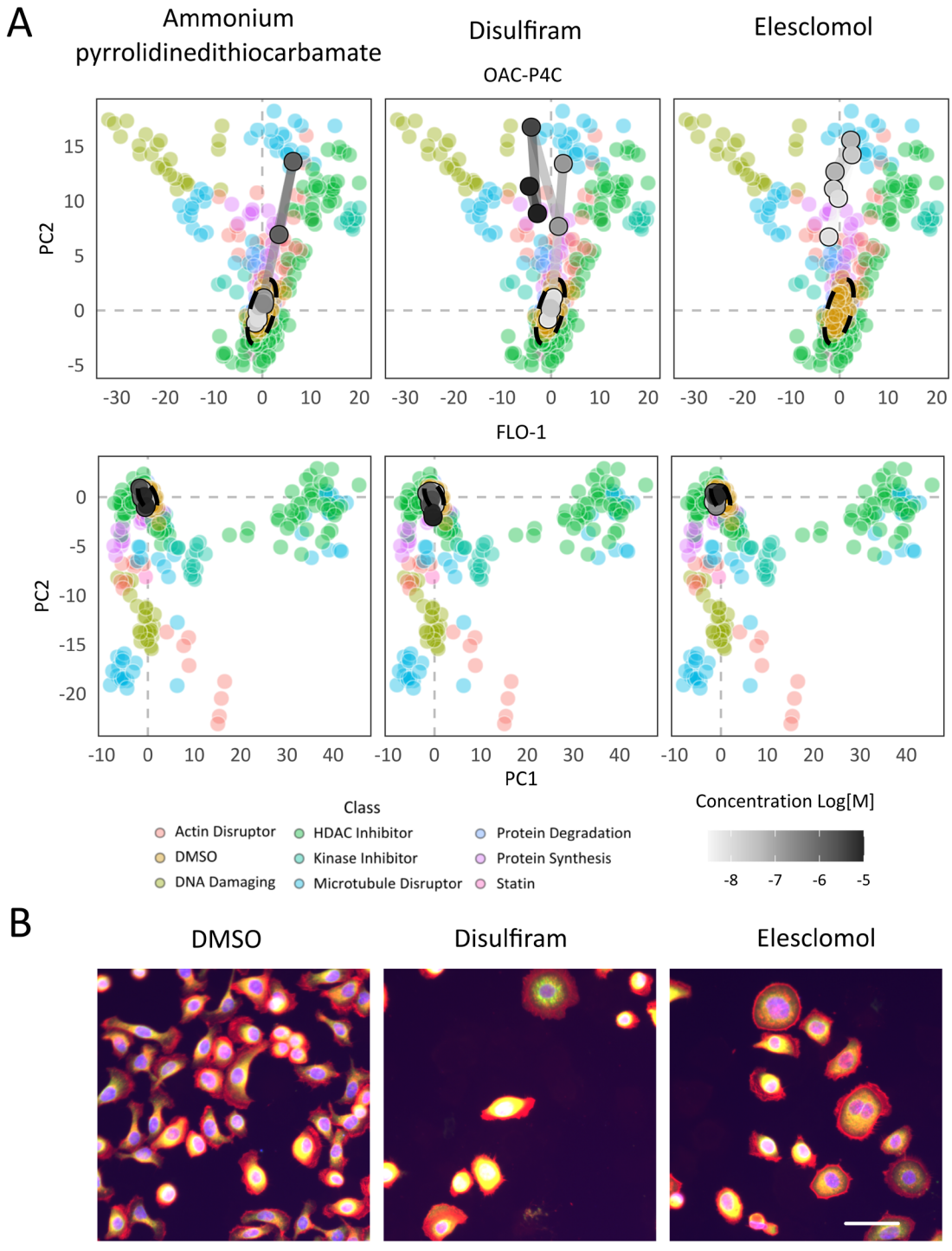



**Figure 5.2: Phenotypic dose responses for ammonium pyrrolidinedithiocarbamate, disulfiram and elesclomol.** A) First two principal components of multivariate phenotypic dose responses in OAC-P4C (most

sensitive) and FLO-1 (least sensitive) overlaid on the reference library of compounds B) Colour combined images for DMSO, disulfiram (100 nM) and elesclomol (3 nM) treatments in the OAC-P4C cell line. Hoescht 33342-nuclei (blue); Phalloidin 594, Wheat germ agglutinin and Alexa Fluor 594- F-actin, golgi and plasma membrane (red); Concanavalin A Alexa Fluor 488- Endoplasmic reticulum (green). Scale bars 50 μm.

Patient derived 3D organoids are thought to better recapitulate the morphology, genomic and transcriptomic landscape of the primary disease. For this reason I worked in collaboration with Professor Rebecca Fitzgerald's group at the MRC Cancer Unit, University of Cambridge, to test patient derived 3D oesophageal adenocarcinoma organoids (Li *et al.*, 2018) against the most potent and selective hit; elesclomol (lab work carried out by Xiaodun Li, University of Cambridge).

Elesclomol showed a similar potency in the organoids, with $IC_{50}$ values in the low nanomolar range (Table 5.2), demonstrating elesclomol is equally active in oesophageal 2D and 3D cultures, as well as across cell lines and patient derived organoids (Figure 5.3). Elesclomol was also highly selective, killing the tumour derived organoids and not the normal gastric organoid NG088 (Figure 5.3B). Overall, both 2D and 3D cell viability studies suggest that these compounds are highly selective and highly potent towards OAC.

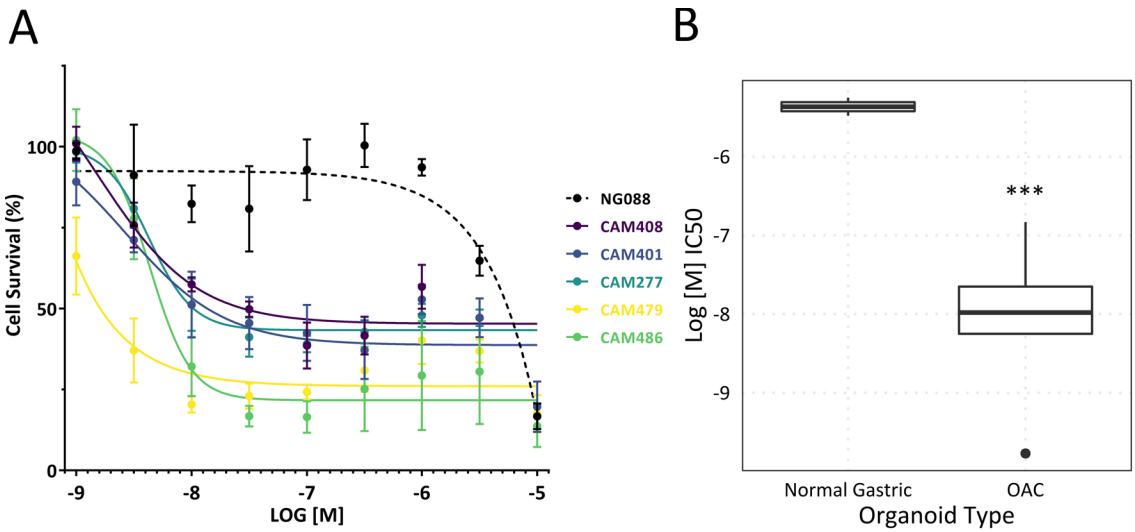



**Figure 5.3: Patient derived organoid dose responses for elesclomol.** A) Elesclomol dose response curves. NG088 - normal gastric control organoid. B) Comparison of $IC_{50}$ of normal gastric to OAC organoids, *** $p < 0.005$. Midline is the median, upper and lower limits are the third and first quartile, whiskers extend 1.5 times the interquartile range and outliers are plotted. Data provided by Xiaodun Li.

**Table 5.2: Patient derived organoids.** Elesclomol $IC_{50}$ values. Lab work carried out by Xiaodun Li, University of Cambridge.

| | | **$IC_{50}$** | |
|---|---|---|---|
| Cell Line | Type | Log [M] | nM |
| NG088 | Normal Gastric | -5.36 | 4325 |
| CAM401 | OAC | -6.83 | 147 |
| CAM408 | OAC | -7.65 | 22 |
| CAM277 | OAC | -7.98 | 10 |
| CAM468 | OAC | -8.25 | 5.6 |
| CAM479 | OAC | -9.77 | 0.17 |

### 5.2.2 Elesclomol, disulfiram and ammonium pyrrolidinedithiocarbamate act through a unified mechanism of action

As described earlier, previous studies indicate that these three compounds have very different targets and therefore likely that they act via multiple distinct mechanisms. However studying our validation data I noticed two things. Firstly, the pattern of sensitivity in both the univariate and multivariate dose responses across all three compounds demonstrates a shared sensitivity profile across the panel of OAC cell lines, from OAC-P4C the most sensitive to FLO-1 the least sensitive. Secondly, all three compounds induce a shared morphological phenotype. Thus I reasoned, that these three compounds were likely acting via the same mechanism in OAC cells, accounting for their shared sensitivity and activity profiles across the panel of cell lines.

To quantify this I assessed the correlation of $IC_{50}$ values for disulfiram and elesclomol using the CellTitre Glo assay across an expanded panel of OAC cell lines (data provided by Richard Elliott, University of Edinburgh). Pearson Correlation of $IC_{50}$ values confirms that the compounds have a shared sensitivity profile across OAC cell lines (Pearson 0.94, p-value <0.001) (Figure 5.4). Similar analysis expanding the cell lines across 373 pan cancer cell lines from the Cancer Dependency Map (DepMap) also demonstrated a significant correlation (Pearson 0.34, p-value <0.001) (Figure 5.4B). Combined, this suggests that the compounds do act through a shared mechanism of action.

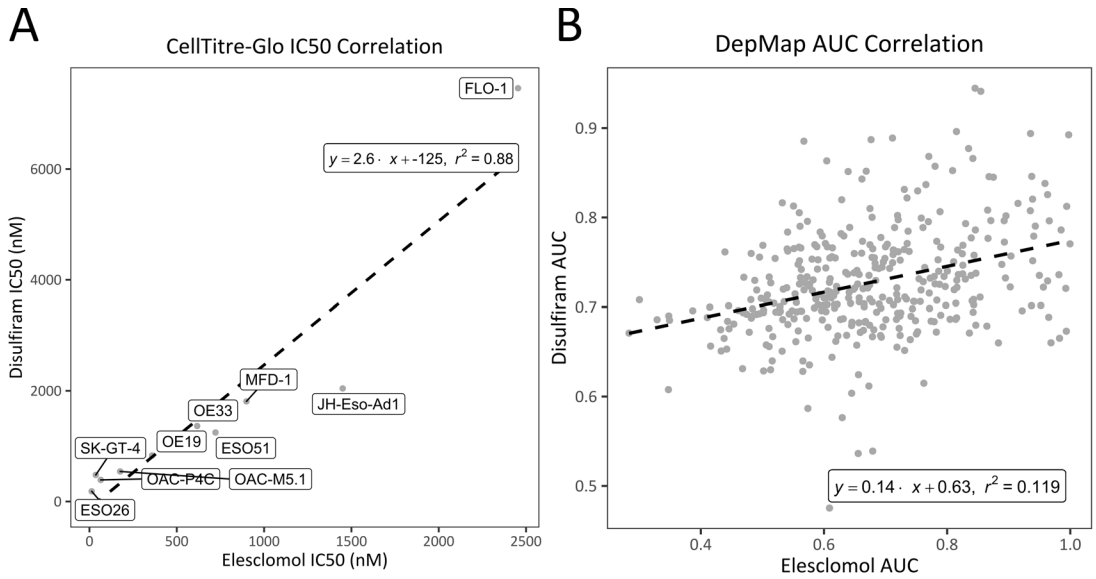



**Figure 5.4: Correlation of disulfiram and elesclomol $IC_{50}$s.** A) Correlation of disulfiram and elesclomol $IC_{50}$s across expanded cell panel using CellTiter-Glo assay (Data from Richard Elliott). B) Pan Cancer correlation of area under curve. Data from DepMap PRISM Repurposing Secondary Screen 19Q4, n=373.

Ammonium pyrrolidinedithiocarbamate is the least active of the three compounds and so was excluded from further analysis, though from the data presented thus far I believe it to act via the same mechanism.

To further confirm a shared mechanism I sought to use gene expression profiling in a compound exposure setting to generate compound induced gene expression signatures. Comparison of signatures can be used to discover new connections among compounds as well as to identify targets and mechanisms of action in a more complex setting than single gene contributions. Here I used gene expression profiling to compare the signatures of disulfiram and elesclomol and identify similarities in gene expression changes across the OAC cell panel. I used the NanoString nCounter platform to quantify the expression of nearly 1500 genes and compared the log2 fold change data for elesclomol (200 nM) and disulfiram (600 nM) treatments. A correlation heatmap of cell lines and treatments shows that the compound induced signatures are more similar within a given cell line and that

the OAC lines form a distinct cluster to the tissue-matched control cell lines (CP-A and EPC2-hTERT) (Figure 5.5A). Pearson correlation and PCA shows there is a very strong relationship between the two treatments within a cell line (Figure 5.5 B and C), demonstrating that disulfiram and elesclomol induce the same gene expression signature. Focusing on the individual cell lines there was a strong and significant correlation between disulfiram and elesclomol induced gene expression changes in the two most sensitive cell lines OAC-P4C and SK-GT-4 (Pearson Correlation 0.93 and 0.95 respectively), the intermediate cell line JH-EsoAd1 also shows strong correlation (0.88) and the resistant line FlO-1 exhibits lower correlation (0.59). It is also apparent that the magnitude of the gene expression changes decreased with decreasing sensitivity to the compounds from OAC-P4C to FLO-1.

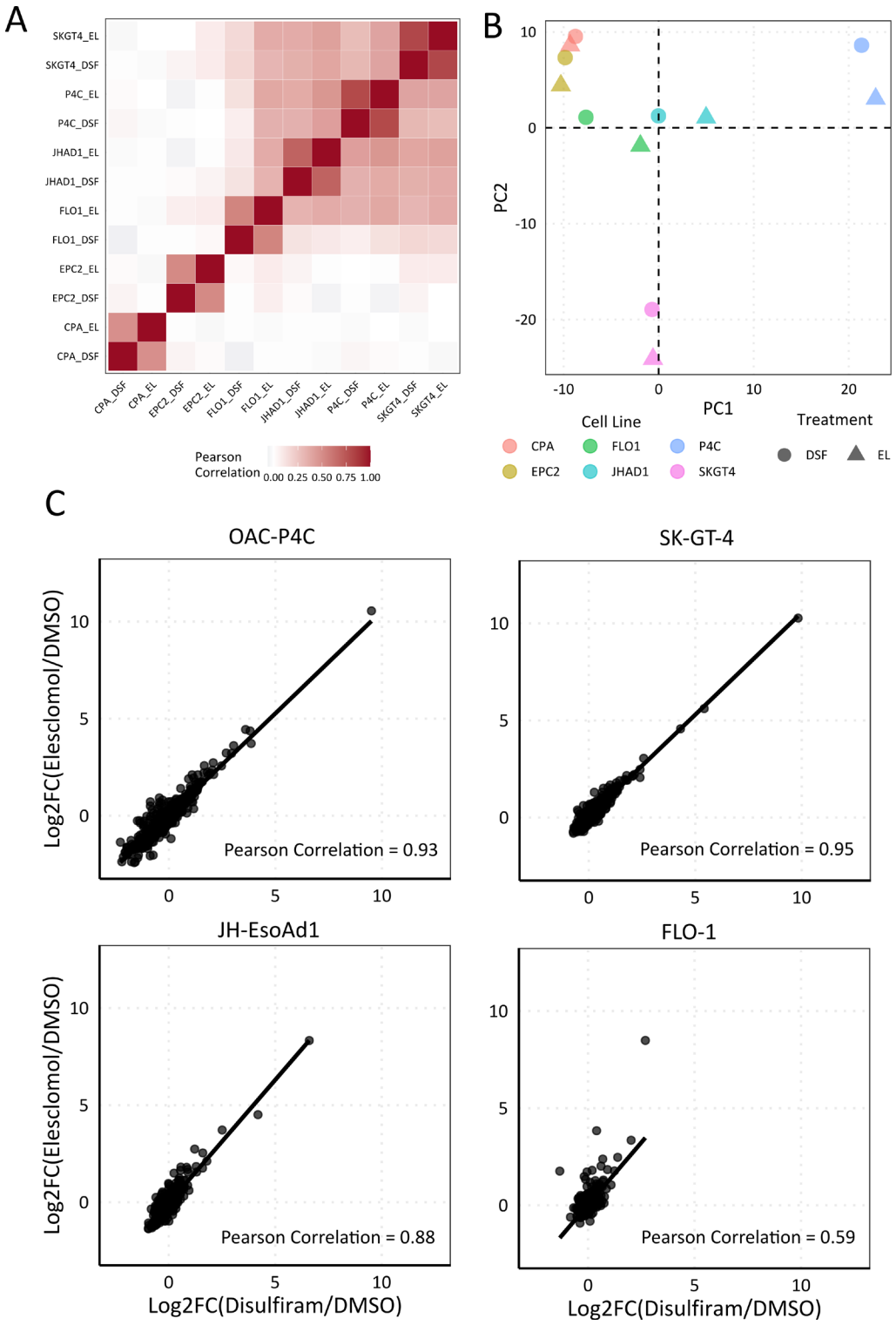

A
SKGT4_EL
SKGT4_DSF
P4C_EL
P4C_DSF
JHAD1_EL
JHAD1_DSF
FLO1_EL
FLO1_DSF
EPC2_EL
EPC2_DSF
CPA_EL
CPA_DSF
Pearson Correlation 0.00 0.25 0.50 0.75 1.00
B
PC2
PC1
Cell Line
CPA
FLO1
P4C
EPC2
JHAD1
SKGT4
Treatment
DSF
EL
C
OAC-P4C
SK-GT-4
JH-EsoAd1
FLO-1
Log2FC(Elesclomol/DMSO)
Log2FC(Disulfiram/DMSO)
Pearson Correlation = 0.93
Pearson Correlation = 0.95
Pearson Correlation = 0.88
Pearson Correlation = 0.59

**Figure 5.5: Disulfiram and elesclomol gene expression changes.** A) Correlation heatmap of log2 fold gene expression changes across both OAC and tissue-matched controls. B) First two principal components of average induced gene expression changes coloured by cell line. DSF - Disulfiram, EL - Elesclomol. C) Correlation of gene expression changes in individual OAC cell lines. n = 3.

It is likely that both compounds have many, possibly shared, off-target mechanisms as evidenced by the body of literature on their activity. However, despite numerous and conflicting targets/mechanisms in the literature, the data presented here shows the compounds share a sensitivity profile and produce almost identical gene expression signatures in OAC cells, providing strong evidence for a shared on-target mechanism for OAC cell death at nanomolar concentrations. Consistent with our findings, comparison of the elesclomol gene expression signature in the Connectivity Map (Lamb *et al.*, 2006; Subramanian *et al.*, 2017) database revealed pyrrolidine-dithiocarbamate to be the most related compound with a similarity score of 99.65 (see Methods 2.1.5 Publically available data for connectivity score explanation).

Since Disulfiram is a well-defined inhibitor of ALDH, three ALDH inhibitors were tested for cytotoxicity across oesophageal cell lines in order to confirm that the effects of disulfiram were not via ALDH inhibition (work carried out by Richard Elliott, University of Edinburgh). CVT-10216, a potent and selective reversible inhibitor of ALDH2 (mitochondrial ALDH), and two potent ALDH1A1 inhibitors (NCT-501 and A37) were tested, none of which had any effect on cell viability in either the tissue matched control or OAC cell lines. We therefore do not believe that disulfiram is acting via its primary annotated target of ALDH2 inhibition.

### 5.2.3 Mechanism of action

The mechanism of action of a compound is defined as the specific biochemical interaction of a set of target and effector proteins through which a drug substance produces its pharmacological effect in a specific cellular context.

## Copper and compound activity

The ability of all three compounds to chelate metals and in particular copper is a unifying link common to all three compounds. I sought to identify its role in the cytotoxic activity of these compounds in OAC. Nagi *et al*. demonstrated that copper is essential for the cytotoxic activity of elesclomol in the melanoma cell line MDA-MB435 and the leukemic cell line HL-60, and that it accumulates within cancer cells specifically (Nagai *et al.*, 2012). In order to confirm that this is also true for oesophageal cancer and that this was not limited to elesclomol but also true for disulfiram I pre-incubated cells with bathocuproinedisulfonic acid (BCS) to remove free copper from the media before compound treatment. This led to complete loss of activity for both disulfiram and elesclomol (Figure 5.6) confirming free extracellular copper is necessary for the cytotoxic activity of both compounds. The iron chelators Deferoxamine (data not shown) and ciclopirox olamine (experiment carried out by Kenji Fujihara, Peter MacCallum Cancer Centre) were unable to rescue cell death, demonstrating the specificity for copper (Figure A.1).

Since this showed that extracellular copper was vital for the activity of these compounds I next investigated the intracellular copper levels both basal and post-treatment using inductively coupled plasma mass spectrometry (ICP-MS) to detect metal ions within cell extracts. Here I found that both disulfiram and elesclomol led to the accumulation of intracellular copper in OAC cells (Figure 5.6B), consistent with previous findings (Nagai *et al.*, 2012). Interestingly, both disulfiram and elesclomol induced significant accumulation of copper in all OAC lines regardless of sensitivity to elesclomol. Consistent with the selectivity for copper chelators, I also found that other metal ions; Fe, Mg, Mn, and Zn, did not increase with compound treatment (Figure A.2B for Fe levels and data not shown for other metal ions), again demonstrating the specificity for copper in their mechanism of action.

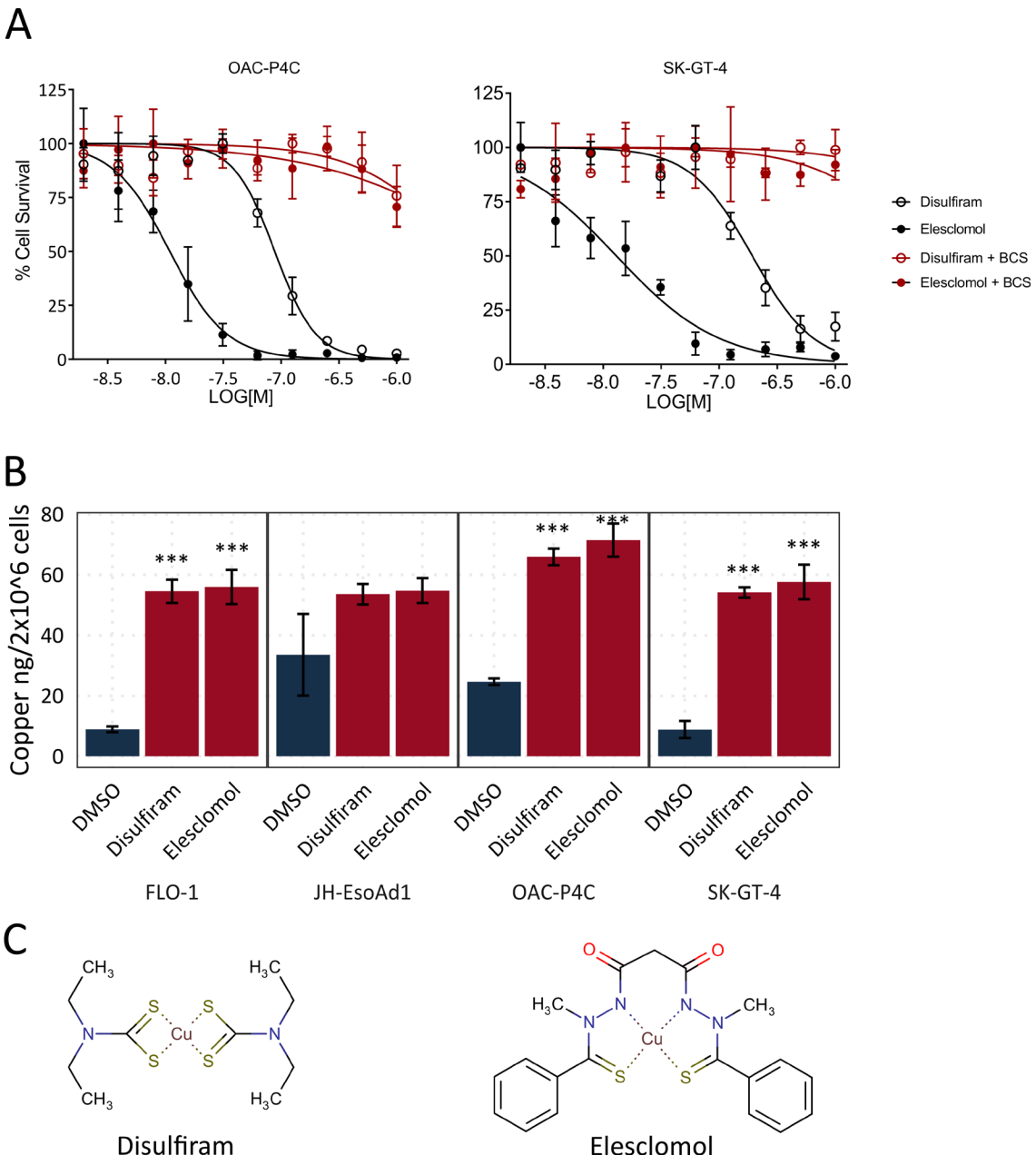



**Figure 5.6: Role of copper in compound activity.** A) Dose responses with and without copper chelator bathocuproinedisulfonic acid (BCS). B) ICP-MS intracellular copper levels. Error bars indicate SE. 1-ANOVA and Tukeys post hoc test for significance. Significance indicated for treatments compared to DMSO. *** $p<0.001$, n = 3. C) Structures of disulfiram and elesclomol as copper chelates.

## N-acetyl-L-cysteine and compound induced toxicity

Although the detailed mechanisms underlying the anticancer activity of these compounds are still not fully understood, several cell culture based studies have indicated that ROS generation may play an important role in their activity. This is thought to be through the redox activity of copper (Kirshner *et al.*, 2008; Barbi de Moura *et al.*, 2012; Nagai *et al.*, 2012; Hasinoff *et al.*, 2014). A major piece of evidence supporting this claim is that pre-incubation of cells with N-acetyl-L-cysteine (NAC) eliminates the cytotoxic effects. Because NAC is a ROS antagonist and precursor to GSH, it has been proposed that the mechanism is due to alleviation of oxidative stress caused by disulfiram and elesclomol. Thus these results have been interpreted to support that induction of ROS may underlie the mechanism by which these compounds exert their cytotoxic activity in cancer cells. However, it is also known that NAC can form conjugates with thiol-reactive compounds (Mi *et al.*, 2010) and is capable of chelating copper (Zheng *et al.*, 2010). I therefore sought to confirm the mechanism by which NAC rescues disulfiram and elesclomol induced cell death, and thereby clarify the roll of ROS in their mechanism.

Firstly, I confirmed that pre-incubation with NAC prevents both disulfiram and elesclomol induced toxicity in OAC cells. Pre-incubation with NAC (10 mM) completely removed the disulfiram and elesclomol dose dependent reduction in cell count in the sensitive cell lines OAC-P4C and SK-GT-4 (Figure 5.7). Conversely, pre-incubation with the antioxidant Trolox has no affect at 1 or 5 mM but is highly toxic at 10 mM (data not shown). Given that NAC may interfere with either free extracellular copper levels or disulfiram and elesclomol itself, I sought to determine the effects of NAC by quantifying the intracellular copper levels induced by elesclomol treatment with and without NAC pre-incubation. ICP-MS revealed that NAC pre-incubation led to a loss of copper accumulation post elesclomol treatment, and resulted in copper levels comparable with those of the untreated cells (Figure 5.7B). Therefore NAC pre-incubation appears to prevent compound

activity and cell death upstream of ROS antagonism, by preventing the increase in intracellular copper levels.

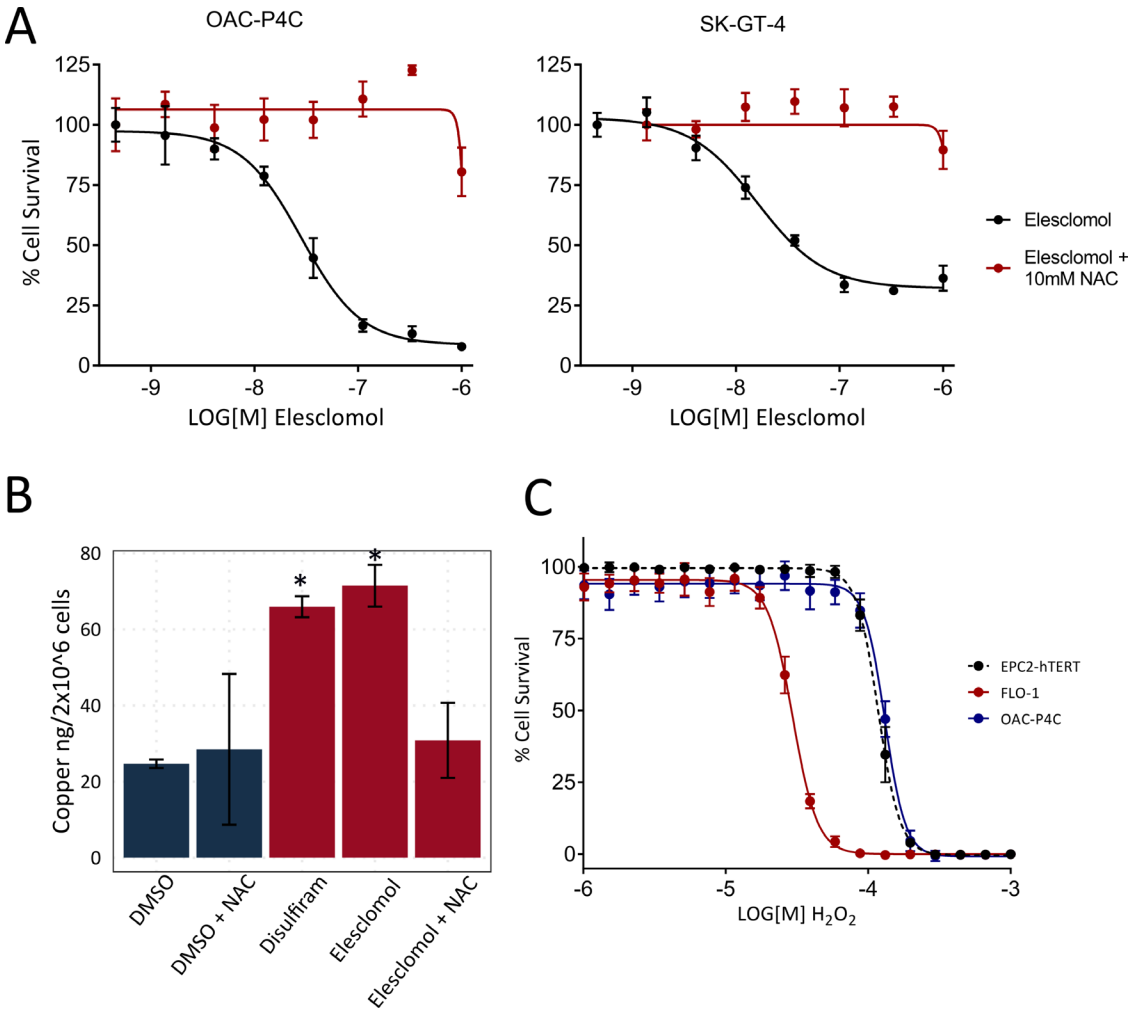



**Figure 5.7: Role of reactive oxygen species in compound activity.** A) Elesclomol dose response with and without N-acetyl-L-cysteine (NAC) pre-treatment * $P < 0.05$ (data provided by Richard Elliott, University of Edinburgh). B) ICP-MS intracellular copper levels in the OAC cell line OAC-P4C with and without 30 min, 10 µM NAC pre-treatment. C) $H_2O_2$ dose responses for OAC cell lines OAC-P4C and FLO-1, and tissue matched control EPC2-hTERT (data provided by Richard Elliott, University of Edinburgh).

### Sensitivity to oxidative stress

While the NAC experiments revealed that compound activity is not conclusively via ROS induction, they do not disprove it either. It is still possible that the increase in intracellular copper with elesclomol treatment exerts its effects through the induction of ROS. To further investigate the role of ROS in disulfiram and elesclomol activity I compared the OAC cell lines sensitivity to $H_2O_2$ treatment. If ROS levels are responsible for the activity of disulfiram and elesclomol, then we would expect the sensitivity of disulfiram and elesclomol to correlate with $H_2O_2$ sensitivity. However, I found that FLO-1 cells (insensitive to disulfiram and elesclomol, Table 5.1) are significantly more sensitive to oxidative stress via $H_2O_2$ than either OAC-P4C or the tissue-matched control EPC2-hTERT (Figure 5.7C). This fits with previous findings that FLO-1 have a lower antioxidant capacity and increased sensitivity to ROS inducing compounds (Liu *et al.*, 2017). This does not rule out a role for ROS or oxidative stress in the mechanism of disulfiram and elesclomol, but it does suggest that it is not simply sensitivity to high levels of ROS that results in cell death as previously suggested (Kirshner *et al.*, 2008; Blackman *et al.*, 2012; Nagai *et al.*, 2012).

### Elesclomol treatment and mitochondrial function

A large randomised phase III trial of elesclomol conducted in patients with metastatic melanoma suggested that serum lactate dehydrogenase (LDH) levels may act as a predictive biomarker for response (Hauschild *et al.*, 2009). They found that only in patients with metastatic melanoma and normal to low serum LDH was the treatment effective at prolonging median progression free survival, suggesting metabolism may play a role in elesclomol's mechanism. Furthermore, it has been found that elesclomol disrupts mitochondrial function and suppresses oxidative phosphorylation in melanoma cell lines (Barbi de Moura *et al.*, 2012). I therefore used the Seahorse Cell Mito Stress test to evaluate elesclomol's effects on metabolism in OAC.

The Seahorse Cell Mito Stress test directly measures oxygen consumption rate to assess mitochondrial function. The assay reports multiplexed key parameters using the compounds; oligomycin, trifluoromethoxy carbonylcyanide phenylhydrazone (FCCP), rotenone, and antimycin A, to measure basal respiration and then ATP-linked respiration, maximal respiration, and non-mitochondrial respiration, respectively. Proton leak and spare respiratory capacity can then be calculated using these parameters (Figure 2.2 for seahorse trace and measurement explanation).

Here I investigated the effects of elesclomol on respiratory function of an elesclomol-sensitive and an -insensitive cell line (SK-GT-4 and FLO-1 respectively). Elesclomol treatment reduced mitochondrial function in both cell lines, indicated by reduced maximal respiration and spare respiratory capacity (Figure 5.8). This provides broad evidence that elesclomol causes interference in mitochondrial respiratory function. In the sensitive cell line only, exposure to elesclomol appeared to eliminate the spare respiratory capacity almost entirely and proton leak was also significantly increased (Figure 5.8C), demonstrating more significant effects on respiration in sensitive cells.

These results suggest that there may be a link between elesclomol treatment and loss of spare respiratory capacity and/or increased proton leak. By contrast, there is no clear effect of elesclomol on ATP-linked respiration in either cell line, and while there is an inhibition of maximal respiration, this effect is similar for both sensitive and insensitive cell lines. Interestingly, elesclomol did not have an effect on basal respiration after the 6 hour incubation at 50 or 200 nM in either cell line. This is in disagreement with experiments in melanoma that revealed a decrease in basal respiration after elesclomol treatment (Barbi de Moura *et al.*, 2012).

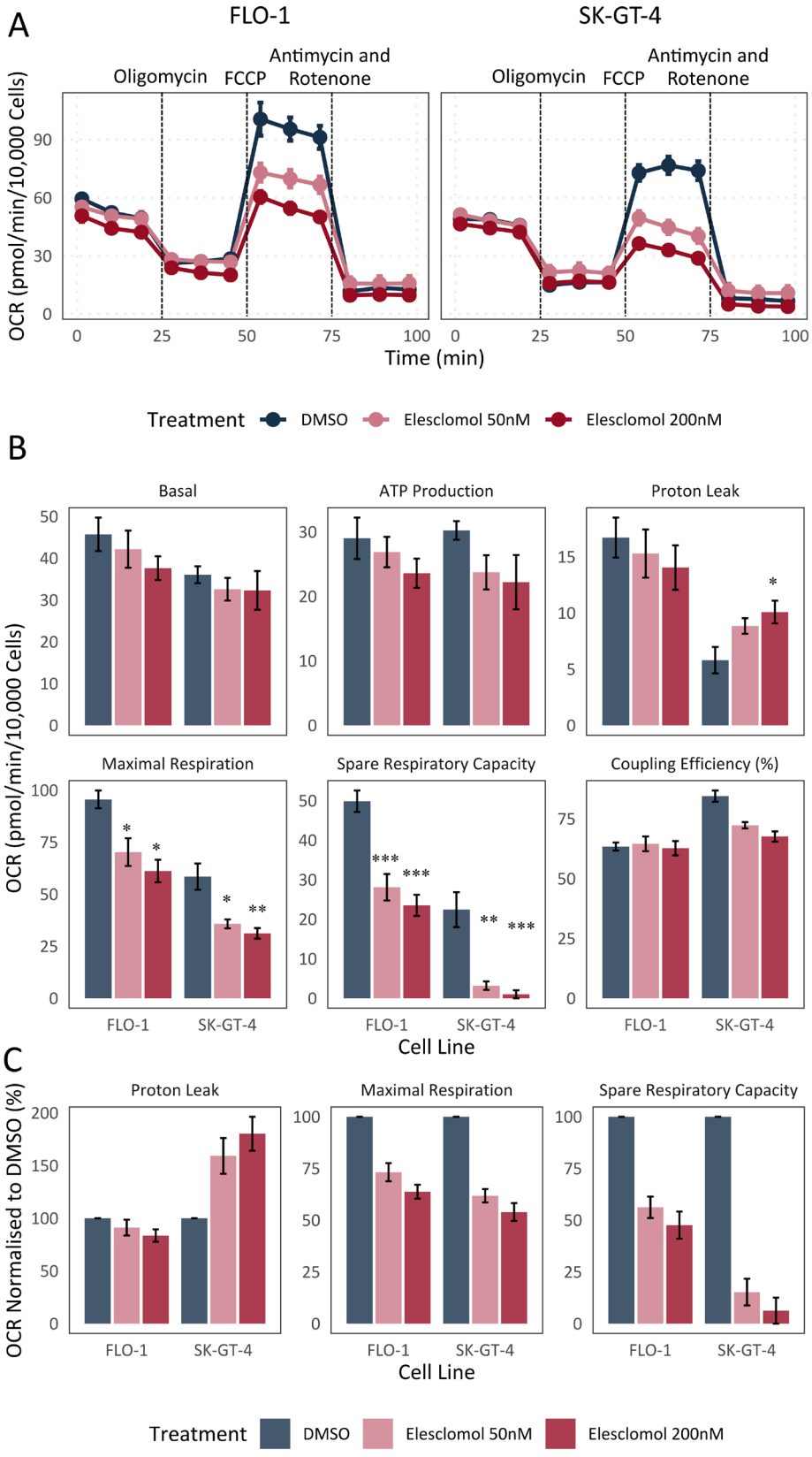

A
FLO-1
SK-GT-4
Oligomycin
FCCP
Antimycin and Rotenone
OCR (pmol/min/10,000 Cells)
Time (min)
Treatment
DMSO
Elesclomol 50nM
Elesclomol 200nM
B
Basal
ATP Production
Proton Leak
Maximal Respiration
Spare Respiratory Capacity
Coupling Efficiency (%)
Cell Line
C
Proton Leak
Maximal Respiration
Spare Respiratory Capacity
OCR Normalised to DMSO (%)

**Figure 5.8: Effect of elesclomol treatment on mitochondrial function.** A) Exemplar oxygen consumption rate (OCR) traces for resistant and sensitive cell lines when treated with varying concentrations of elesclomol for 6 hrs prior to assay. B) Parameters for each cell line normalised to 10,000 cells n=3. Error bars indicate SEM. 1-ANOVA and Tukeys post hoc test to test significance. Significance indicated for treatments compared to DMSO. * P<0.05, ** p<0.01, *** p<0.005. C) Significant parameters from B normalised to DMSO values for comparison of elesclomol effects between cell lines.

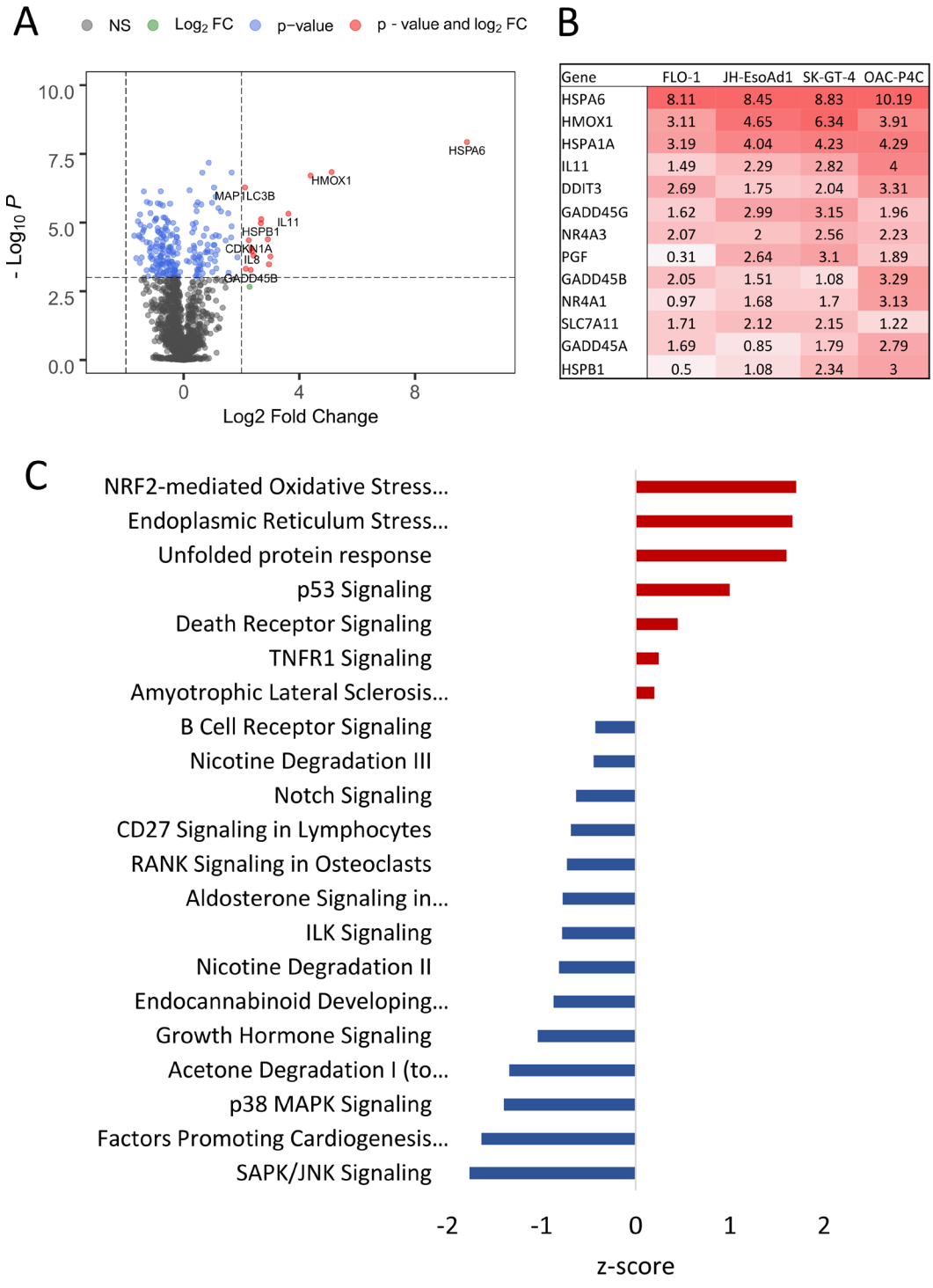
The mechanism through which elesclomol exerts its effects on mitochondrial respiration remain to be fully elucidated and whether these effects are the cause of elesclomol induced toxicity or not remain to be confirmed. However, it is my hypothesis that unless elesclomol affects basal or ATP-linked respiration at longer incubation periods, then this is unlikely the cause of cell death since both 50 and 200 nM are above the $IC_{50}$ in the SK-GT-4 cell line. It is also more likely that effects observed at higher concentrations or longer incubation periods maybe a consequence of cell death rather than the cause.

### Compound induced gene expression profiles

In a more unbiased approach to studying the mechanism of action of elesclomol, I focused on the transcriptomic profile of elesclomol treatment in OAC. As described previously I have used the NanoString nCounter platform to determine a shared mechanism of action between disulfiram and elesclomol. To further elucidate the mechanism of action of elesclomol at the transcriptomic level I carried out differential expression analysis (Wang *et al.*, 2016) on the gene expression changes induced following a 6 hour treatment with elesclomol (200nM). Differential expression analysis revealed strong and consistent activation of the heat shock response (HSR) and the unfolded protein response (UPR). Specifically heat shock protein 70 (HSP70) subunits, HMOX1, GADD45B/G, and DDIT3 were upregulated (Figure 5.9). These genes were induced in all four OAC cell lines, but the fold change (FC) tended to be larger in the more sensitive cell lines (Figure 5.9B). Of note, most of the gene expression changes involved an increase, not a decrease in gene expression level (Figure 5.9). A similar result was found in a recent study of

copper induced protein expression changes in *E.coli* (Wiebelhaus *et al.*, 2020), possibly suggesting elesclomol treatment results in a copper exposure gene signature.

Pathway analysis integrates gene expression data with databases of known biologically relevant pathways to identify changes associated with treatment conditions at a higher level than individual genes, giving a more holistic view of the impact of treatment conditions. Here I used Ingenuity Pathway Analysis (IPA) to define deregulated pathways in a compound exposure setting to elucidate mechanism of action (see Methods 2.1.4 Data analysis for details). Pathway analysis of the elesclomol induced gene expression signature highlights SAPK/JNK signalling, NRF2-mediated Oxidative Stress Response, unfolded protein response (UPR), endoplasmic reticulum (ER) stress, and p53 signalling as the top five altered pathways after elesclomol treatment. NRF2-mediated Oxidative Stress Response, unfolded protein response (UPR), endoplasmic reticulum (ER) stress, and p53 signalling are all upregulated. Similarly, recent microarray analysis of copper induced gene expression has revealed HSR, UPR and ubiquitin/proteasomal degradation as the top upregulated pathways (Saporito-Magriñá *et al.*, 2018), suggesting the elesclomol gene expression signature is most heavily influenced by the general presence of excess copper rather than the compound itself.




| Gene | FLO-1 | JH-EsoAd1 | SK-GT-4 | OAC-P4C |
|---|---|---|---|---|
| HSPA6 | 8.11 | 8.45 | 8.83 | 10.19 |
| HMOX1 | 3.11 | 4.65 | 6.34 | 3.91 |
| HSPA1A | 3.19 | 4.04 | 4.23 | 4.29 |
| IL11 | 1.49 | 2.29 | 2.82 | 4 |
| DDIT3 | 2.69 | 1.75 | 2.04 | 3.31 |
| GADD45G | 1.62 | 2.99 | 3.15 | 1.96 |
| NR4A3 | 2.07 | 2 | 2.56 | 2.23 |
| PGF | 0.31 | 2.64 | 3.1 | 1.89 |
| GADD45B | 2.05 | 1.51 | 1.08 | 3.29 |
| NR4A1 | 0.97 | 1.68 | 1.7 | 3.13 |
| SLC7A11 | 1.71 | 2.12 | 2.15 | 1.22 |
| GADD45A | 1.69 | 0.85 | 1.79 | 2.79 |
| HSPB1 | 0.5 | 1.08 | 2.34 | 3 |

**Figure 5.9: Elesclomol induced gene expression signature.** A) Differential expression of 6 h DMSO and elesclomol (200 nM) treatment in the two most sensitive cell lines OAC-P4C and SK-GT-4. P value cut off equivalent to adjusted p

value 0.05. n = 3. (Table A.3 for list of significant genes) B) Log2 fold change for top significant genes C) Top 30 (by p-value) pathway Z-scores for pathways identified in IPA analysis in the sensitive cell lines (OAC-P4C and SK-GT-4).

A more in-depth analysis of these pathways and comparison with the microarray analysis of copper overload (Saporito-Magriñá *et al.*, 2018) revealed further similarities. Firstly, assessment of the enzymes involved in maintenance of redox homeostasis did not reveal any major induction of antioxidant enzymes (Figure 5.10). In fact, $H_2O_2$ sink enzymes glutathione peroxidase 1 and 4 (GPX1, GPX4), and catalase (CAT) are transcriptionally downregulated (Figure 5.10**A**), reflecting the results from the copper induced microarray analysis (Saporito-Magriñá *et al.*, 2018). Conversely, several aspects of thiol (–SH) homeostasis are upregulated. This suggests that while the antioxidant response is not overwhelmed by elesclomol treatment, there are perhaps higher levels of protein oxidation occurring. Secondly, further similarities to copper treatment can be found in the strong induction of the HSR and the UPR, known to arise from increased levels of misfolded proteins in the cytosol and ER respectively (Taylor, Berendzen and Dillin, 2014). Perhaps surprisingly given that elesclomol has been shown to cause copper to accumulate selectively within the mitochondria of melanoma cells (Nagai *et al.*, 2012), a recently proposed target, FDX1, is a mitochondrial protein (Tsvetkov *et al.*, 2019), and links with metabolism and mitochondrial electron transport (Blackman *et al.*, 2012), the mitochondrial UPR was not induced in this study of elesclomol in OAC cells. Again reflecting the previous study of copper induced toxicity (Saporito-Magriñá *et al.*, 2018) (Figure 5.10), and suggesting the response is not confined to the mitochondria, consistent with our Seahorse findings above.

A

| | Gene | FLO-1 | JH-EsoAD1 | SK-GT-4 | OAC-P4C |
|---|---|---|---|---|---|
| | | Cell Lines | | | |
| $O_2$ Sinks | SOD1 | 0.07 | 0.58 | 0.60 | 0.38 |
| | SOD2 | -0.15 | 0.14 | 0.12 | 0.35 |
| | SOD3 | 0.34 | 0.07 | 0.56 | 0.31 |
| $H_2O_2$ Sinks | CAT | 0.02 | -0.19 | -0.32 | -0.35 |
| | GPX4 | -0.03 | 0.01 | -0.26 | -0.27 |
| | GPX1 | -0.27 | 0.04 | -0.18 | -0.29 |
| | PRDX1 | -0.07 | 0.36 | 0.70 | 0.60 |
| | PRDX5 | 0.03 | 0.06 | -0.07 | -0.11 |
| -SH homeostasis | GCLC | 0.44 | 0.88 | 1.28 | 0.69 |
| | TXN | 0.00 | 0.27 | 0.30 | 0.44 |
| | TXN2 | -0.15 | -0.04 | -0.30 | -0.16 |
| | TXNRD1 | 0.49 | 1.57 | 1.84 | 1.51 |

B

| | Gene | FLO-1 | JH-EsoAD1 | SK-GT-4 | OAC-P4C |
|---|---|---|---|---|---|
| | | Cell Lines | | | |
| Heat Shock | HSPA6 | 7.95 | 8.60 | 8.97 | 10.47 |
| | HSPA1A | 2.77 | 4.16 | 4.26 | 4.40 |
| | HSPB1 | 0.09 | 1.16 | 2.45 | 3.40 |
| mtUPR | HSPE1 | -0.08 | 0.60 | 1.20 | 0.91 |
| | NDUFB2 | 0.03 | -0.05 | -0.16 | -0.05 |
| | TXN2 | -0.15 | -0.04 | -0.30 | -0.16 |
| | ENDOG | -0.42 | -0.36 | -0.54 | -0.89 |
| UPR | ATF4 | 0.53 | 0.07 | 1.18 | 0.51 |
| | DDIT3 | 2.12 | 1.76 | 2.18 | 3.71 |
| | DDIT4 | 1.59 | 0.22 | 2.66 | 1.06 |

C

| | Gene | FLO-1 | JH-EsoAD1 | SK-GT-4 | OAC-P4C |
|---|---|---|---|---|---|
| | | Cell Lines | | | |
| Ubiquitin | UBB | -0.14 | 0.94 | 1.22 | 1.12 |
| Autophagy | MAP1LC3B | 0.54 | 1.58 | 2.01 | 2.23 |
| | GABARAP | 0.06 | 0.06 | -0.07 | -0.07 |

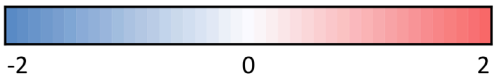



**Figure 5.10: Effect of elesclomol on highlighted cellular pathways.** Effect of elesclomol treatment on the transcriptional profile of A) Antioxidant enzymes B) Heat shock response (HSP) and unfolded protein response (UPR). C) Autophagy and proteasomal degradation. Log2 fold change values for gene expression across OAC cell lines using NanoString data. Elesclomol treatment was 200 nM for 6 h.

Taken together these finding correlate very well with previous studies on copper induced toxicity and suggest that the mechanism of elesclomol induced toxicity may be entirely down to the accumulation of excess copper within the cancer cells. Furthermore, the gene expression studies demonstrate that elesclomol treatment leads to strong induction of the HSR and the UPR, suggesting the mechanism is associated with these pathways. It also further strengthens my hypothesis that the cytotoxicity is not associated with ROS as previously thought due to the lack of $O_2$ and $H_2O_2$ sink genes.

### Comparison to proteasome inhibitor activity

In light of the above findings, I went back to study potential links to protein misfolding and proteostasis. Disulfiram and ammonium pyrrolidinedithiocarbamate have previously been shown to inhibit proteolytic protein degradation (Lövborg *et al.*, 2006). However, it is not known what the mechanism is and further the relevance of this to the selective toxicity of OAC is unknown.

I started by looking for similarities between the proteasome inhibitors in the reference library (MG132 and Lactacystin) in the class 'Protein Degradation' (Figure 5.11, green points) and disulfiram, elesclomol, and ammonium pyrrolidinedithiocarbamate. Phenotypically they do not cluster together, indicating none of these compounds induce a proteasome inhibitor like phenotype (Figure 5.11). This would suggest this is not their mechanism of action.

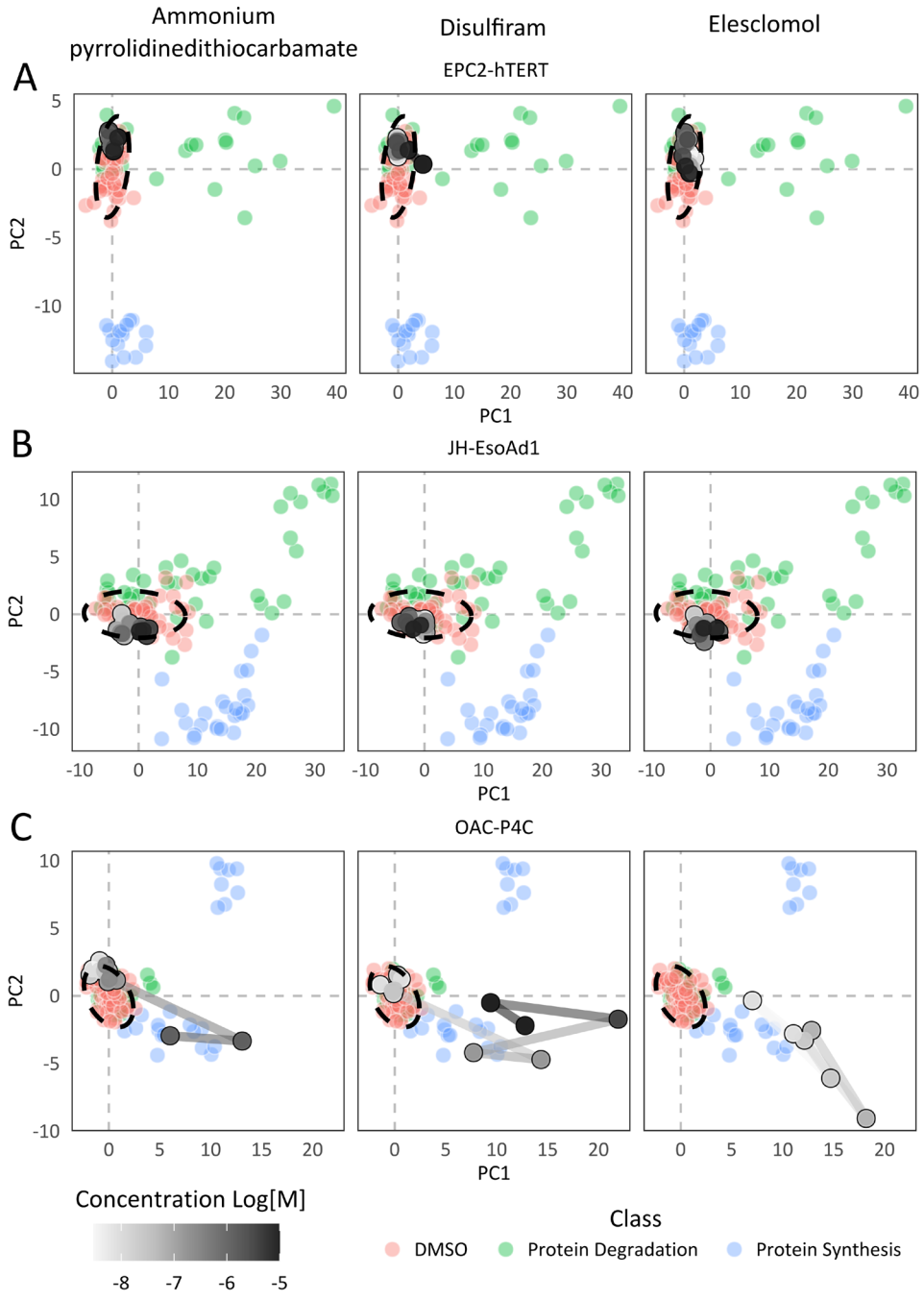



**Figure 5.11: Comparison of phenotypes induced by compounds affecting proteostasis.** First two principal components for A) Tissue matched non transformed cell line EPC2-hTERT. B) Elesclomol resistant cell line JH-EsoAd1. C) Elesclomol sensitive cell line OAC-P4C. Reference library coloured my mechanistic class, dose response coloured by concentration.

Furthermore, phenotypically the proteasome inhibitors in the reference library caused very limited changes in the OAC cell lines, with the exception of JH-EsoAD1, but caused strong phenotypic changes in the tissue-matched control cell lines, possibly suggesting increased proteasome inhibitor resistance in the majority of OAC cell lines (Figure 5.11). Interestingly, there is almost an opposing effect when comparing this to the strength of the induced phenotypes of disulfiram, elesclomol, and ammonium pyrrolidinedithiocarbamate, which caused strong phenotypic changes in the cell lines except for the FLO-1 and JH-EsoAd1 OAC lines and the tissue-matched control cell lines (Figure 5.11). This opposing effect has been proposed between bortezomib, a proteasome inhibitor, and elesclomol previously (Tsvetkov *et al.*, 2019). Tsvetkov *et al.* found that resistance to bortezomib, caused by loss of the 19S regulatory complex of the proteasome (Lo19S), confers a selective sensitivity to elesclomol. Further they found that elesclomol enhanced the effect of the proteasome inhibitor bortezomib, specifically in a proteasome inhibitor resistant Lo19S state. This suggests a link between the two groups of compounds despite the lack of phenotypic similarity.

To further explore this link I studied the gene expression signatures induced by bortezomib and elesclomol treatment (data from Tsvetkov et al., 2019). Unexpectedly given the lack of similarity in phenotypic response, the two gene expression signatures show significant correlation (Pearson correlation 0.49, $p < 0.001$) (Figure 5.12). One possible explanation, and my personal hypothesis, is that while not acting as proteasome inhibitors, disulfiram and elesclomol act via another mechanism to elicit the same downstream response e.g. an imbalance in proteostasis, resulting in a similar gene expression profile despite differing mechanisms and therefore phenotypes. This may also help explain why elesclomol has been shown to enhance the cytotoxicity of bortezomib, since both compounds would act in synergy to overwhelm proteostasis.

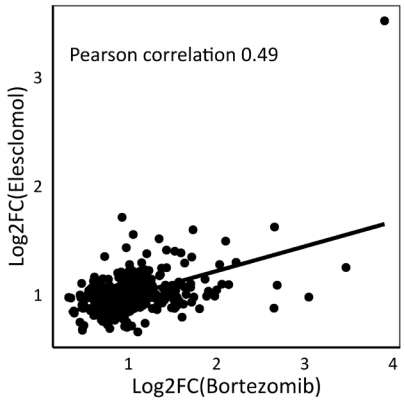



**Figure 5.12: Correlation of bortezomib and elesclomol induced gene expression changes.** Data from Tsvetkov et al., 2019.

## 5.3 Discussion

In summary, comparison of compound activity across a panel of distinct cell lines led to the hypothesis of a shared sensitivity and activity profile of three compounds; elesclomol, disulfiram and ammonium pyrrolidinedithiocarbamate. Further analysis revealed a strong correlation between the gene expression signatures of the compounds. In pursuit of identification of the mechanism responsible for their activity this work has confirmed that the mechanism is copper dependent, and leads to the accumulation of intracellular copper. Further, rescue of intracellular copper levels reverses the effects of the compounds, suggesting that intracellular copper accumulation is necessary for their activity and not an off-target consequence of the treatments. I believe that all cellular responses are likely entirely due to the accumulation of copper and the compounds just act as a shuttle, bypassing the copper regulatory mechanisms within a cell. This would explain why the induced gene expression signature is so similar to that of the very high levels of extracellular copper in other published work (Saporito-Magriñá *et al.*, 2018) (Figure 5.10).

It has previously been thought that NAC is able to rescue compound induced cell death due to relief of oxidative stress caused by the accumulation of ROS (Kirshner *et al.*, 2008; Blackman *et al.*, 2012; Nagai *et al.*, 2012). However, work carried out here demonstrates that NAC is actually interfering upstream of ROS induction, and

prevents the accumulation of intracellular copper. While this finding renders the role of ROS inconclusive, the finding that the most sensitive cell line (OAC-P4C) is more resistant to oxidative stress induction by $H_2O_2$ treatment than the elesclomol resistant cell line FLO-1, suggests that ROS accumulation may not fully explain the activity of these compounds. One possible explanation for the ability of NAC to inhibit elesclomol dependent intracellular copper accumulation is that NAC is thought to be capable of chelating copper, specifically Cu(II) (Offen *et al.*, 2004). A second possibility is that NAC has been shown to potently reduce Cu(II) to Cu(I) (Saporito-Magriñá *et al.*, 2018) and this may prevent the elesclomol-Cu chelate formation or prevent it from undergoing redox reactions within the cell and releasing the copper. Regarding disulfiram, disulfiram requires Cu(II) for its conversion to DETC (Kragh, 2008) so in the presence of Cu(I) it may not form copper chelates either, explaining its lack of activity.

Gene expression studies revealed strong induction of the heat shock response (HSR), the NRF2-mediated Oxidative Stress Response, unfolded protein response (UPR), and endoplasmic reticulum (ER) stress. Comparison to the proteasome inhibitor bortezomib demonstrated similarities in the gene expression profiles but not to the same extent as was evident between disulfiram and elesclomol. Furthermore, high content phenotypic fingerprints for the compounds do not match. Of note, disulfiram and elesclomol show opposite activity to bortezomib in the OAC cell lines, consistent with previous findings that the two have opposing sensitivities in cell lines (Tsvetkov *et al.*, 2019). These studies therefore indicate that the mechanism-of-action of disulfiram and elesclomol is distinct from that of the proteasome inhibitors but they do share similar downstream activity at a transcriptomic pathway level. One intriguing explanation for these findings could be that disulfiram and elesclomol somehow impair protein folding or enhance protein aggregation, possibly through the non-specific interaction of copper and intracellular proteins, leading to an imbalance in proteostasis. This would explain the pathways identified in the gene expression studies, the increase in expression of thiol homeostasis proteins, and the correlation of induced gene expression

signatures between elesclomol and bortezomib; since both inhibition of the proteasome and increased protein aggregation lead to the same imbalance in proteostasis. Since both compounds act via completely different mechanisms this could also explain why their induced phenotypes do not match. This also fits with studies of copper induced protein aggregation (Saporito-Magriñá *et al.*, 2018). This work highlights the value of multiparametric phenotypic profiling to differentiate distinct compound mechanisms of action which may not be clearly apparent at a transcriptomic level.

In contrast to modern molecular targeted therapies designed against specific proteins or enzyme families, defining the mechanism of action of disulfiram and elesclomol in cells represents a significant challenge. While defining these exact mechanisms requires further study, combining the work carried out here with previously published studies creates a more complete picture of their mechanism. I propose that all three compounds chelate copper in the extracellular environment, are able to cross the plasma membrane, react with unknown cellular targets leading to the reduction of Cu(II) to Cu(I) and its release into the cellular environment. The compound, once released from its complex is then able to efflux back out of the cell starting the cycle again, leading to an accumulation of copper over time. I believe the cytotoxic activity of the compounds is entirely due to the accumulation of copper. Finally, this accumulation of copper leads to upregulation of the HSR and the UPR, possibly through protein interactions and misfolding/aggregation, and it is this that eventually causes the cytotoxicity.

Many of the proposed effects of disulfiram, elesclomol and ammonium pyrrolidinedithiocarbamate are in relation to the mitochondria; the accumulation of copper specifically within the mitochondria (Nagai *et al.*, 2012), FDX1; a proposed target, is a mitochondrial protein (Tsvetkov *et al.*, 2019), and the effects of elesclomol on mitochondrial respiration in melanoma cells (Barbi de Moura *et al.*, 2012). However the mitochondrial function studies carried out here and the lack of induction of the mitochondrial UPR suggests this is not what causes cellular toxicity (at least at nanomolar concentrations).

Future investigation of FDX1 is essential to further understanding of the mechanism and sensitivity of disulfiram, elesclomol, and ammonium pyrrolidinedithiocarbamate. It has been shown that knockout of FDX1 increases resistance to elesclomol (Tsvetkov *et al.*, 2019), although it does not completely remove sensitivity and the expression of many other proteins have also been shown to effect disulfiram and elesclomol sensitivity. Perhaps FDX1 facilitates the reduction of the copper chelates from Cu(II) to Cu(I), through its redox activity, leading to the intracellular release of copper. Determining whether knockout of FDX1 leads to a loss or reduction in copper accumulation after elesclomol treatment could answer this question.

It has been shown previously that the redox potential of elesclomol is vital for the increase in intracellular copper and associated toxicity (Nagai *et al.*, 2012). It is therefore possible that FDX1 is just one such cellular target with a corresponding potential and there may be other proteins capable of interacting with the complex and releasing copper, explaining why there is only a reduction in sensitivity with FDX1 knockout and not complete resistance.

Molecular targeted therapies have made limited impact upon the treatment of many complex cancers such as OAC, and many other cancer subtypes which do not fall into clearly defined molecular subgroups. This lack of response is presumably due to the high degree of inter- and intra-patient tumour heterogeneity observed (Secrier *et al.*, 2016). It is therefore critical to identify a strategy define those patients which are most likely respond to targeted therapies.

Given the clear variability in OAC cell line sensitivity to disulfiram and elesclomol, understanding the differences accounting for this variability and identifying the mechanisms which confer sensitivity to these compounds is vital, and could lead to the identification of a biomarker for patient stratification. Secondly, understanding the lack of activity at both the phenotypic and transcriptomic levels explored in tissue-matched control cells could identify cancer specific pathways for future drug development programs. These ideas will be explored in the final chapter.

# *6| Pharmacogenomics of Disulfiram and Elesclomol*

Linking phenotype to genotype

## 6.1 Introduction

With the development of genome sequencing came the promise of functional genomics and the identification of gene and protein functions; genotype to phenotype. While we are still a long way from understanding the function of many genes, substantial progress has been made, owing to advances in next generation sequencing and gene editing technologies that enable the study of single genes in model organisms such as drosophila (Benfey and Mitchell-Olds, 2008) and zebrafish (Varshney *et al.*, 2016). However, the reverse of this process; going from phenotype to genotype is far more complex. This is because phenotypes are most often the result of complex interactions across multiple genes and biological pathways (Carlborg and Haley, 2004).

Pharmacogenomics, an example of phenotype to genotype assessment, is the study of how drug response is affected by genes, and is the key to personalised medicine. For example, when a gene variant is associated with a particular drug response in a patient, there is the potential for making clinical decisions based on genetics by

choosing which patients to give the drug too. The objective of personalised medicine is to identify those subgroups of patients which are most likely to respond to a therapy and thus increase clinical trial success rates and improve patient outcomes.

Designing personalised medicine clinical trials requires the identification of a biomarker at the pre-clinical stage. Expansion of screening studies across heterogeneous panels of cell lines allows this through the quantification of variations in response across distinct cell lines. By including multiple cell lines it allows associations to be made linking phenotypes to genotypes utilising basal genomic, transcriptomic and proteomic data (Hart, 2005; Schirle and Jenkins, 2016). This in turn, further supports compound mechanism of action studies, biomarker discovery, and target ID, as part of a larger personalized medicine strategy.

Recent large scale initiatives have produced a number of resources to help connect phenotype to genotype through the identification of genetic dependencies and small molecule sensitivities, including DepMap (Tsherniak *et al.*, 2017) and the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang *et al.*, 2013). By combining large scale drug screening and genomic analyses they aim to provide resources for therapeutic biomarker discovery in cancer cells. These databases, along with others such as Connectivity Map (Lamb *et al.*, 2006) for perturbation induced gene expression profiling, cover many hundreds of cell lines screened against many thousands of perturbations. They have been used in phenotypic drug discovery for compound similarity searching, mechanism of action elucidation, drug target identification and biomarker prediction (Rees *et al.*, 2016; Nguyen, Dang and Ballester, 2017; Suphavilai, Bertrand and Nagarajan, 2018; Wei *et al.*, 2019; Li *et al.*, 2020).

In this chapter, we shall explore both selectivity and sensitivity to disulfiram and elesclomol via integration of gene expression data across both OAC specific datasets that I have generated and larger pan cancer datasets which are publicly available.

Firstly selectivity over the tissue-matched control cell lines is addressed in order to better understand the anti-cancer mechanism of action of disulfiram and elesclomol but also to identify potential cancer specific pathways that could be leveraged in further drug discovery programs. Secondly, efficacy of disulfiram, elesclomol, and ammonium pyrrolidinedithiocarbamate across the OAC cell panel shows marked variation in sensitivity. Understanding the reason for this is critical both to better understanding the drug mechanism and also biomarker identification for patient stratification to better target those who would benefit from such a treatment.

## 6.2 Results

### 6.2.1 Selectivity over tissue-matched control cells

As discussed in the previous chapter, elesclomol, disulfiram and ammonium pyrrolidinedithiocarbamate are highly selective for OAC over tissue-matched controls. This is true for both the univariate; survival analysis, and the multivariate, phenotypic analysis. The dose responses show that there is no toxicity in either of the tissue-matched control cell lines (CP-A and EPC2-hTERT) and further that they continue to proliferate at their normal rate under 10 μM treatment for 48 hrs (Figure 6.1A). Moreover, the compounds did not cause any phenotypic changes in the EPC2-hTERT or the CP-A cell lines across the dose responses either (Figure 6.1B). This demonstrates strong selectivity, as phenotypic changes due to stress are usually present at lower concentrations than cell death or changes in proliferation.

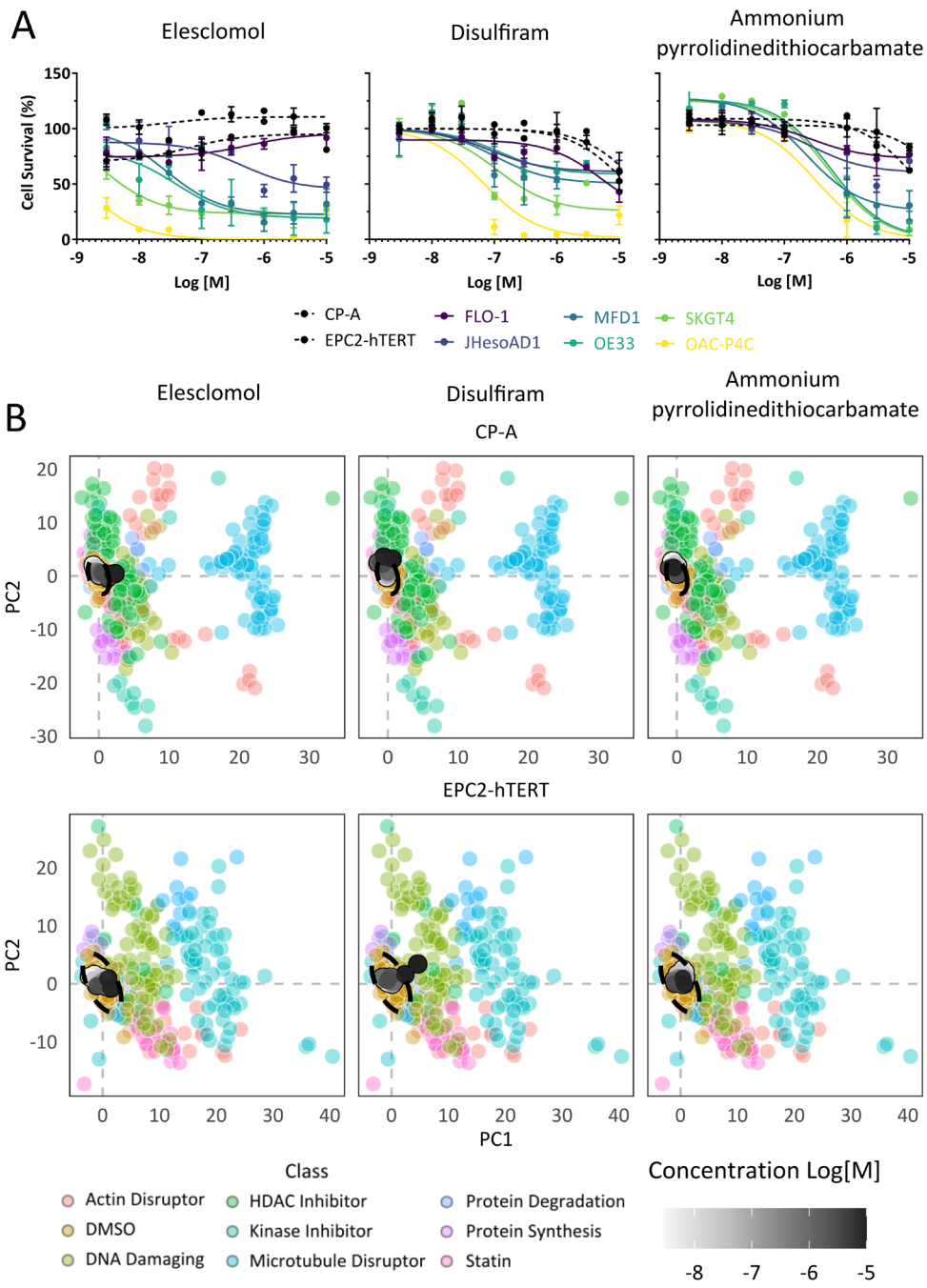



**Figure 6.1: Selectivity over tissue-matched control cell lines**. A) Univariate dose responses and B) Phenotypic dose responses, for elesclomol, disulfiram and ammonium pyrrolidinedithiocarbamate in the tissue-matched control cell lines. Library of reference compounds coloured by mechanistic class. Compound dose response overlay coloured by concentration.

It has been suggested that the selective toxicity of disulfiram and elesclomol towards cancer cells is due to the increased sensitivity of cancer cells to ROS (Kirshner *et al.*, 2008; Barbi de Moura *et al.*, 2012; Nagai *et al.*, 2012; Hasinoff *et al.*, 2014). However, our findings suggest that this is not the case. Firstly, looking at NanoString compound induced gene expression profiles neither elesclomol nor disulfiram induce any significant gene expression changes in the two tissue-matched control cell lines (Figure 6.2). This inability to elicit a response in gene expression suggests that the compounds do not disrupt major biological processes or pathways within tissue-matched control cells.

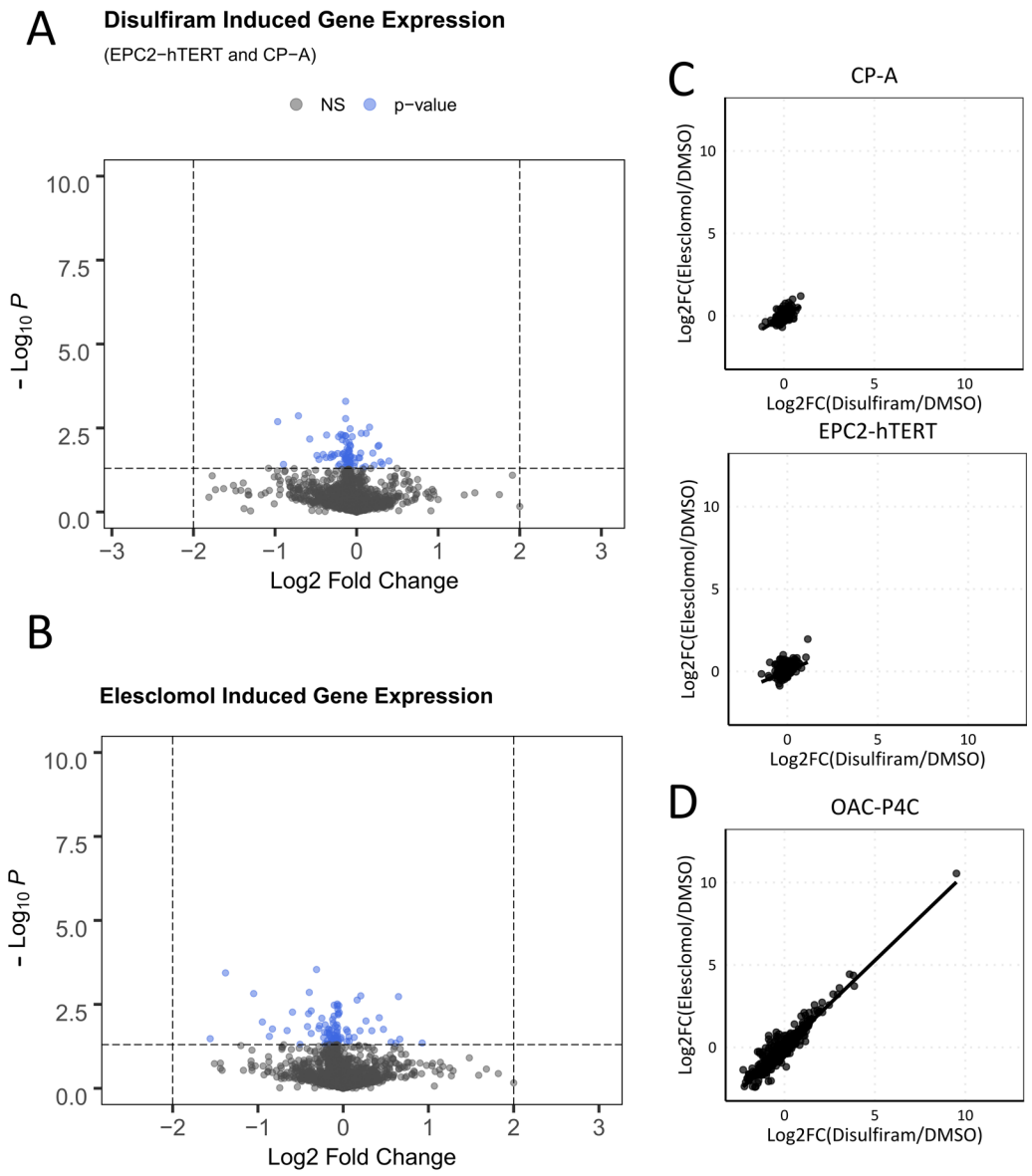



**Figure 6.2: Treatment induced gene expression studies in tissue-matched control cell lines.** Differential expression analysis for A) Disulfiram, B) Elesclomol, induced gene expression for EPC2-hTERT and CP-A. n = 1. No significant genes after correcting for multiple testing. Correlation plot for Disulfiram and Elesclomol induced log2 fold change gene expression changes in C) tissue-matched controls CP-A and EPC2-hTERT, D) Sensitive OAC cell line OAC-P4C for comparison.

To further probe this theory I assessed the intracellular copper concentration of the tissue-matched control cell line EPC2-hTERT to determine if selectivity is due to a lack of copper accumulation. ICP-MS showed no accumulation of copper in the tissue-matched control cells after incubation with either disulfiram or elesclomol (Figure 6.3). This is consistent with work carried out in peripheral blood mononuclear cells which were also found to lack copper accumulation (Nagai *et al.*, 2012). This finding is in line with both our phenotypic dose response data and the gene expression data, and suggests that these compounds are inactive in tissue-matched control cells. This is in contrast to the resistant OAC cell line, FLO-1, which is highly resistant to the compounds and shows no morphological response but it does accumulate copper (Figure 6.1, and Figure 5.2 and Figure 5.6- chapter 5) and displays a compound induced gene expression signature (Figure 5.5, chapter 5). These findings support the idea that unlike the resistant cells which are able to respond to and survive treatment, the compounds are inactive in the tissue-matched control cells. I also determined whether basal intracellular copper levels may differ between the tissue-matched control cell line ECP2-hTERT and the OAC cell lines accounting for the differential sensitivity. However overall, there is no consistent difference between the tissue-matched control and the OAC cell lines. There is also no correlation between basal intracellular copper levels and sensitivity to elesclomol or disulfiram across the OAC cell lines (Figure 6.3B).

Overall, this demonstrates that there is a lack of copper accumulation within tissue-matched control cells leading to the lack of toxicity, rather than the cells being more resistant to ROS stress. However, the mechanism by which elesclomol leads to the accumulation of copper within cancer cells specifically is still unknown and warrants further investigation.

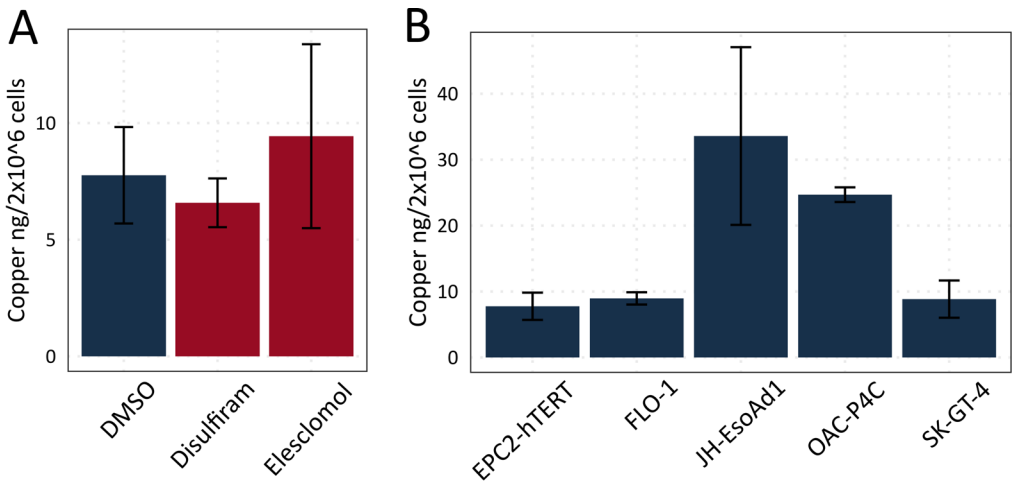



**Figure 6.3: ICP-MS copper**. A) Intracellular copper levels for EPC2-hTERT with DMSO, disulfiram (600 nM) and elesclomol (200 nM) treatment. B) Basal intracellular copper levels across tissue-matched control and OAC cell lines. Error bars indicate SE. 1-ANOVA and Tukeys post hoc test to test significance. No significance for treatments compared to DMSO or EPC2-hTERT basal compared to grouped OAC cell lines. n = 3.

In order to begin to unravel this mechanism of disulfiram and elesclomol selectivity towards OAC cells I used basal gene expression data to identify differentially expressed genes between the OAC and tissue-matched control cell lines. As expected, most of the top differentially expressed genes (although none were significant after correcting for multiple testing) are cancer-related (according to the human protein atlas) (Figure 6.4). However, interestingly the most significant gene, ID2, has been previously implicated in the progression of OAC (Gibson *et al.*, 2013), demonstrating the ability to identify functional genes that differentiate OAC and tissue-matched control cells. Another one of the top genes identified was the transferrin receptor (TFRC) (Figure 6.4), a cell surface receptor necessary for the cellular uptake of iron by receptor-mediated endocytosis. Among the top down regulated genes in OAC when compared to the tissue-matched control cells are two genes involved in lipid metabolism; fatty acid synthase (FASN) and stearoyl-CoA desaturase (SCD). To gain biological meaning from these findings I carried out pathway analysis using IPA.

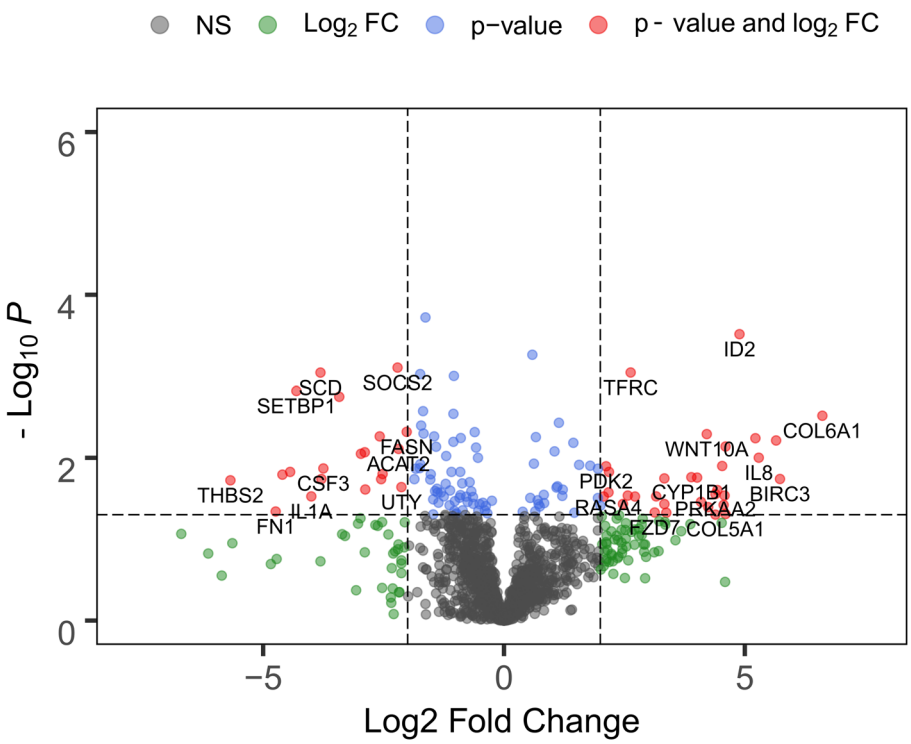



**Figure 6.4: OAC versus tissue-matched control basal differential expression analysis.** OAC (OAC-P4C, SK-GT-4, JH-EsoAd1, FLO-1) versus tissue-matched control (EPC2-hTERT, CP-A). *none are significant after correcting for multiple testing. n =1

Pathway analysis of the basal NanoString expression data revealed Liver X Receptor-Retinoid X Receptor (LXR/RXR) activation to be the most significant pathway differentiating OAC from tissue-matched control cell lines (Figure 6.5). The pathway is significantly overexpressed in the tissue-matched control cells compared to the OAC and is an important regulator of cholesterol, fatty acid and glucose homeostasis (Laffitte *et al.*, 2003; Laurencikiene and Rydén, 2012; Hamilton *et al.*, 2016). This is in line with the NanoString differential expression analysis (Figure 6.4), where FASN and SCD were among the top down regulated genes in OAC when compared to the tissue-matched control cells, both are part of the LXR-RXR activation pathway defined in IPA. Intriguingly, the pathway has also recently been identified as important in Wilson disease; a disorder caused by mutations in the copper-transporter ATPase copper-transporting beta (ATP7B),

that leads to copper accumulation within the liver (Hamilton *et al.*, 2016). Here they found that reduced LXR/RXR nuclear receptor function triggered disease progression and furthermore that upregulation of LXR/RXR function ameliorates the disease, demonstrating a connection between LXR/RXR pathways and copper sensitivity in a disease situation. This fits with our findings that OAC cells with reduced LXR/RXR are more sensitive to copper accumulation in a drug induced situation. However, while Hamilton *et al.* showed that chemical activation of LXR ameliorated disease manifestations, it did not lead to loss of copper accumulation in APT7B null mice suggesting this pathway does not explain the lack of copper accumulation in elesclomol treated tissue-matched control cells, but may explain the increased sensitivity of OAC to copper accumulation.

The fourth most significant pathway was caveolar-mediated endocytosis (Figure 6.5). CAV-1, a major component of caveolar-mediated endocytosis, has very recently been shown to stabilise the copper transporter ATPase copper-transporting alpha (ATP7A) (Sudhahar *et al.*, 2020), demonstrating both pathways have been linked to copper homeostasis previously. Intriguingly, TFRC expression (identified through the NanoString differential gene expression analysis) has been linked to regulation of caveolin-dependent endocytosis (Lei *et al.*, 2016), supporting its involvement in elesclomol selectivity. Something both these identified pathways have in common is a link to lipid metabolism and cholesterol homeostasis (Luo *et al.*, 2010). Furthermore, lipid metabolism and in particular cholesterol biosynthesis have been identified as major pathways inhibited in response to copper accumulation in Wilson disease (Hamilton *et al.*, 2016). This data provides preliminary evidence for further exploration of the role these pathways play in the selectivity of disulfiram and elesclomol for cancer cell cytotoxicity, whether this is linked to lipid metabolism, and how these pathways prevent the accumulation of copper in tissue-matched control cells.

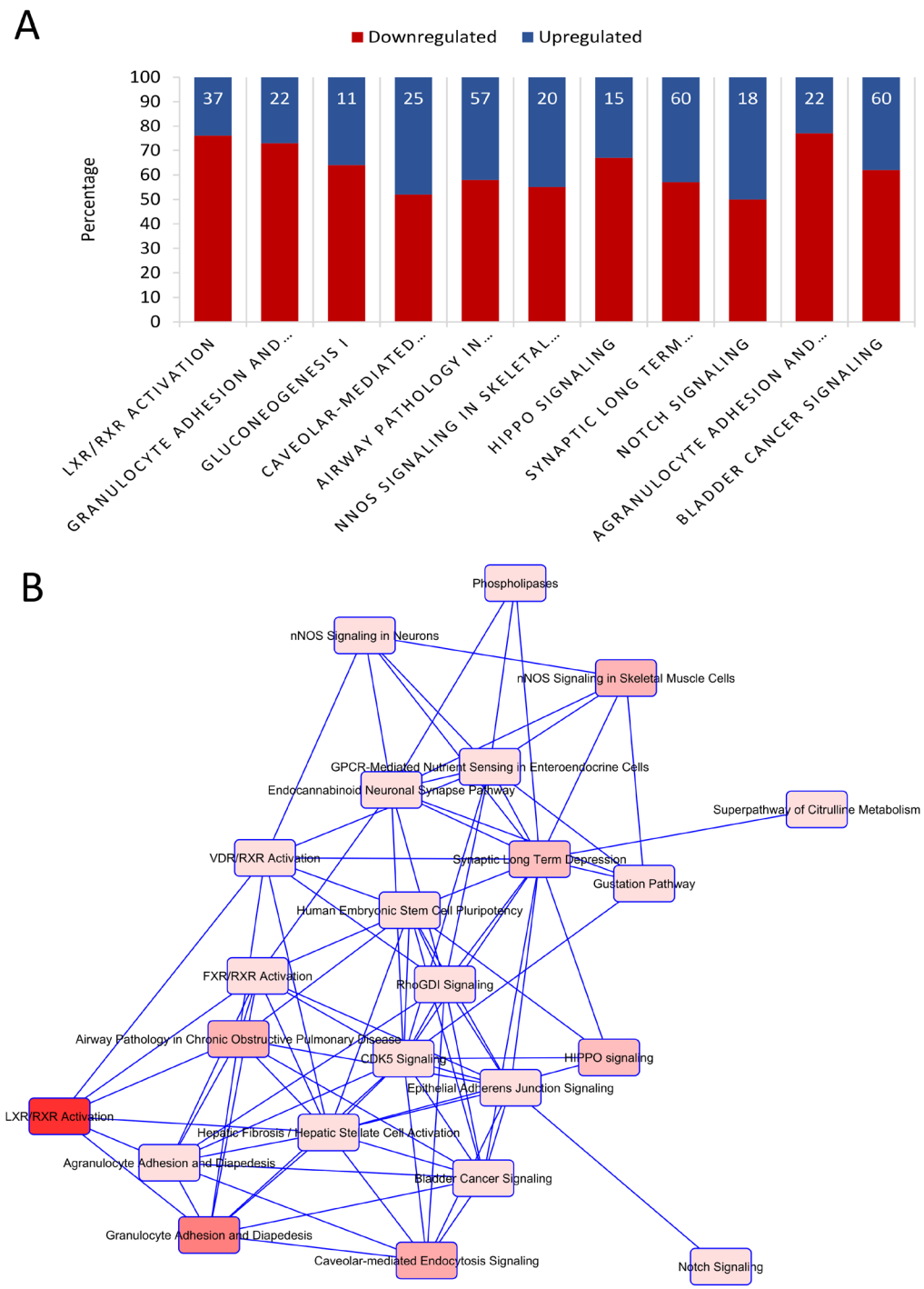



**Figure 6.5: OAC versus tissue-matched control basal gene expression IPA pathway analysis.** A) Up and down regulated genes in OAC compared to tissue-matched controls for significant pathways ranked by significance left to right.

Numbers over bars signify total number of genes in pathway. B) Network generated from top 25 pathways that differentiate OAC from tissue-matched control cells. Minimum number of shared genes between pathways is 2. Degree of red shading represents significance, with more significant pathways in darker red.

In summary, this work in conjunction with studies performed in chapter 3 provides evidence that the selectivity of disulfiram and elesclomol is unrelated to the sensitivity of cancer cells to ROS, but rather is due to a lack of copper accumulation in tissue-matched control cells. Study of gene expression profiles has revealed several pathways, including an overarching association to lipid metabolism, that may be responsible for the selectivity but further experimental work is required.

### 6.2.2 Sensitivity to disulfiram and elesclomol

Given the clear variability in OAC cell line sensitivity to disulfiram and elesclomol, understanding the differences accounting for this variability and identifying the mechanisms which confer sensitivity to these compounds is critical to both better understanding their mechanism of action, and identification of patient stratification hypotheses.

From the mechanistic copper studies (Figure 5.6 and Figure 6.3) it is clear that neither basal nor post-treatment intracellular copper levels could explain the differential sensitivity of the OAC panel to disulfiram and elesclomol, since both sensitive and resistant cell lines accumulate comparable levels of copper and there is no clear separation in basal levels either. Secondly, it also appears unlikely that sensitivity to these compounds is linked to general reactive oxygen species (ROS) sensitivity. I therefore sought to identify the mechanisms which confer sensitivity through the integration of basal transcriptomic data.

Firstly, to assess the sensitivity of our OAC cell lines compared across a larger panel of pan cancer cell lines, I made use of data from the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang *et al.*, 2013). This dataset describes the drug sensitivity of 1000 cell lines to 100's of compounds. The dataset includes the sensitivity of 881

pan cancer cell lines to elesclomol. Sensitivity ranking shows the sensitive OAC lines identified in my studies to be in the top 200 most sensitive cell lines and the resistant OAC lines to be in the top 200 least sensitive cell lines (Figure 6.6). Given the obvious separation in sensitivity of our OAC lines, this should allow us to determine differences driving sensitivity. This analysis also highlights the need for a biomarker or patient stratification hypothesis since it is clear only a subset of cancers are likely to response to elesclomol treatment.

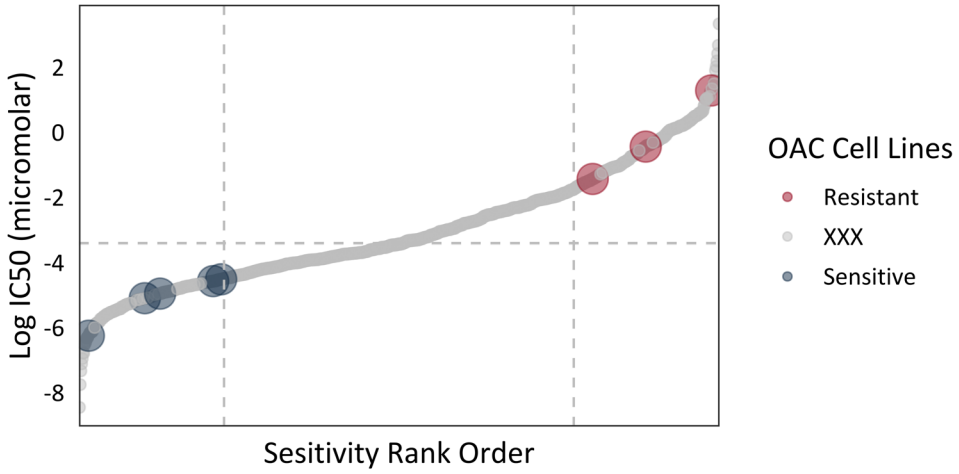


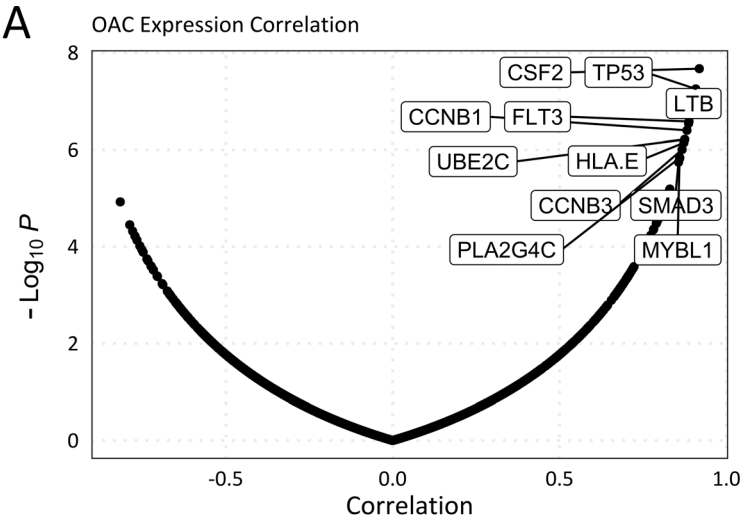

**Figure 6.6: OAC comparison to pan cancer elesclomol sensitivity**. Genomics of Drug Sensitivity in Cancer GDSC1 data. OAC cell lines are highlighted and classified as either sensitive or resistant to elesclomol. Vertical lines represent top and bottom 200 cell lines by elesclomol sensitivity ranking.

### Basal gene expression

It has recently been established that basal gene expression can be used not only to infer mechanism of action (Rees *et al.*, 2016) but also to identify biomarkers of drug sensitivity (Li *et al.*, 2020). Using the NanoString nCounter platform I quantified the basal gene expression profile of nine OAC cell lines across the Human PanCancer Pathways and Metabolic Pathways panels, encompassing around 1,400 genes. Correlating basal gene expression with elesclomol sensitivity, as determined by CellTitre Glo viability assay (chapter 5) revealed CSF2, UBE2C, CCNB1 and TP53 as the top hits (Figure 6.7A). However, after correcting for multiple testing none of the genes were significant.

While correlation of basal gene expression across OAC cell lines did not identify any significant genes, grouping the sensitive and resistant cell lines for differential expression analysis identified a number of potential biomarkers of sensitivity (Figure 6.7B). Methionine sulfoxide reductase B2 (MSRB2) was one of the most significant genes.

To gain better predictive power, I also expanded the cell lines to a pan cancer correlation of gene expression analysis. Combining the two approaches could identify robust gene markers of sensitivity. In order to do this I made further use of the data from the GDSC, encompassing both gene expression from 17715 genes and elesclomol area under the curve (AUC) values for an expanded pan cancer analysis using 916 cell lines. Of note, this analysis also identified a methionine sulfoxide reductase-MSRB3 (Figure 6.7C). Taken together, these two analyses, suggest methionine sulfoxide reductases may play a role in sensitivity to elesclomol and copper.



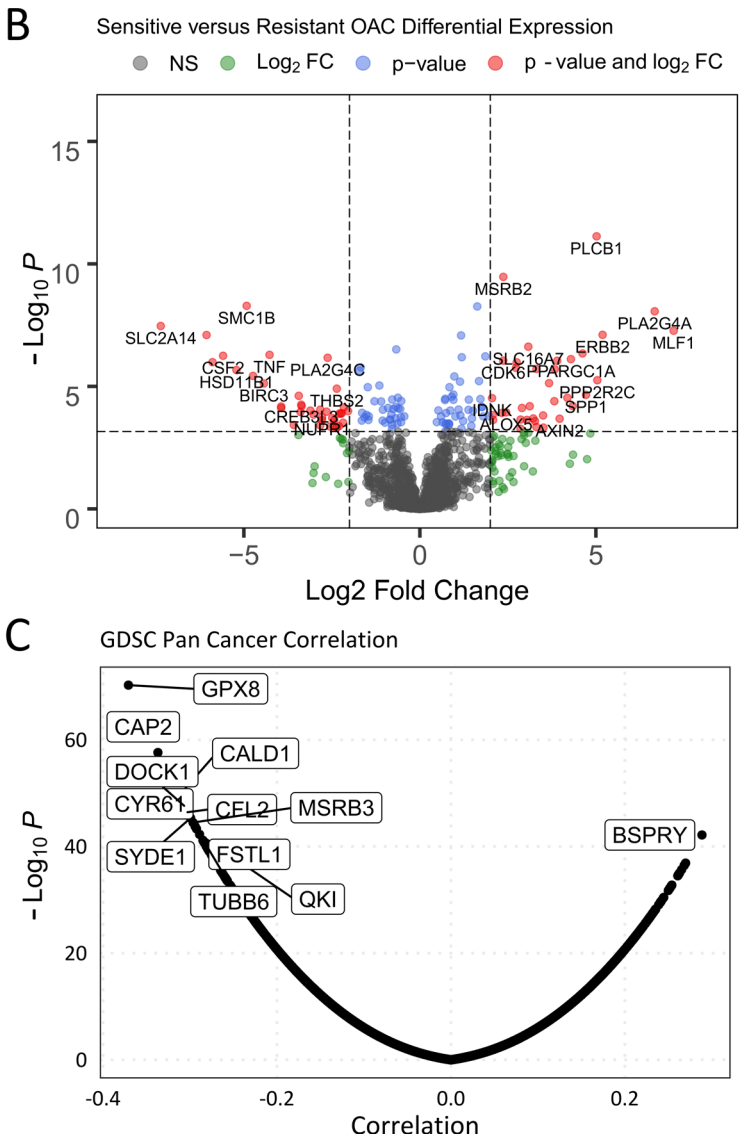



**Figure 6.7: Gene expression defining elesclomol sensitivity.** A) Correlation analyses of basal gene expression and elesclomol $IC_{50}$ (as determined by CellTitre Glo, chapter 5) across nine OAC cell lines (OAC-P4C, SK-GT-4, ESO26, OE19, OE33, ESO51, MFD-1, JH-EsoAd1, FLO-1) gene expression n = 3 B) Differential expression analysis. Sensitive cell lines: OAC-P4C, SK-GT-4, OE19, ESO26. Resistant cell lines: FLO-1, JH-EsoAd1, ESO51. N=3. P value threshold equivalent to adjusted p value 0.05. Gene total = 1397. NanoString nSolver software. C) Correlation analyses of Genomics of Drug Sensitivity in Cancer. RMA normalised expression values for 17,737 genes against elesclomol area under the curve for 916 cell lines.

## Pathway and gene set enrichment analysis

It has been noted that where single-gene analysis finds few or weak associations in gene expression, pathway and gene set enrichment analysis (GSEA) may reveal many significant associations at the biological pathway level (Subramanian *et al.*, 2005). To this end, I made use of RNAseq data for seven OAC cell lines from the GDSC and used GSEA, a method that focuses on gene sets, that is, groups of genes that share common biological function and their enrichment in any given phenotype, to identify gene sets whose expression differentiates sensitive and resistant cell lines, as determined by CellTitre Glo (see chapter 5). GSEA hallmark analysis revealed 10 gene sets that positively correlated with $IC_{50}$ and 11 that negatively correlated, to be significantly enriched ($p < 0.05$) (Figure 6.8). MYC targets are the top two gene sets positively associated with elesclomol sensitivity. Within these gene sets the top differentially expressed genes are ubiquitin enzymes and proteasome subunits (Table A.4). In concordance with this the unfolded protein response is the fifth gene set identified.

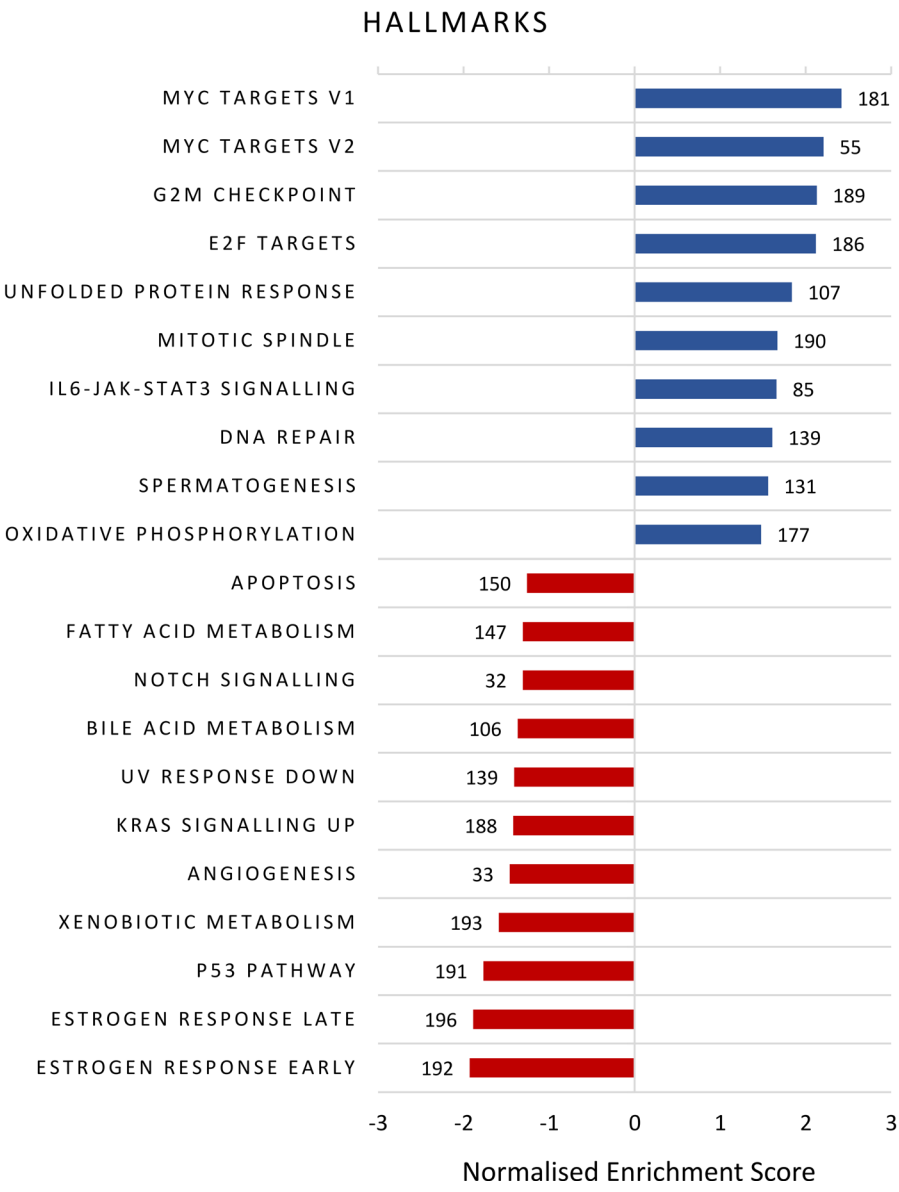



**Figure 6.8: GSEA Hallmarks analysis.** Using the MSigDB Hallmark collection of 50 gene sets. Numbers next to bars represents gene set size. Pearson was used as a rank metric.

Furthermore, GSEA of KEGG (Kyoto Encyclopedia of Genes and Genomes) gene sets, revealed 18 gene sets significantly enriched and positively correlated with elesclomol sensitivity, of which the proteasome was the top gene set. This analysis

corroborates the findings in the previous chapter, suggesting that proteostasis may be involved in both the mechanism and sensitivity to elesclomol though this remains to be confirmed experimentally. Of note, cysteine and methionine metabolism was identified as the sixth most significant KEGG pathway from the GSEA analysis, giving further confidence in the identification of the methionine sulfoxide reductase enzymes at the single gene level (Figure 6.9).

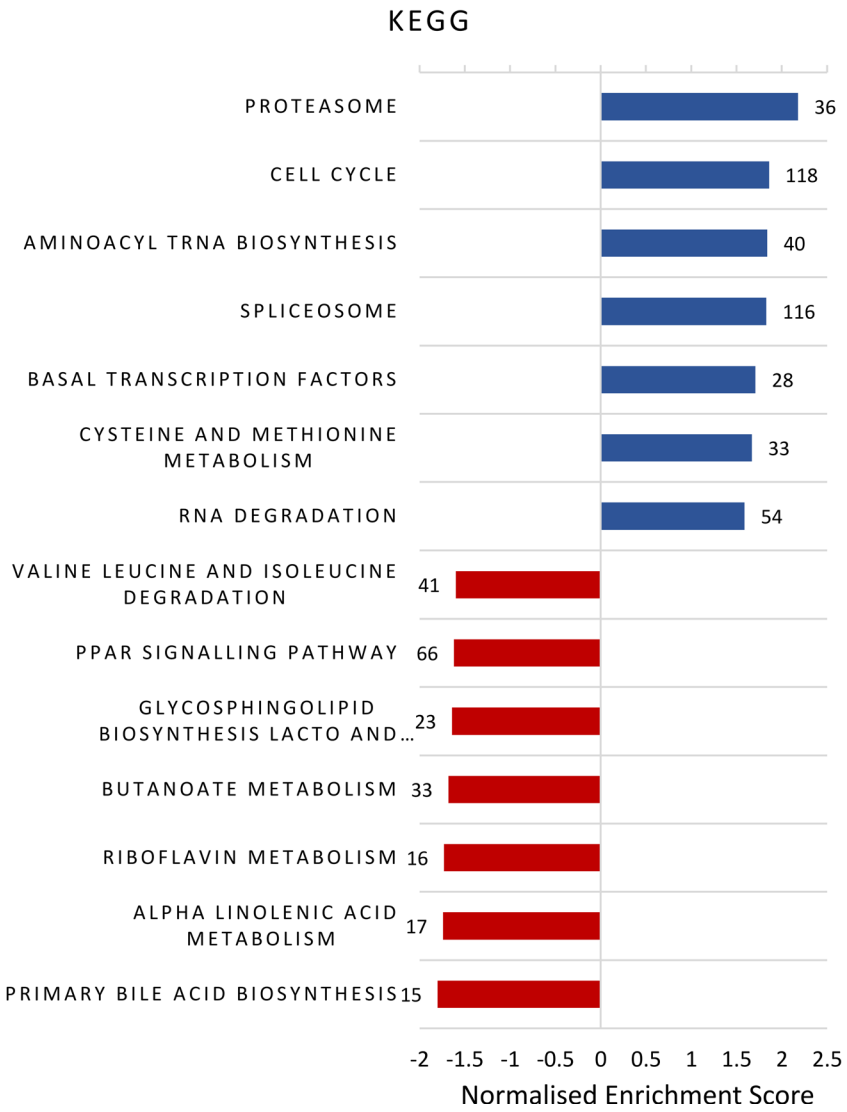



**Figure 6.9: GSEA KEGG analysis.** Using the canonical pathways KEGG collection of 186 gene sets. Numbers next to bars represents gene set size. Pearson was used as a rank metric.

Of the pathways negatively associated with sensitivity. The top two hallmarks gene sets are the oestrogen response (early and late), followed by the TP53 pathway, xenobiotic metabolism, bile acid metabolism and fatty acid metabolism (Figure 6.8), while the top KEGG gene sets are bile acid biosynthesis, linolenic acid metabolism, riboflavin metabolism, butanoate metabolism and glycosphingolipid biosynthesis (Figure 6.9)(Table A.4-Table A.8 for gene rankings within genesets with enrichment scores greater than 2 or less than -2). These are all metabolic pathways concerning carbohydrate and lipid metabolism, and present further evidence for an association of lipid metabolism and sensitivity to disulfiram and elesclomol. In line with this Phospholipase C Beta 1 (PLCB1) was the most significant gene identified at the single gene level (Figure 6.7B), although not statistically significant after multiple testing corrections.

The role of TP53 in sensitivity to elesclomol was further studied (in collaboration with Kenji Fujihara, Peter MacCallum Cancer Centre, Australia) due to its identification as one of the top genes correlated with sensitivity (Figure 6.7A), its identification at the geneset level, and its known role in iron sensitivity and ferroptosis (Kang, Kroemer and Tang, 2019). However we found that TP53 knockout in the TP53 high, elesclomol resistant, cell line JH-EsoAd1 did not alter sensitivity to elesclomol (Figure A.3).

### Basal mitochondrial metabolism

Given the identification of MSR genes, some of which are known to localise to the mitochondria and the potential involvement of lipid metabolism; along with the proposed role of FDX1, a mitochondrial Fe-S cluster biogenesis protein and the identification that LDH low patients respond more favourably to elesclomol in melanoma trials (Hauschild *et al.*, 2009), I thought there may be metabolic differences underlying the sensitivity of OAC cells to disulfiram and elesclomol. I therefore explored respiration parameters across two sensitive (OAC-P4C and SK-GT-4) and two resistance cell lines (FLO-1 and JH-EsoAd1) to identify differences in mitochondrial function that may explain the differential sensitivity (**Figure 6.10**). Samples were prepared and analysed on a Seahorse Bioanalyzer at the Queens

Medical Research Institute with the assistance of Dr. Roderick Carter as described in materials and methods.

In contrast to hypotheses in the literature (Barbi de Moura *et al.*, 2012; Blackman *et al.*, 2012), there was no significant difference in basal respiration between the sensitive and resistant OAC lines. However there are significant differences in the maximal respiration rate and the spare respiratory capacity, and both parameters are also significantly correlated with sensitivity, as determined by $IC_{50}$ correlation (Kendal tau max respiration = 0.87, spare capacity = 0.89). An inability to increase respiration rates beyond that of the basal rate suggests a cell line is energetically inflexible, leaving them unable to adapt in response to stress.

Overall, it appears that there are differences in metabolism and that there may be a reduced mitochondrial energy flexibility and spare respiratory capacity in sensitive cell lines, possibly leaving cells vulnerable to further disruption of mitochondrial respiration upon elesclomol treatment as seen in the previous chapter. This may be linked to the identification of lipid metabolism pathways from GSEA.

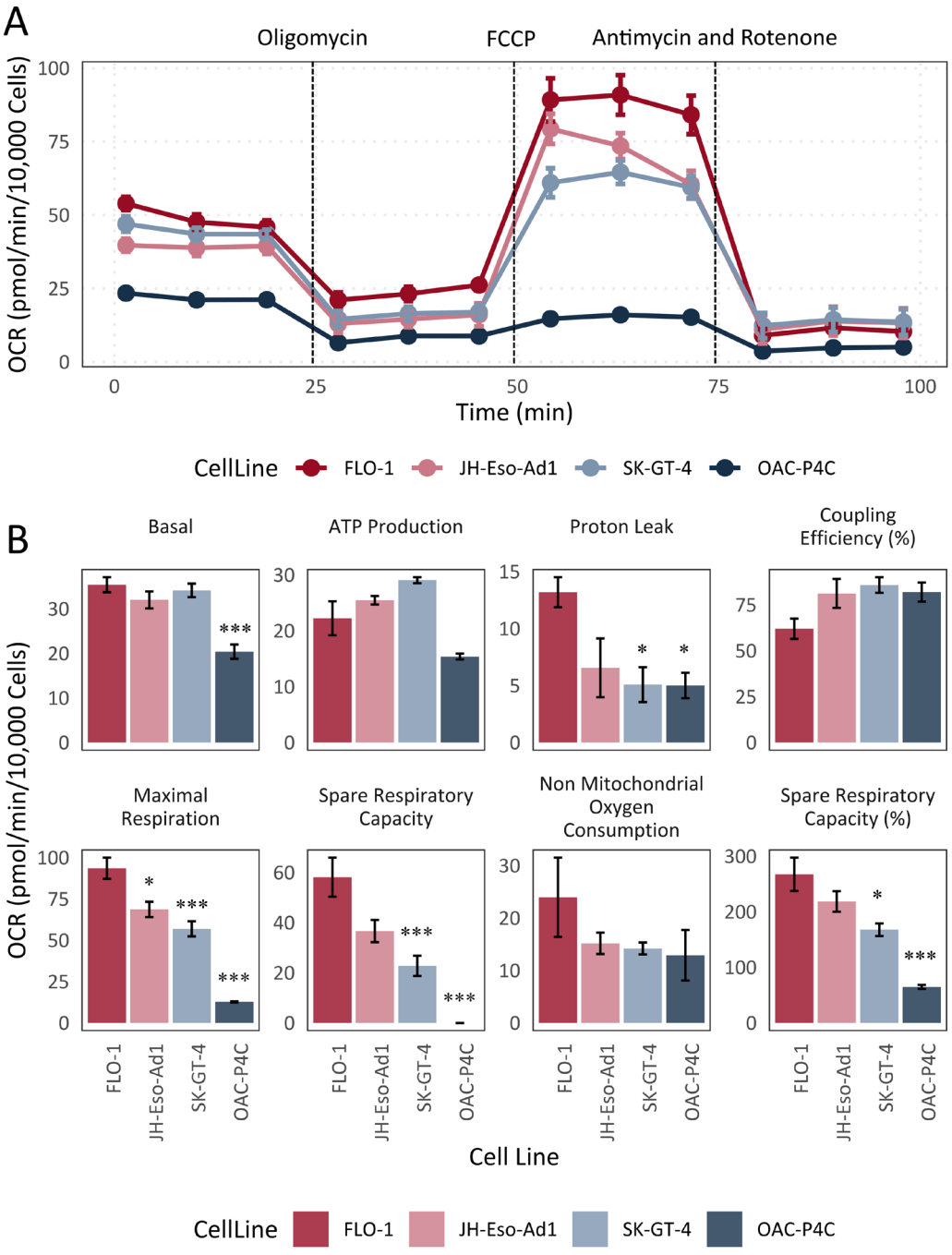



**Figure 6.10: Basal mitochondrial function.** Seahorse Oxygen Consumption Rate (OCR) test. A) Exemplar OCR trace for each cell line. OCR normalised per 10,000 cells. Dark red = FLO-1 (most resistant), light red = JH-EsoAD1 (second most resistant), dark blue = OAC-P4C (most sensitive), light blue = SK-GT-4 (second most sensitive). B) Parameters for each cell line normalised to 10,000 cells n=3. Error bars indicate SEM. 1-ANOVA and Tukeys post hoc test for significance. Significance

indicated when compared to most resistant cell line, FLO-1. * P>0.05, ** p>0.01, *** p>0.005.

## 6.3 Discussion

In conclusion, while these findings create many more questions than they answer, they do provide a number of testable hypotheses that could be answered in the future. Firstly, regarding the selectivity of disulfiram and elesclomol for cancer, the work here demonstrates that there is a lack of toxicity and phenotypic changes in tissue-matched control cells and further that there are no changes in gene expression upon treatment. In this work I confirmed that this selectivity is likely due to the lack of copper accumulation, rather than an increased resistance to ROS as previously postulated in the literature. The question remains as to whether this is due to efficient export of copper or a lack of release in the first place. One piece of evidence suggesting that elesclomol can release copper in non-cancerous cells is the discovery that elesclomol restores cellular copper levels in non-cancerous copper-deficient cells such as seen in Menkes disease (Soma *et al.*, 2018; Guthrie *et al.*, 2020). This would suggest that tissue-matched control cells are more capable of regulating copper levels than cancerous cells and that it is this that lends itself to the selectivity of the compounds. Inhibition of copper export mechanisms in conjunction with elesclomol treatment may determine whether this is true. If found to be true, this might fit with the pathway analysis that highlighted caveolar-mediated endocytosis as significantly overexpressed in the tissue-matched control cells.

Overall, pathway analysis highlighted several pathways that are linked to lipid metabolism and are also linked to the ATPase copper-transporting genes ATP7A and ATP7B that reside in the trans-Golgi network. There is also a link between the iron transporter TFRC (identified by differential expression) and these pathways (Lei *et al.*, 2016). While it is known that iron and lipid trafficking are closely linked (Rockfield *et al.*, 2018), we did follow up any links to iron levels by quantification of intracellular levels by ICP-MS, and iron chelation and ferroptosis inhibitors (in collaboration with Kenji Fujihara, Peter MacCallum Cancer Centre, Australia), to

find no associations of elesclomol activity or selectivity with intracellular iron levels (Figure A.1, Figure A.2). But perhaps it is the lipid trafficking aspect that is linked to iron that is important here rather than the iron itself, something we have yet to follow-up. If and how these pathways play a role in elesclomol selectivity and prevent the accumulation of copper within tissue-matched control cell lines remains to be explored.

Secondly, with regards to the sensitivity of OAC cell lines to disulfiram and elesclomol, this work has identified a number of potential genes, pathways and mechanisms which may define sensitivity. A the single gene level, both OAC specific and pan cancer analyses revealed Methionine sulfoxide reductase (MSR) genes as potential biomarkers of sensitivity. Methionine metabolism was also identified in the gene set enrichment analysis (GSEA). MSR genes have been identified as upregulated in cancer previously, for example MSRB3 is significantly upregulated in clear cell renal carcinoma (RCC) samples compared to adjacent tissues (Ye *et al.*, 2020). Furthermore, MSRA has also been shown to be upregulated in human epidermis in response to UVA and $H_2O_2$ (Ogawa *et al.*, 2006), which may provide an explanation for the sensitivity of melanoma to elesclomol (Barbi de Moura *et al.*, 2012; Blackman *et al.*, 2012). In line with these findings, it has been shown that copper affects the stability and/or assembly of Fe-S proteins in eukaryotic cells (Garcia-Santamarina *et al.*, 2017), and that knockout of the MSR genes in yeast leads to a copper resistance phenotype and is related to Fe-S cluster function (Sideri, Willetts and Avery, 2009). These previous studies demonstrate a relationship between copper sensitivity and MSR gene expression in yeast and other organisms. Our finding here suggest that this extends to elesclomol induced copper toxicity and that this occurs in pan human cancers, not just OAC. Manipulation of this pathway would provide further evidence confirming this hypothesis.

At the pathway and gene set level, GSEA highlights the proteasome as one of the pathways that is significantly positively correlated with sensitivity and this is in line with the identification of Ubiquitin Conjugating Enzyme E2 C (UBE2C) at the single gene level (although it was not significant after multiple testing). In support

of this association is the identification of proteostasis as one on the major pathways disrupted in response to disulfiram and elesclomol treatment in OAC cell lines discussed in the previous chapter. GSEA also identified a number of pathways that negatively correlated with elesclomol sensitivity, namely bile acid biosynthesis, linolenic acid metabolism, riboflavin metabolism, butanoate metabolism and glycosphingolipid biosynthesis. The identification of glycosphingolipid biosynthesis is intriguing since according to the Connectivity Map database, the fifth most related compound to the elesclomol gene expression signature, after pyrrolidine dithiocarbamate, is that of sphingosine, a ceramidase inhibitor and a primary constituent of sphingolipids, with a connectivity score of 98.03 (Table A. 9), suggesting a shared mechanism of action.

Finally, mitochondrial metabolism studies suggest that there are metabolic differences between the sensitive and resistant cells, but that this is not simply due to a difference in basal mitochondrial respiration levels as previously thought. Perhaps more in-depth study of the relative contribution of primary substrates that fuel the mitochondria, e.g. fatty acids (lipid metabolism), glucose, pyruvate and glutamine, across sensitive and resistant cells may further elucidate how these pathways contribute to sensitivity of disulfiram and elesclomol and how we may best exploit them to treat patients in the future.

# 7| *Conclusions*

In this work I have attempted to identify drug repurposing opportunities in OAC and provide patient stratification hypotheses to support clinical translation through the development and validation of a high content profiling assay with integrated machine learning and transcriptomic pathway analysis.

The first results chapter builds on published literature to further investigate how compound induced morphological changes can be used to inform compound mechanism of action and high content assay quality. While the work described here, and that carried out by others (Ljosa *et al.*, 2013) demonstrates that cellular morphology does indeed predict compound mechanism of action, it has only been studied in the context of clearly defined, discrete mechanisms and reference libraries. Future work incorporating larger target-annotated reference compound libraries with less well defined classes, such as compounds with polypharmacology and combination treatments is required to determine how the interplay of multiple mechanisms and targets, both multiple on-target and off-target effects, can be captured and determined from cellular morphology. This will be pivotal for the study of high content and phenotypic programs for heterogeneous and complex diseases such as OAC, where successful treatments will likely involve combination therapies and/or compounds displaying polypharmacology.

This work confirms that compounds induce, at least to some extent, core morphological changes present across heterogeneous cell lines. Certainly, one very useful tool for the research community would be a “morphological atlas” to

compare compound morphological profiles and identify similarities between unknown hits and reference compound or functional genomic datasets. Although in its infancy, one such tool may arise from the JUMP-CP consortium announced last year (https://jump-cellpainting.broadinstitute.org/), where they plan to create the largest perturbation induced Cell Painting dataset including both chemical and genetic perturbations, to match diseases, drugs and genes. Combining one such morphological profiling database with compound induced gene expression databases as seen in the Connectivity Map (Lamb *et al.*, 2006) could prove even more powerful as a tool to discover new connections among compounds as well as to identify targets and mechanisms of action, as has recently been explored by Lapins and Spjuth. In this work they found that while machine learning models built from either gene expression or morphological profiles performed equally well with a number of mechanisms of action, they performed notably different with other mechanisms (Lapins and Spjuth, 2019). This suggests that the two profiling approaches are capable of capturing distinct information content and highlights the importance of integrating datasets in the future.

Although not explored in this thesis, hit triage is a key aspect of drug discovery. Since not all compound hits can be followed up, having assessment criteria that lead to the selection of efficacious, selective, non-toxic and novel compounds is critical to reducing attrition rates and costs, and increasing the success of drug discovery programs. Morphological reference libraries could be applied as one such hit triage strategy. By building a reference library of known desirable and/or undesirable mechanisms of action, we could use this to triage screening hits to remove unwanted mechanisms or off-target activities, and instead select for desirable or novel mechanisms. This may prove a particularly beneficial strategy in instances when novel, first-in-class drug are a priority, such as with antibiotics, due to the rapid development of resistance within a given class of antibiotics, or in areas of unmet therapeutic need such as OAC, where current treatments have not proved efficacious, so as to avoid duplication of failed classes of compounds.

The second chapter describes the application of the optimised high content assay to a comprehensive small-molecule screen encompassing 19,555 small molecules,

eight cell lines, 512x384 well plates, 3.9 million images and 36 TB of data for the identification of new drug discovery and drug repurposing opportunities in OAC. The purpose of this section of work was to demonstrate the assay platform was capable of identifying OAC selective compounds with a high validation rate, using the subset of 3000 small-molecules suitable for repurposing opportunities. This gives us the confidence to explore the novel chemical space encompassed in the diversity libraries in ongoing work in the Carragher group. Further, it demonstrates the platforms suitability for even larger scale studies and CRISPR screens in the future. These could identify new therapeutic opportunities and genetic vulnerabilities in OAC to support the development of future treatments.

Work in the second chapter also flags a warning to researchers hoping to transfer learning to new cell lines or compare morphological signatures among cell lines. As is perceivable when studying any compound induced signatures, including morphological, not all cell lines respond equally, as exemplified by the differing responses of the cell panel to the well characterized DNA damaging agents Methotrexate, Raltitrexed and Pemetrexed. Nevertheless, this is almost certainly true when using gene expression signatures too and yet it is now common practice to compare gene expression signatures using online databases as seen in the Connectivity Map (Lamb *et al.*, 2006). This risk will most easily be mitigated through the incorporation of cell panels, reflecting multiple genetic backgrounds[v].

In the third chapter, I attempted to better define the mechanism of action of ammonium pyrrolidinedithiocarbamate, disulfiram and elesclomol; three of the most potent and selective hits identified from the repurposing libraries. While it is not a prerequisite for the progression of candidate drugs, there is a preference to carry out target deconvolution and identify the mechanisms responsible for a compounds phenotypic activity (Moffat *et al.*, 2017; Vincent *et al.*, 2020). This is because it increases our understanding of the disease mechanisms at play, affords

[v] There will always be exceptions to the rule. All we can do as researchers is try to ensure our models reflect the rule rather than the exception.

us the ability to integrate *in silico* and target based approaches for the development of second-in-class and follower molecules, and aids the identification of biomarkers. However, this proves no mean feat when compounds from phenotypic screens can, and often do, display activity against multiple on- and off-targets (Lin *et al.*, 2012; Lu *et al.*, 2012).

The work presented here provides evidence of a universal mechanism of action shared by all three compounds in OAC. Delving into the mechanism itself, the findings of this chapter lead me to conclude that the cell death associated with elesclomol and disulfiram is copper dependent and due to loss of proteostasis, rather than oxidative stress, as is proposed in the literature (Kirshner *et al.*, 2008; Barbi de Moura *et al.*, 2012; Nagai *et al.*, 2012; Hasinoff *et al.*, 2014). Consequently, cells mostly react to the accumulation of copper by mounting a heat shock response and trying to restore protein homeostasis. This fits with the transcriptomic analysis revealing strong activation of the heat shock response (HSR) and the unfolded protein response (UPR), while pathway analysis identified SAPK/JNK signalling, NRF2-mediated Oxidative Stress Response, unfolded protein response (UPR), endoplasmic reticulum (ER) stress, and p53 signalling as the top five altered pathways after elesclomol treatment. This signature fits very strongly with the signature induced by excess copper in a recent study of the effects of Glutathione (GSH) on excess copper in cells (Saporito-Magriñá *et al.*, 2018), providing further evidence to the requirements of copper in the mechanism of action. Interestingly, and in line with my findings, this study further demonstrated that GSH prevents copper toxicity, but not through its role in oxidative stress relief as previously believed (Mostofa, Seraj and Fujita, 2014). Instead, Saporito-Magriñá *et al.* provide evidence that cell death is driven by the interaction of cuprous ions with proteins which impairs protein folding and promotes aggregation. The protective role of GSH is based on their binding to cuprous ions, thus preventing the metal interacting with proteins.

In relation to the above findings, N-acetyl-L-cysteine (NAC), is a commonly used reagent in biology to assess the effects or contribution of oxidative stress and ROS to a phenotype of interest (Wang *et al.*, 2018; Xie *et al.*, 2018). It has been used in

numerous studies for the activity of metal ions, including copper (Zheng *et al.*, 2010) and also the activity of elesclomol (Kirshner *et al.*, 2008). However, rather unexpectedly, in this work I discovered NAC pre-incubation prevents elesclomol activity and cell death upstream of ROS antagonism, by preventing the increase in intracellular copper levels. This means that when studying the effects of copper (and possibly other metal ions) it should not be used to confirm the effects of ROS. Further I believe this may hold true for many other molecules, particularly thiols (Carrasco-Pozo *et al.*, 2008), including GSH (Freedman, Ciriolo and Peisach, 1989; Prutz, 1994; Carrasco-Pozo *et al.*, 2008). This highlights the importance as a researcher of knowing what you are measuring in complex cell based assays - a problem perhaps difficult to overcome in the field of biology, where we are often measuring things indirectly.

Finally, molecular targeted therapies have made limited impact upon the treatment of many complex cancers such as OAC, and many other cancer subtypes which do not fall into clearly defined molecular subgroups. This lack of response is presumably due to the high degree of inter- and intra-patient tumour heterogeneity observed (Secrier *et al.*, 2016). Cancer heterogeneity is associated with multiple pathway redundancies and rapid tumour evolution contributing to disease progression and therapeutic resistance. The identification of drug combinations or therapeutic agents exhibiting polypharmacology and pleiotropic effects such as that exemplified here by disulfiram and elesclomol may prove more successful in the treatment of such complex diseases. However, with such multi-targeted therapeutic approaches it is critical to develop patient stratification hypotheses to identify those patients which are most likely respond.

In the final chapter, I attempted to gain insight into drug selectively and potential patient stratification hypotheses to support clinical translation through the integration of transcriptomic pathway analyses. While the majority of published Cell Painting assays has largely been limited to a single cell line, in the work described here I have applied this across a panel of genetically and morphologically distinct OAC cell lines and tissue-matched control cells and demonstrated the ability to link compound induced phenotypes to transcriptomic

signatures to identify biomarkers of phenotypic response, critical for the success of therapies for heterogeneous diseases such as OAC.

Here I identified a lack of copper accumulation in tissue-matched control cells, and I propose this to be the mechanism purveying the selectivity of disulfiram and elesclomol for cancer cells. The mechanisms responsible for the lack of copper accumulation in tissue-matched control cells remain to be identified and confirmed. In this work through the use of basal transcriptomic studies and pathway analysis I highlight two potential pathways; LXR/RXR activation and caveolar-mediated endocytosis, both of which have ties to cholesterol, fatty acid and glucose homeostasis (Laffitte *et al.*, 2003; Luo *et al.*, 2010; Laurencikiene and Rydén, 2012; Hamilton *et al.*, 2016), as well as links to two diseases associated with dysregulation of copper homeostasis; Wilson disease (Hamilton *et al.*, 2016) and Menkes disease (Sudhahar *et al.*, 2020). Future studies are needed to unravel any potential roles of these pathways in copper accumulation and elesclomol selectivity, which may reveal biological differences between cancer and tissue-matched control cells to support future drug discovery and drug repurposing efforts. These studies may also provide insights and potential therapeutic strategies for two currently untreatable diseases; Wilson disease and Menkes disease.

As previously mentioned it is critical to identify the mechanisms of sensitivity to disulfiram and elesclomol, as well as selectivity, since it is clear these compounds will not be an effective therapeutic strategy for all cancer patients. In this work, GSEA highlights the proteasome as one of the pathways that is significantly positively correlated with sensitivity and this is in line with the identification of Ubiquitin Conjugating Enzyme E2 C (UBE2C) at the single gene level. Furthermore, this fits mechanistically with the mechanism I propose to be responsible for the cytotoxic activity of the compounds- loss of proteostasis. GSEA also highlighted several pathways that are negatively associated with disulfiram and elesclomol sensitivity, mostly relating to metabolism. Identifying any overlap and pathway cross-talk between metabolism and proteostasis will lead to a better

understanding of the pathways involved in sensitivity and provide patient stratification criteria for future clinical trials.

In conclusion, this body of work validates a high quality, high content assay platform for the discovery of new therapeutic agents and strategies for the treatment of OAC, identifies several repurposing opportunities, and provides new mechanistic insights into the universal mechanism of action of ammonium pyrrolidinedithiocarbamate, disulfiram and elesclomol, and potential patient stratification hypotheses for future clinical trials.

# 8| *References*


Abraham, E. P. and Chain, E. (1940) 'An enzyme from bacteria able to destroy penicillin', *Nature*. Nature Publishing Group, p. 837. doi: 10.1038/146837a0.

Absi, A., Adelstein, D. J. and Rice, T. (2010) *Esophageal Cancer*. Available at: http://www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/hematology-oncology/esophageal-cancer/ (Accessed: 1 November 2017).

Aguirre-Plans, J. *et al.* (2018) 'Proximal pathway enrichment analysis for targeting comorbid diseases via network endopharmacology', *Pharmaceuticals*. MDPI AG, 11(3). doi: 10.3390/ph11030061.

Alderson, D. *et al.* (2017) 'Neoadjuvant cisplatin and fluorouracil versus epirubicin, cisplatin, and capecitabine followed by resection in patients with oesophageal adenocarcinoma (UK MRC OE05): an open-label, randomised phase 3 trial', *The Lancet Oncology*, 18(9), pp. 1249–1260. doi: 10.1016/S1470-2045(17)30447-3.

Allum, W. H. *et al.* (2011) 'Guidelines for the management of oesophageal and gastric cancer', *Gut*, 60(11), pp. 1449–1472. doi: 10.1136/gut.2010.228254.

Ando, D. M., McLean, C. Y. and Berndl, M. (2017) 'Improving Phenotypic Measurements in High-Content Imaging Screens', *bioRxiv*. bioRxiv, p. 161422. doi: 10.1101/161422.

Atkinson, A. J. *et al.* (2001) 'Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework', *Clinical Pharmacology & Therapeutics*. John Wiley & Sons, Ltd, 69(3), pp. 89–95. doi: 10.1067/mcp.2001.113989.

Bang, Y.-J. *et al.* (2010) 'Trastuzumab in combination with chemotherapy versus

chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial', *The Lancet*, 376(9742), pp. 687–697. doi: 10.1016/S0140-6736(10)61121-X.

Barbi de Moura, M. *et al.* (2012) 'Mitochondrial Respiration - An Important Therapeutic Target in Melanoma', *PLoS ONE*. Edited by J. Santos. Public Library of Science, 7(8), p. e40690. doi: 10.1371/journal.pone.0040690.

Benfey, P. N. and Mitchell-Olds, T. (2008) 'From genotype to phenotype: Systems biology meets natural variation', *Science*. NIH Public Access, pp. 495–497. doi: 10.1126/science.1153716.

Benjamini, Y. and Yekutieli, D. (2001) 'The control of the false discovery rate in multiple testing under dependency', *Annals of Statistics*. Institute of Mathematical Statistics, 29(4), pp. 1165–1188. doi: 10.1214/aos/1013699998.

Bickle, M. (2010) 'The beautiful cell: high-content screening in drug discovery', *Analytical and Bioanalytical Chemistry*. Springer-Verlag, 398(1), pp. 219–226. doi: 10.1007/s00216-010-3788-3.

Birmingham, A. *et al.* (2009) 'Statistical methods for analysis of high-throughput RNA interference screens', *Nature Methods*, 6(8), pp. 569–575. doi: 10.1038/nmeth.1351.

Blackman, R. K. *et al.* (2012) 'Mitochondrial Electron Transport Is the Cellular Target of the Oncology Drug Elesclomol', *PLoS ONE*. Edited by J. Santos. Public Library of Science, 7(1), p. e29798. doi: 10.1371/journal.pone.0029798.

Bose, K. *et al.* (2017) 'Perioperative Therapy of Oesophagogastric Adenocarcinoma: Mainstay and Future Directions', *Gastroenterology Research and Practice*. Hindawi Limited. doi: 10.1155/2017/5651903.

Bray, M.-A. *et al.* (2016) 'Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes.', *Nature protocols*. NIH Public Access, 11(9), pp. 1757–74. doi: 10.1038/nprot.2016.105.

Caicedo, J. C. *et al.* (2017) 'Data-analysis strategies for image-based cell profiling', *Nature Methods*, 14(9). doi: 10.1038/nmeth.4397.

Caie, P. D. *et al.* (2010) 'High-content phenotypic profiling of drug response signatures across distinct cancer cells.', *Molecular cancer therapeutics*. American Association for Cancer Research, 9(6), pp. 1913–26. doi: 10.1158/1535-7163.MCT-09-1148.

Campbell, J. *et al.* (2016) 'Large-Scale Profiling of Kinase Dependencies in Cancer Cell Lines', *Cell Reports*. Elsevier B.V., 14(10), pp. 2490–2501. doi: 10.1016/j.celrep.2016.02.023.

Carlborg, Ö. and Haley, C. S. (2004) 'Epistasis: Too often neglected in complex trait studies?', *Nature Reviews Genetics*, pp. 618–625. doi: 10.1038/nrg1407.

Carpenter, A. E. *et al.* (2006) 'CellProfiler: image analysis software for identifying and quantifying cell phenotypes.', *Genome Biology*, 7(10), p. R100. doi: 10.1186/gb-2006-7-10-r100.

Carrasco-Pozo, C. *et al.* (2008) 'Double edge redox-implications for the interaction between endogenous thiols and copper ions: In vitro studies', *Bioorganic and Medicinal Chemistry*. Pergamon, 16(22), pp. 9795–9803. doi: 10.1016/j.bmc.2008.09.068.

Carson, P. E. *et al.* (1956) 'Enzymatic deficiency in primaquine-sensitive erythrocytes', *Science*. Science, 124(3220), pp. 484–485. doi: 10.1126/science.124.3220.484-a.

Chapman, P. B. *et al.* (2011) 'Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation', *New England Journal of Medicine*. New England Journal of Medicine (NEJM/MMS), 364(26), pp. 2507–2516. doi: 10.1056/NEJMoa1103782.

Chen, S. *et al.* (2013) 'Syntheses and antitumor activities of N'1,N'3-dialkyl-N'1,N'3-di-(alkylcarbonothioyl) malonohydrazide: The discovery of elesclomol', *Bioorganic & Medicinal Chemistry Letters*, 23(18), pp. 5070–5076. doi: 10.1016/j.bmcl.2013.07.032.

Chih-Hsin Yang, J. *et al.* (2013) 'Symptom Control and Quality of Life in LUX-Lung 3: A Phase III Study of Afatinib or Cisplatin/Pemetrexed in Patients With Advanced

Lung Adenocarcinoma With EGFR Mutations', *J Clin Oncol*, p. 31. doi: 10.1200/JCO.2012.44.2806.

Chong, I. Y. *et al.* (2018) 'Mapping genetic vulnerabilities reveals BTK as a novel therapeutic target in oesophageal cancer', *Gut*, 67, pp. 1780–1792. doi: 10.1136/gutjnl-2017-314408.

Chung, S. M. *et al.* (2007) 'p53 in esophageal adenocarcinoma: a critical reassessment of mutation frequency and identification of 72Arg as the dominant allele.', *International journal of oncology*, 31(6), pp. 1351–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17982662 (Accessed: 11 May 2017).

Cong, L. *et al.* (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science*. American Association for the Advancement of Science, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.

Contino, G. *et al.* (2016) 'Whole-genome sequencing of nine esophageal adenocarcinoma cell lines', *F1000Research*, 5, p. 1336. doi: 10.12688/f1000research.7033.1.

Croft, D. *et al.* (2014) 'The Reactome pathway knowledgebase', *Nucleic Acids Research*. Oxford Academic, 42(D1), pp. D472–D477. doi: 10.1093/nar/gkt1102.

Croston, G. E. (2017) 'The utility of target-based discovery', *Expert Opinion on Drug Discovery*. Taylor and Francis Ltd, 12(5), pp. 427–429. doi: 10.1080/17460441.2017.1308351.

Dalecki, A. G. *et al.* (2015) 'Disulfiram and copper ions kill Mycobacterium tuberculosis in a synergistic manner', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology, 59(8), pp. 4835–4844. doi: 10.1128/AAC.00692-15.

Dawson, J. C. *et al.* (2020) 'A synergistic anticancer FAK and HDAC inhibitor combination discovered by a novel chemical–genetic high-content phenotypic screen', *Molecular Cancer Therapeutics*. American Association for Cancer Research Inc., 19(2), pp. 637–649. doi: 10.1158/1535-7163.MCT-19-0330.

Drews, J. (2000) 'Drug discovery: A historical perspective', *Science*. American

Association for the Advancement of Science, pp. 1960–1964. doi: 10.1126/science.287.5460.1960.

Du, Y. *et al.* (2020) 'Development of a miniaturized 3D organoid culture platform for ultra-high-throughput screening', *Journal of Molecular Cell Biology*. Edited by X. Yao. Oxford University Press, 12(8), pp. 630–643. doi: 10.1093/jmcb/mjaa036.

Duggan, S. P. *et al.* (2018) 'siRNA Library Screening Identifies a Druggable Immune-Signature Driving Esophageal Adenocarcinoma Cell Growth', *CMGH*. Elsevier Inc, 5(4), pp. 569–590. doi: 10.1016/j.jcmgh.2018.01.012.

Dulak, A. M. *et al.* (2012) 'Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis.', *Cancer research*, 72(17), pp. 4383–93. doi: 10.1158/0008-5472.CAN-11-3893.

Dutton, S. J. *et al.* (2014) 'Gefitinib for oesophageal cancer progressing after chemotherapy (COG): a phase 3, multicentre, double-blind, placebo-controlled randomised trial', *The Lancet Oncology*, 15(8), pp. 894–904. doi: 10.1016/S1470-2045(14)70024-5.

Echeverri, C. J. *et al.* (2006) 'Minimizing the risk of reporting false positives in large-scale RNAi screens', *Nature Methods*. Nat Methods, 3(10), pp. 777–779. doi: 10.1038/nmeth1006-777.

Eder, J., Sedrani, R. and Wiesmann, C. (2014) 'The discovery of first-in-class drugs: Origins and evolution', *Nature Reviews Drug Discovery*. Nature Publishing Group, 13(8), pp. 577–587. doi: 10.1038/nrd4336.

Erl, W., Weber, C. and Hansson, G. K. (2000) 'Pyrrolidine dithiocarbamate-induced apoptosis depends on cell type, density, and the presence of Cu2+ and Zn2+', *American Journal of Physiology - Cell Physiology*. American Physiological Society, 278(6 47-6). doi: 10.1152/ajpcell.2000.278.6.c1116.

Farmer, H. *et al.* (2005) 'Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy', *Nature*. Nature Publishing Group, 434(7035), pp. 917–921. doi: 10.1038/nature03445.

Fisher, O. M. *et al.* (2017) 'The prognostic value of TP53 mutations in oesophageal adenocarcinoma: a systematic review and meta-analysis', *Gut*, 66(3), pp. 399–410. doi: 10.1136/gutjnl-2015-310888.

Fitzmaurice, C. *et al.* (2015) 'The Global Burden of Cancer 2013', *JAMA Oncology*. American Medical Association, 1(4), pp. 505–527. doi: 10.1001/jamaoncol.2015.0735.

Flaherty, K. T. *et al.* (2012) 'Improved Survival with MEK Inhibition in BRAF-Mutated Melanoma', *New England Journal of Medicine*. Massachusetts Medical Society, 367(2), pp. 107–114. doi: 10.1056/NEJMoa1203421.

Frankell, A. M. *et al.* (2019) 'The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic', *Nature Genetics*. Nature Publishing Group, 51(3), pp. 506–516. doi: 10.1038/s41588-018-0331-5.

Freedman, J. H., Ciriolo, M. R. and Peisach, J. (1989) 'The role of glutathione in copper metabolism and toxicity', *Journal of Biological Chemistry*. Elsevier, 264(10), pp. 5598–5605. doi: 10.1016/s0021-9258(18)83589-x.

Garcia-Santamarina, S. *et al.* (2017) 'Cryptococcus neoformans iron-sulfur protein biogenesis machinery is a novel layer of protection against Cu stress', *mBio*. American Society for Microbiology, 8(5). doi: 10.1128/mBio.01742-17.

Garnett, M. J. *et al.* (2012) 'Systematic identification of genomic markers of drug sensitivity in cancer cells', *Nature*. Nature Publishing Group, 483(7391), pp. 570–575. doi: 10.1038/nature11005.

Geiss, G. K. *et al.* (2008) 'Direct multiplexed measurement of gene expression with color-coded probe pairs', *Nature Biotechnology*. Nature Publishing Group, 26(3), pp. 317–325. doi: 10.1038/nbt1385.

Geoffrey Burns, C. *et al.* (2005) 'High-Throughput Assay for Small Molecules That Modulate Zebrafish Embryonic Heart Rate', *Nature Chemical Biology*. Nature Publishing Group, 1(5), pp. 263–264. doi: 10.1038/nchembio732.

Gibson, M. K. *et al.* (2013) 'Barrett's esophagus: cancer and molecular biology', *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc., 1300(1), pp.

296–314. doi: 10.1111/nyas.12252.

Giuliano, K. A. *et al.* (1997) 'High-Content Screening: A New Approach to Easing Key Bottlenecks in the Drug Discovery Process', *Journal of Biomolecular Screening*. Mary Ann Liebert Inc., 2(4), pp. 249–259. doi: 10.1177/108705719700200410.

Godinez, W. J. *et al.* (2017) 'A multi-scale convolutional neural network for phenotyping high-content cellular images', *Bioinformatics*. Oxford University Press, 33(13), pp. 2010–2019. doi: 10.1093/bioinformatics/btx069.

Godinez, W. J., Hossain, I. and Zhang, X. (2018) 'Unsupervised phenotypic analysis of cellular images with multi-scale convolutional neural networks', *bioRxiv*. bioRxiv, p. 361410. doi: 10.1101/361410.

Gonçalves, E. *et al.* (2020) 'Drug mechanism-of-action discovery through the integration of pharmacological and <scp>CRISPR</scp> screens', *Molecular Systems Biology*. EMBO, 16(7), p. e9405. doi: 10.15252/msb.20199405.

Guillem, P. G. (2005) 'How to Make a Barrett Esophagus: Pathophysiology of Columnar Metaplasia of the Esophagus', *Digestive Diseases and Sciences*, 50(3), pp. 415–424. doi: 10.1007/s10620-005-2451-x.

Gustafsdottir, S. M. *et al.* (2013) 'Multiplex Cytological Profiling Assay to Measure Diverse Cellular States', *PLoS ONE*. Edited by M. A. Mancini. Public Library of Science, 8(12), p. e80999. doi: 10.1371/journal.pone.0080999.

Guthrie, L. M. *et al.* (2020) 'Elesclomol alleviates Menkes pathology and mortality by escorting Cu to cuproenzymes in mice', *Science*. American Association for the Advancement of Science, 368(6491), pp. 620–625. doi: 10.1126/science.aaz8899.

Hald, J. and Jacobsen, E. (1948) 'A Drug Sensitising the Organism to Ethyl Alcohol', *The Lancet*. Elsevier, 252(6539), pp. 1001–1004. doi: 10.1016/S0140-6736(48)91514-1.

Hameeteman, W. *et al.* (1989) 'Barrett's Esophagus; Development of Dysplasia and Adenocarcinoma', *Gastroenterology*, 96(5), pp. 1249–1256. doi: 10.1016/S0016-5085(89)80011-3.

Hamilton, J. P. *et al.* (2016) 'Activation of liver X receptor/retinoid X receptor pathway ameliorates liver disease in Atp7B−/− (Wilson disease) mice', *Hepatology*. John Wiley and Sons Inc., 63(6), pp. 1828–1841. doi: 10.1002/hep.28406.

Hart, C. P. (2005) 'Finding the target after screening the phenotype', *Drug Discovery Today*. Elsevier Current Trends, pp. 513–519. doi: 10.1016/S1359-6446(05)03415-X.

Hasinoff, B. B. *et al.* (2014) 'The cytotoxicity of the anticancer drug elesclomol is due to oxidative stress indirectly mediated through its complex with Cu(II)', *Journal of Inorganic Biochemistry*. Elsevier Inc., 137, pp. 22–30. doi: 10.1016/j.jinorgbio.2014.04.004.

Hauschild, A. *et al.* (2009) 'Phase III, randomized, double-blind study of elesclomol and paclitaxel versus paclitaxel alone in stage IV metastatic melanoma (MM)', *Journal of Clinical Oncology*. American Society of Clinical Oncology (ASCO), 27(18_suppl), pp. LBA9012–LBA9012. doi: 10.1200/jco.2009.27.18_suppl.lba9012.

Hauschild, A. *et al.* (2012) 'Dabrafenib in BRAF-mutated metastatic melanoma: A multicentre, open-label, phase 3 randomised controlled trial', *The Lancet*. Lancet Publishing Group, 380(9839), pp. 358–365. doi: 10.1016/S0140-6736(12)60868-X.

Hecht, J. R. *et al.* (2016) 'Lapatinib in Combination With Capecitabine Plus Oxaliplatin in Human Epidermal Growth Factor Receptor 2–Positive Advanced or Metastatic Gastric, Esophageal, or Gastroesophageal Adenocarcinoma: TRIO-013/LOGiC—A Randomized Phase III Trial', *Journal of Clinical Oncology*, 34(5), pp. 443–451. doi: 10.1200/JCO.2015.62.6598.

Hernández-Lemus, E. and Martínez-García, M. (2021) 'Pathway-Based Drug-Repurposing Schemes in Cancer: The Role of Translational Bioinformatics', *Frontiers in Oncology*. Frontiers Media SA, 10, p. 2996. doi: 10.3389/fonc.2020.605680.

Herszényi, L. and Tulassay, Z. (2010) 'Epidemiology of gastrointestinal and liver tumors.', *European review for medical and pharmacological sciences*, 14(4), pp. 249–58. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20496531 (Accessed: 1 November 2017).

Hopkins, A. L. (2008) 'Network pharmacology: The next paradigm in drug discovery', *Nature Chemical Biology*. Nature Publishing Group, pp. 682–690. doi: 10.1038/nchembio.118.

Hughes, R. E. *et al.* (2020) 'High-Content Phenotypic Profiling in Esophageal Adenocarcinoma Identifies Selectively Active Pharmacological Classes of Drugs for Repurposing and Chemical Starting Points for Novel Drug Discovery', *SLAS Discovery*. SAGE Publications Inc., 25(7), pp. 770–782. doi: 10.1177/2472555220917115.

Hughes, T. R. *et al.* (2000) 'Functional discovery via a compendium of expression profiles', *Cell*. Cell Press, 102(1), pp. 109–126. doi: 10.1016/S0092-8674(00)00015-5.

Jackson, A. L. *et al.* (2006) 'Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity', *RNA*. RNA, 12(7), pp. 1179–1187. doi: 10.1261/rna.25706.

Jacobsen, E. (1952) 'The Metabolism of Ethyl Alcohol', *Nature*, 169(4303). doi: 10.1038/169645a0.

Janssens, R. *et al.* (2020) 'Fully unsupervised deep mode of action learning for phenotyping high-content cellular images', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.07.22.215459. doi: 10.1101/2020.07.22.215459.

Jinek, M. *et al.* (2012) 'A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity', *Science*. American Association for the Advancement of Science, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.

Kamath, R. S. and Ahringer, J. (2003) 'Genome-wide RNAi screening in Caenorhabditis elegans', *Methods*. Academic Press Inc., 30(4), pp. 313–321. doi: 10.1016/S1046-2023(03)00050-1.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*. Oxford University Press, pp. 27–30. doi: 10.1093/nar/28.1.27.

Kang, R., Kroemer, G. and Tang, D. (2019) 'The tumor suppressor protein p53 and the ferroptosis network', *Free Radical Biology and Medicine*, 133, pp. 162–168. doi:

10.1016/j.freeradbiomed.2018.05.074.

Karasic, T. B. *et al.* (2018) 'Phase II trial of palbociclib in patients with advanced esophageal or gastric cancer.', *Journal of Clinical Oncology*. American Society of Clinical Oncology (ASCO), 36(4_suppl), pp. 68–68. doi: 10.1200/jco.2018.36.4_suppl.68.

Karasic, T. B. *et al.* (2020) ' Phase II Trial of Palbociclib in Patients with Advanced Esophageal or Gastric Cancer ', *The Oncologist*. Wiley, 25(12). doi: 10.1634/theoncologist.2020-0681.

Kaufman, B. *et al.* (2015) 'Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation', *Journal of Clinical Oncology*. American Society of Clinical Oncology, 33(3), pp. 244–250. doi: 10.1200/JCO.2014.56.2728.

Kiger, A. A. *et al.* (2003) 'A functional genomic analysis of cell morphology using RNA interference', *Journal of Biology*. J Biol, 2(4). doi: 10.1186/1475-4924-2-27.

Kirshner, J. R. *et al.* (2008) 'Elesclomol induces cancer cell apoptosis through oxidative stress', *Molecular Cancer Therapeutics*. American Association for Cancer Research, 7(8), pp. 2319–2327. doi: 10.1158/1535-7163.MCT-08-0298.

Kragh, H. S. (2008) 'From Disulfiram to Antabuse: The Invention of a Drug', *Bulletin for the History of Chemistry*, 33(2), pp. 82–88.

Krämer, A. *et al.* (2014) 'Causal analysis approaches in Ingenuity Pathway Analysis', *Bioinformatics*. Oxford Academic, 30(4), pp. 523–530. doi: 10.1093/bioinformatics/btt703.

Kraus, O. Z., Ba, J. L. and Frey, B. J. (2016) 'Classifying and segmenting microscopy images with deep multiple instance learning', *Bioinformatics*. Oxford University Press, 32(12), pp. i52–i59. doi: 10.1093/bioinformatics/btw252.

Kummel, A. *et al.* (2010) 'Integration of Multiple Readouts into the Z' Factor for Assay Quality Assessment', *Journal of Biomolecular Screening*. SAGE PublicationsSage CA: Los Angeles, CA, 15(1), pp. 95–101. doi: 10.1177/1087057109351311.

Laffitte, B. A. *et al.* (2003) 'Activation of liver X receptor improves glucose tolerance through coordinate regulation of glucose metabolism in liver and adipose tissue', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 100(9), pp. 5419–5424. doi: 10.1073/pnas.0830671100.

Lamb, J. *et al.* (2006) 'The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease', *Science*. Science, 313(5795), pp. 1929–1935. doi: 10.1126/science.1132939.

Lansing Taylor, D. (2010) 'A personal perspective on high-content screening (HCS): From the beginning', *Journal of Biomolecular Screening*. SAGE PublicationsSage CA: Los Angeles, CA, pp. 720–725. doi: 10.1177/1087057110374995.

Lapins, M. and Spjuth, O. (2019) 'Evaluation of Gene Expression and Phenotypic Profiling Data as Quantitative Descriptors for Predicting Drug Targets and Mechanisms of Action', *bioRxiv*. bioRxiv, p. 580654. doi: 10.1101/580654.

Laurencikiene, J. and Rydén, M. (2012) 'Liver X receptors and fat cell metabolism', *International Journal of Obesity*. Nature Publishing Group, pp. 1494–1502. doi: 10.1038/ijo.2012.21.

Lee, J. M., Ledermann, J. A. and Kohn, E. C. (2014) 'PARP inhibitors for BRCA1/2 mutation-associated and BRCA-like malignancies', *Annals of Oncology*. Oxford University Press, 25(1), pp. 32–40. doi: 10.1093/annonc/mdt384.

Lei, R. *et al.* (2016) 'Transferrin receptor facilitates TGF-β and BMP signaling activation to control craniofacial morphogenesis', *Cell Death and Disease*. Nature Publishing Group, 7(6), pp. e2282–e2282. doi: 10.1038/cddis.2016.170.

Lepage, C. *et al.* (2008) 'Continuing Rapid Increase in Esophageal Adenocarcinoma in England and Wales', *The American Journal of Gastroenterology*, 103(11), pp. 2694–2699. doi: 10.1111/j.1572-0241.2008.02191.x.

Li, X. *et al.* (2018) 'Organoid cultures recapitulate esophageal adenocarcinoma heterogeneity providing a model for clonality studies and precision therapeutics', *Nature Communications*. Nature Publishing Group, 9(1), p. 2983. doi:

10.1038/s41467-018-05190-9.

Li, Y. *et al.* (2020) 'Predicting Tumor Response to Drugs based on Gene-Expression Biomarkers of Sensitivity Learned from Cancer Cell Lines', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.07.03.180620. doi: 10.1101/2020.07.03.180620.

Liberzon, A. *et al.* (2015) 'The Molecular Signatures Database Hallmark Gene Set Collection', *Cell Systems*. Cell Press, 1(6), pp. 417–425. doi: 10.1016/j.cels.2015.12.004.

Lin, X. *et al.* (2012) 'Life beyond kinases: Structure-based discovery of sorafenib as nanomolar antagonist of 5-HT receptors', *Journal of Medicinal Chemistry*. American Chemical Society, 55(12), pp. 5749–5759. doi: 10.1021/jm300338m.

Liu, D. S. *et al.* (2017) 'Inhibiting the system xC- /glutathione axis selectively targets cancers with mutant-p53 accumulation', *Nature Communications*. Nature Publishing Group, 8(1), pp. 1–14. doi: 10.1038/ncomms14844.

Ljosa, V. *et al.* (2013) 'Comparison of Methods for Image-Based Profiling of Cellular Morphological Responses to Small-Molecule Treatment', *Journal of Biomolecular Screening*. SAGE PublicationsSage CA: Los Angeles, CA, 18(10), pp. 1321–1329. doi: 10.1177/1087057113503553.

Lohse, I. *et al.* (2018) 'Ex vivo drug sensitivity testing as a means for drug repurposing in esophageal adenocarcinoma', *PLOS ONE*. Edited by J. S. Castresana. Public Library of Science, 13(9), p. e0203173. doi: 10.1371/journal.pone.0203173.

Lövborg, H. *et al.* (2006) 'Inhibition of proteasome activity, nuclear factor-KB translocation and cell survival by the antialcoholism drug disulfiram', *International Journal of Cancer*, 118(6), pp. 1577–1580. doi: 10.1002/ijc.21534.

Lu, J.-J. *et al.* (2012) 'Multi-Target Drugs: The Trend of Drug Research and Development', *PLoS ONE*. Edited by B. B. Aggarwal. Public Library of Science, 7(6), p. e40262. doi: 10.1371/journal.pone.0040262.

Luo, D. X. *et al.* (2010) 'A novel model of cholesterol efflux from lipid-loaded cells', *Acta Pharmacologica Sinica*. Nature Publishing Group, pp. 1243–1257. doi: 10.1038/aps.2010.93.

Mahalanobis, P. C. (1936) ‘On the generalized distance in statistics.’, *Proceedings of the National Institute of Sciences of India*, 2(1), pp. 49–55.

Mali, P. *et al.* (2013) ‘RNA-guided human genome engineering via Cas9’, *Science*. American Association for the Advancement of Science, 339(6121), pp. 823–826. doi: 10.1126/science.1232033.

Martensen-Larsen, O. (1948) ‘Treatment of Alcoholism with a Sensitising Drug’, *The Lancet*. Elsevier, 252(6539), pp. 1004–1005. doi: 10.1016/S0140-6736(48)91515-3.

McDermott, U. (2015) ‘Next-generation sequencing and empowering personalised cancer medicine’, *Drug Discovery Today*. Elsevier Ltd, pp. 1470–1475. doi: 10.1016/j.drudis.2015.10.008.

McFarland, J. M. *et al.* (2020) ‘Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action’, *Nature Communications*. Nature Research, 11(1), pp. 1–15. doi: 10.1038/s41467-020-17440-w.

Messager, M. *et al.* (2016) ‘Variations among 5 European countries for curative treatment of resectable oesophageal and gastric cancer: A survey from the EURECCA Upper GI Group (EUropean REgistration of Cancer CAre)’, *European Journal of Surgical Oncology (EJSO)*, 42(1), pp. 116–122. doi: 10.1016/j.ejso.2015.09.017.

Mi, L. *et al.* (2010) ‘A cautionary note on using N-acetylcysteine as an antagonist to assess isothiocyanate-induced reactive oxygen species-mediated apoptosis’, *Analytical Biochemistry*. Academic Press Inc., 405(2), pp. 269–271. doi: 10.1016/j.ab.2010.06.015.

Milioli, H. H. *et al.* (2015) ‘The discovery of novel biomarkers improves breast cancer intrinsic subtype prediction and reconciles the labels in the METABRIC data set’, *PLoS ONE*. Public Library of Science, 10(7). doi: 10.1371/journal.pone.0129711.

Moffat, J. G. *et al.* (2017) ‘Opportunities and challenges in phenotypic drug discovery: An industry perspective’, *Nature Reviews Drug Discovery*. Nature Publishing Group, pp. 531–543. doi: 10.1038/nrd.2017.111.

Moffat, J. G., Rudolph, J. and Bailey, D. (2014) 'Phenotypic screening in cancer drug discovery-past, present and future', *Nature Reviews Drug Discovery*. Nature Publishing Group, 13(8), pp. 588–602. doi: 10.1038/nrd4366.

Mostofa, M. G., Seraj, Z. I. and Fujita, M. (2014) 'Exogenous sodium nitroprusside and glutathione alleviate copper toxicity by reducing copper uptake and oxidative damage in rice (Oryza sativa L.) seedlings', *Protoplasma*. Springer-Verlag Wien, 251(6), pp. 1373–1386. doi: 10.1007/s00709-014-0639-7.

Nagai, M. *et al.* (2012) 'The oncology drug elesclomol selectively transports copper to the mitochondria to induce oxidative stress in cancer cells', *Free Radical Biology and Medicine*. Pergamon, 52(10), pp. 2142–2150. doi: 10.1016/J.FREERADBIOMED.2012.03.017.

Napier, K. J., Scheerer, M. and Misra, S. (2014) 'Esophageal cancer: A Review of epidemiology, pathogenesis, staging workup and treatment modalities.', *World journal of gastrointestinal oncology*. Baishideng Publishing Group Inc, 6(5), pp. 112–20. doi: 10.4251/wjgo.v6.i5.112.

Nebert, D. W. (1999) 'Pharmacogenetics and pharmacogenomics: why is this relevant to the clinical geneticist?', *Clinical Genetics*. John Wiley & Sons, Ltd, 56(4), pp. 247–258. doi: 10.1034/j.1399-0004.1999.560401.x.

Nguyen, L., Dang, C. C. and Ballester, P. J. (2017) 'Systematic assessment of multi-gene predictors of pan-cancer cell line sensitivity to drugs exploiting gene expression data', *F1000Research*. F1000 Research Ltd, 5, p. 2927. doi: 10.12688/f1000research.10529.2.

Nishimura, D. (2001) 'BioCarta', *Biotech Software & Internet Report*. Mary Ann Liebert Inc, 2(3), pp. 117–120. doi: 10.1089/152791601750294344.

Noble, F. *et al.* (2017) 'Multicentre cohort study to define and validate pathological assessment of response to neoadjuvant therapy in oesophagogastric adenocarcinoma', *British Journal of Surgery*, 104(13), pp. 1816–1828. doi: 10.1002/bjs.10627.

O'Neill, J. R. *et al.* (2017) 'Patients unfit for neoadjuvant therapy may still undergo

resection of locally advanced esophageal or esophagogastric junctional cancer with acceptable oncological results.', *International journal of surgery. Oncology*. Wolters Kluwer Health, 2(2), p. e09. doi: 10.1097/IJ9.0000000000000009.

Offen, D. *et al.* (2004) 'A low molecular weight copper chelator crosses the blood-brain barrier and attenuates experimental autoimmune encephalomyelitis', *Journal of Neurochemistry*. J Neurochem, 89(5), pp. 1241–1251. doi: 10.1111/j.1471-4159.2004.02428.x.

Ogawa, F. *et al.* (2006) 'The repair enzyme peptide methionine-S-sulfoxide reductase is expressed in human epidermis and upregulated by UVA radiation', *Journal of Investigative Dermatology*. Elsevier, 126(5), pp. 1128–1134. doi: 10.1038/sj.jid.5700116.

Okines, A. F. C. *et al.* (2010) 'Epirubicin, Oxaliplatin, and Capecitabine With or Without Panitumumab for Advanced Esophagogastric Cancer: Dose-Finding Study for the Prospective Multicenter, Randomized, Phase II/III REAL-3 Trial', *Journal of Clinical Oncology*, 28(25), pp. 3945–3950. doi: 10.1200/JCO.2010.29.2847.

Otsu, N. (1979) 'Threshold selection method from gray-level histograms', *IEEE Trans Syst Man Cybern*, SMC-9(1), pp. 62–66. doi: 10.1109/tsmc.1979.4310076.

Padmanabhan, K., Eddy, W. F. and Crowley, J. C. (2010) 'A novel algorithm for optimal image thresholding of biological data', *Journal of Neuroscience Methods*. J Neurosci Methods, 193(2), pp. 380–384. doi: 10.1016/j.jneumeth.2010.08.031.

Paterson, A. L. *et al.* (2013) 'Characterization of the timing and prevalence of receptor tyrosine kinase expression changes in oesophageal carcinogenesis', *The Journal of Pathology*. John Wiley & Sons, Ltd, 230(1), pp. 118–128. doi: 10.1002/path.4044.

Pennathur, A. *et al.* (2013) 'Oesophageal carcinoma', *The Lancet*. British Society of Gastroenterology, London, 381(9864), pp. 400–412. doi: 10.1016/S0140-6736(12)60643-6.

Perlman, Z. E. *et al.* (2004) 'Multidimensional Drug Profiling By Automated Microscopy', *Science*, 306(5699), pp. 1194–1198. doi: 10.1126/science.1100709.

Pirmohamed, M. (2001) 'Pharmacogenetics and pharmacogenomics', *British Journal of Clinical Pharmacology*. Wiley-Blackwell, pp. 345–347. doi: 10.1046/j.0306-5251.2001.01498.x.

Prutz, W. A. (1994) 'Interaction between glutathione and Cu(II) in the vicinity of nucleic acids', *Biochemical Journal*. Portland Press Ltd, 302(2), pp. 373–382. doi: 10.1042/bj3020373.

Rees, M. G. *et al.* (2016) 'Correlating chemical sensitivity and basal gene expression reveals mechanism of action', *Nature Chemical Biology*. Nature Publishing Group, 12(2), pp. 109–116. doi: 10.1038/nchembio.1986.

Renner, H. *et al.* (2020) 'A fully automated high-throughput workflow for 3d-based chemical screening in human midbrain organoids', *eLife*. eLife Sciences Publications Ltd, 9, pp. 1–39. doi: 10.7554/eLife.52904.

Reynolds, J. *et al.* (2017) 'ICORG 10-14: NEOadjuvant trial in Adenocarcinoma of the oEsophagus and oesophagoGastric junction International Study (Neo-AEGIS)', *BMC cancer*. BioMed Central, 17(1), pp. 401–401. doi: 10.1186/S12885-017-3386-2.

Robert, C. *et al.* (2015) 'Improved Overall Survival in Melanoma with Combined Dabrafenib and Trametinib', *New England Journal of Medicine*. Massachusetts Medical Society, 372(1), pp. 30–39. doi: 10.1056/NEJMoa1412690.

Rockfield, S. *et al.* (2018) 'Links between iron and lipids: Implications in some major human diseases', *Pharmaceuticals*. MDPI AG. doi: 10.3390/ph11040113.

Rose, F. *et al.* (2018) 'Compound Functional Prediction Using Multiple Unrelated Morphological Profiling Assays', *SLAS TECHNOLOGY: Translating Life Sciences Innovation*. SAGE PublicationsSage CA: Los Angeles, CA, 23(3), pp. 243–251. doi: 10.1177/2472630317740831.

Saporito-Magriñá, C. M. *et al.* (2018) 'Copper-induced cell death and the protective role of glutathione: The implication of impaired protein folding rather than oxidative stress', *Metallomics*. Royal Society of Chemistry, 10(12), pp. 1743–1754. doi: 10.1039/c8mt00182k.

Schirle, M. and Jenkins, J. L. (2016) 'Identifying compound efficacy targets in

phenotypic drug discovery', *Drug Discovery Today*. Elsevier Ltd, pp. 82–89. doi: 10.1016/j.drudis.2015.08.001.

Schneidewind, T. *et al.* (2020) 'Morphological Profiling Identifies a Common Mode of Action for Small Molecules with Different Targets', *ChemBioChem*. Wiley-VCH Verlag, 21(22), pp. 3197–3207. doi: 10.1002/cbic.202000381.

Schreck, R. *et al.* (1992) 'Dithiocarbamates as potent inhibitors of nuclear factor κb activation in intact cells', *Journal of Experimental Medicine*. The Rockefeller University Press, 175(5), pp. 1181–1194. doi: 10.1084/jem.175.5.1181.

Secrier, M. *et al.* (2016) 'Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance', *Nature Genetics*. Nature Research, 48(10), pp. 1131–1141. doi: 10.1038/ng.3659.

Sequist, L. V. *et al.* (2013) 'Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations', *Journal of Clinical Oncology*. American Society of Clinical Oncology, 31(27), pp. 3327–3334. doi: 10.1200/JCO.2012.44.2806.

Shalem, O. *et al.* (2014) 'Genome-scale CRISPR-Cas9 knockout screening in human cells', *Science*. American Association for the Advancement of Science, 343(6166), pp. 84–87. doi: 10.1126/science.1247005.

Sideri, T. C., Willetts, S. A. and Avery, S. V. (2009) 'Methionine sulphoxide reductases protect iron-sulphur clusters from oxidative inactivation in yeast', *Microbiology*. NIH Public Access, 155(2), pp. 612–623. doi: 10.1099/mic.0.022665-0.

Snyder, L. H. (1932) 'Studies in Human Inheritance. IX, The Inheritance of Taste Deficiency in Man', *The Ohio Journal of Science*, 32(5), pp. 436–440. Available at: https://kb.osu.edu/handle/1811/2588 (Accessed: 11 February 2021).

Soma, S. *et al.* (2018) 'Elesclomol restores mitochondrial function in genetic models of copper deficiency', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(32), pp. 8161–8166. doi: 10.1073/pnas.1806296115.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: A knowledge-based

approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 102(43), pp. 15545–15550. doi: 10.1073/pnas.0506580102.

Subramanian, A. *et al.* (2017) 'A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles', *Cell*. Cell Press, 171(6), pp. 1437-1452.e17. doi: 10.1016/j.cell.2017.10.049.

Sudhahar, V. *et al.* (2020) 'Caveolin-1 stabilizes ATP7A, a copper transporter for extracellular SOD, in vascular tissue to maintain endothelial function', *American Journal of Physiology - Cell Physiology*. American Physiological Society, 319(5), pp. C933–C944. doi: 10.1152/ajpcell.00151.2020.

Sui, Y. and Wu, Z. (2007) 'Alternative Statistical Parameter for High-Throughput Screening Assay Quality Assessment', *Journal of Biomolecular Screening*, 12(2), pp. 229–234. doi: 10.1177/1087057106296498.

Suphavilai, C., Bertrand, D. and Nagarajan, N. (2018) 'Predicting Cancer Drug Response using a Recommender System', *Bioinformatics*. Edited by J. Wren. Oxford University Press, 34(22), pp. 3907–3914. doi: 10.1093/bioinformatics/bty452.

Swinney, D. C. and Anthony, J. (2011) 'How were new medicines discovered?', *Nature Reviews Drug Discovery*. Nature Publishing Group, 10(7), pp. 507–519. doi: 10.1038/nrd3480.

Taylor, R. C., Berendzen, K. M. and Dillin, A. (2014) 'Systemic stress signalling: Understanding the cell non-autonomous control of proteostasis', *Nature Reviews Molecular Cell Biology*. Nat Rev Mol Cell Biol, pp. 211–217. doi: 10.1038/nrm3752.

Tsherniak, A. *et al.* (2017) 'Defining a Cancer Dependency Map', *Cell*. Cell Press, 170(3), pp. 564-576.e16. doi: 10.1016/J.CELL.2017.06.010.

Tsvetkov, P. *et al.* (2019) 'Mitochondrial metabolism promotes adaptation to proteotoxic stress', *Nature Chemical Biology*. Nature Publishing Group, 15(7), pp. 681–689. doi: 10.1038/s41589-019-0291-9.

Twarog, N. R. *et al.* (2016) 'Robust Classification of Small-Molecule Mechanism of Action Using a Minimalist High-Content Microscopy Screen and Multidimensional

Phenotypic Trajectory Analysis', *PLOS ONE*. Edited by M. Bogyo. Public Library of Science, 11(2), p. e0149439. doi: 10.1371/journal.pone.0149439.

Varshney, G. K. *et al.* (2016) 'A high-throughput functional genomics workflow based on CRISPR/Cas9-mediated targeted mutagenesis in zebrafish', *Nature Protocols 2016 11:12*. Nature Publishing Group, 11(12), pp. 2357–2375. doi: 10.1038/nprot.2016.141.

Vastrik, I. *et al.* (2007) 'Reactome: A knowledge base of biologic pathways and processes', *Genome Biology*. BioMed Central, 8(3), p. R39. doi: 10.1186/gb-2007-8-3-r39.

Verma, S. *et al.* (2012) 'Trastuzumab Emtansine for HER2-Positive Advanced Breast Cancer', *New England Journal of Medicine*. New England Journal of Medicine (NEJM/MMS), 367(19), pp. 1783–1791. doi: 10.1056/NEJMoa1209124.

Vincent, F. *et al.* (2020) 'Hit Triage and Validation in Phenotypic Screening: Considerations and Strategies', *Cell Chemical Biology*. Elsevier Ltd, pp. 1332–1346. doi: 10.1016/j.chembiol.2020.08.009.

Wang, H. *et al.* (2016) 'NanoStringDiff: A novel statistical method for differential expression analysis based on NanoString nCounter data', *Nucleic Acids Research*. Oxford University Press, 44(20), p. e151. doi: 10.1093/nar/gkw677.

Wang, J. *et al.* (2018) 'Protective effect of N-acetylcysteine against oxidative stress induced by zearalenone via mitochondrial apoptosis pathway in SIEC02 cells', *Toxins*. MDPI AG, 10(10). doi: 10.3390/toxins10100407.

Wang, T. *et al.* (2014) 'Genetic Screens in Human Cells Using the CRISPR-Cas9 System', *Science*, 343(6166), pp. 80–84. doi: 10.1126/science.1246981.

Warchal, S. J., Dawson, J. C. and Carragher, N. O. (2019) 'Evaluation of Machine Learning Classifiers to Predict Compound Mechanism of Action When Transferred across Distinct Cell Lines.', *SLAS discovery : advancing life sciences R & D*. SAGE Publications Inc., 24(3), pp. 224–233. doi: 10.1177/2472555218820805.

Ward, J. H. (1963) 'Hierarchical Grouping to Optimize an Objective Function', *Journal of the American Statistical Association*, 58(301), pp. 236–244. doi:

10.1080/01621459.1963.10500845.

Waring, M. J. *et al.* (2015) 'An analysis of the attrition of drug candidates from four major pharmaceutical companies', *Nature Reviews Drug Discovery*. Nature Publishing Group, pp. 475–486. doi: 10.1038/nrd4609.

Weaver, J. M. J. *et al.* (2014) 'Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis', *Nature Genetics*. Nature Research, 46(8), pp. 837–843. doi: 10.1038/ng.3013.

Wei, D. *et al.* (2019) 'Comprehensive anticancer drug response prediction based on a simple cell line-drug complex network model', *BMC Bioinformatics*. BioMed Central Ltd., 20(1), p. 44. doi: 10.1186/s12859-019-2608-9.

Wiebelhaus, N. *et al.* (2020) 'Protein Folding Stability Changes Across the Proteome Reveal Targets of Cu Toxicity in *E. coli*', *ACS Chemical Biology*. American Chemical Society, p. acschembio.0c00900. doi: 10.1021/acschembio.0c00900.

Wilschefski, S. and Baxter, M. (2019) 'Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects', *Clinical Biochemist Reviews*. Australasian Association for Clinical Biochemistry and Laboratory Medicine, 40(3), pp. 115–133. doi: 10.33176/aacb-19-00024.

Winters, C. *et al.* (1987) 'Barrett's esophagus. A prevalent, occult complication of gastroesophageal reflux disease', *Gastroenterology*, 92(1), pp. 118–124. doi: 10.1016/0016-5085(87)90847-X.

De Wolf, H. *et al.* (2018) 'High-Throughput Gene Expression Profiles to Define Drug Similarity and Predict Compound Activity', *Assay and Drug Development Technologies*. Mary Ann Liebert Inc., 16(3), pp. 162–176. doi: 10.1089/adt.2018.845.

Wu, L. *et al.* (2011) 'LC–MS/MS and density functional theory study of copper(II) and nickel(II) chelating complexes of elesclomol (a novel anticancer agent)', *Journal of Pharmaceutical and Biomedical Analysis*. Elsevier, 54(2), pp. 331–336. doi: 10.1016/J.JPBA.2010.09.007.

Xie, C. *et al.* (2018) 'N-Acetylcysteine reduces ROS-mediated oxidative DNA damage and PI3K/Akt pathway activation induced by Helicobacter pylori infection',

*Oxidative Medicine and Cellular Longevity*. Hindawi Limited, 2018. doi: 10.1155/2018/1874985.

Yang, W. *et al.* (2013) 'Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells.', *Nucleic acids research*. Oxford University Press, 41(Database issue), pp. D955-61. doi: 10.1093/nar/gks1111.

Ye, C. *et al.* (2018) 'DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery', *Nature Communications*. Nature Publishing Group, 9(1), pp. 1–9. doi: 10.1038/s41467-018-06500-x.

Ye, X. *et al.* (2020) 'MSRB3 promotes the progression of clear cell renal cell carcinoma via regulating endoplasmic reticulum stress', *Pathology Research and Practice*. Elsevier GmbH, 216(2), p. 152780. doi: 10.1016/j.prp.2019.152780.

Young, D. W. *et al.* (2008) 'Integrating high-content screening and ligand-target prediction to identify mechanism of action', *Nature Chemical Biology*. Nature Publishing Group, 4(1), pp. 59–68. doi: 10.1038/nchembio.2007.53.

Zheng, J. *et al.* (2010) 'N-Acetylcysteine interacts with copper to generate hydrogen peroxide and selectively induce cancer cell death', *Cancer Letters*. NIH Public Access, 298(2), pp. 186–194. doi: 10.1016/j.canlet.2010.07.003.

Ziegler-Heitbrock, H. W. *et al.* (1993) 'Pyrrolidine dithiocarbamate inhibits NF-kappa B mobilization and TNF production in human monocytes.', *The Journal of Immunology*, 151(12).

# *A.1 | Appendices*

1. Supplementary Information.

**Table A.1. Primary screen compound hits**

| | Phenotypic hits | cell survival hits |
|---|---|---|
| 1 | (S)-(+)-Camptothecin | |
| 2 | 5-Fluorouracil | Adrucil(Fluorouracil) |
| 3 | Alcuronium chloride | |
| 4 | Aminopterin | |
| 5 | | Ammonium pyrrolidinedithiocarbamate |
| 6 | Amonafide | |
| 7 | Amsacrine hydrochloride | |
| 8 | Ancitabine hydrochloride | |
| 9 | Ascorbic acid | Ascorbic acid |
| 10 | | Aurora A Inhibitor I |
| 11 | Azapropazone | |
| 12 | AZD7762 | |
| 13 | Bacampicillin hydrochloride | Bacampicillin hydrochloride |
| 14 | Benzylpenicillin sodium | Benzydamine hydrochloride |
| 15 | Betahistine mesylate | |
| 16 | BI 2536 | |
| 17 | BIX-01294·3HCl | |
| 18 | Calcimycin | |
| 19 | Cantharidic Acid | |
| 20 | Cantharidin | |
| 21 | Carmofur | |
| 22 | CCT137690 | |

| | | |
|---|---|---|
| 23 | Ceforanide | |
| 24 | CID 11210285 hydrochloride | |
| 25 | | Clindamycin hydrochloride |
| 26 | Clofarabine | |
| 27 | | Clorgyline hydrochloride |
| 28 | Colchicine | |
| 29 | Cytosine-1-beta-D-arabinofuranoside hydrochloride | |
| 30 | Deptropine citrate | Deptropine citrate |
| 31 | Dipivefrin hydrochloride | Dipivefrin hydrochloride |
| 32 | | Disulfiram |
| 33 | | Elesclomol |
| 34 | Ellence | |
| 35 | | Etifenin |
| 36 | Etoposide | |
| 37 | Floxuridine | Floxuridine |
| 38 | Fluoxetine hydrochloride | Fluoxetine hydrochloride |
| 39 | Gemcitabine hydrochloride | |
| 40 | Ibuprofen | |
| 41 | JNJ-26481585 | |
| 42 | Letrozole | |
| 43 | Methotrexate | Methotrexate |
| 44 | Mitoxantrone | |
| 45 | Mizolastine | |
| 46 | MK-2206 | MK-2206 |
| 47 | MLN8237 | |
| 48 | Molsidomine | |
| 49 | | Nicergoline |
| 50 | NSC-3852 | NSC-3852 |
| 51 | Obatoclax Mesylate | |
| 52 | | Oxytetracycline dihydrate |

| | | |
|---|---|---|
| 53 | Paclitaxel | |
| 54 | Pemetrexed disodium (LY-231514) | Pemetrexed disodium (LY-231514) |
| 55 | Pirenperone | |
| 56 | PKC-412 | |
| 57 | PMEG hydrate | |
| 58 | Podophyllotoxin | |
| 59 | | Propylthiouracil |
| 60 | Raltitrexed (Tomudex) | Raltitrexed(Tomudex) |
| 61 | SB743921 hydrochloride | |
| 62 | SB939\| | |
| 63 | SNS-032(BMS-387032) | |
| 64 | SNS-314 Mesylate | |
| 65 | Suprafenacine | |
| 66 | | Terbutaline hemisulfate |
| 67 | Thiocolchicine | |
| 68 | Thiorphan | Thiorphan |
| 69 | Topotecan hydrochloride hydrate | |
| 70 | Trichostatin A | Trichostatin A |
| 71 | Vincristine sulfate | |
| 72 | Vinorelbine (Navelbine) | |
| 73 | VX-680 | |
| 74 | | YM155 |

1.

**Table A.2. Validation compounds.** (Red- not selective)

| | validation compounds | Phenotypic Validation | Cell survival Validation |
|---|---|---|---|
| 1 | (S)-(+)-Camptothecin | ✓ | ✓ |
| 2 | 5-Fluorouracil | ✓ | ✓ |
| 3 | alcuronium chloride | | |
| 4 | Aminopterin | | ✓ |

| | | | |
|---|---|---|---|
| 5 | Ammonium pyrrolidinedithiocarbamate | ✓ | ✓ |
| 6 | Amonafide | ✓ | ✓ |
| 7 | Amsacrine hydrochloride | ✓ | ✓ |
| 8 | Ancitabine hydrochloride | ✓ | ✓ |
| 9 | Ascorbic acid | | |
| 10 | Aurora A Inhibitor I | ✓ | ✓ |
| 11 | Azapropazone | | |
| 12 | AZD7762 | ✓ | ✓ |
| 13 | Benzydamine hydrochloride | ✓ | |
| 14 | Benzylpenicillin sodium | | |
| 15 | Betahistine mesylate | | |
| 16 | BI 2536 | ✓ | ✓ |
| 17 | BIX 01294 trihydrochloride hydrate | ✓ | |
| 18 | Calcimycin | ✓ | ✓ |
| 19 | Cantharidic Acid | ✓ | |
| 20 | Cantharidin | ✓ | ✓ |
| 21 | Carmofur | ✓ | ✓ |
| 22 | CCT137690 | ✓ | ✓ |
| 23 | Ceforanide | | |
| 24 | CID 11210285 hydrochloride | ✓ | ✓ |
| 25 | Clindamycin hydrochloride | | |
| 26 | Clofarabine | ✓ | ✓ |
| 27 | Clorgyline hydrochloride | | |
| 28 | Colchicine | ✓ | ✓ |
| 29 | Cytosine-1-beta-D-arabinofuranoside hydrochloride | ✓ | ✓ |
| 30 | deptropine citrate | | |
| 31 | Dipivefrin hydrochloride | | |
| 32 | Disulfiram | ✓ | ✓ |
| 33 | Elesclomol | ✓ | ✓ |
| 34 | Ellence | ✓ | ✓ |

| | | | |
|---|---|---|---|
| 35 | Etifenin | | |
| 36 | Etoposide | ✓ | ✓ |
| 37 | Floxuridine | ✓ | ✓ |
| 38 | Fluoxetine hydrochloride | ✓ | |
| 39 | Gemcitabine hydrochloride | ✓ | ✓ |
| 40 | Ibuprofen (racemic) | | |
| 41 | JNJ-26481585 | ✓ | ✓ |
| 42 | Letrozole | | |
| 43 | Methotrexate | ✓ | ✓ |
| 44 | Mitoxantrone | ✓ | ✓ |
| 45 | Mizolastine | | |
| 46 | MK-2206 | ✓ | ✓ |
| 47 | MLN8237 | ✓ | ✓ |
| 48 | Molsidomine | | |
| 49 | Nicergoline | | |
| 50 | NSC-3852 | ✓ | ✓ |
| 51 | Obatoclax Mesylate | ✓ | ✓ |
| 52 | Oxytetracycline dihydrate | | |
| 53 | Paclitaxel | ✓ | ✓ |
| 54 | Pemetrexed disodium (LY-231514) | ✓ | ✓ |
| 55 | pirenperone | | |
| 56 | PKC-412 | | ✓ |
| 57 | Podophyllotoxin | | ✓ |
| 58 | Propylthiouracil | | |
| 69 | Raltitrexed(Tomudex) | ✓ | ✓ |
| 60 | SB743921 hydrochloride | ✓ | ✓ |
| 61 | SNS-032(BMS-387032) | ✓ | ✓ |
| 62 | SNS-314 (mesylate) | ✓ | ✓ |
| 63 | Suprafenacine | ✓ | ✓ |
| 64 | Terbutaline hemisulfate | | |
| 65 | Thiocolchicine | ✓ | ✓ |

| | | | |
|---|---|---|---|
| 66 | Thiorphan | | |
| 67 | Topotecan hydrochloride hydrate | ✓ | ✓ |
| 68 | Trichostatin A | ✓ | ✓ |
| 69 | Vincristine sulfate | ✓ | ✓ |
| 70 | Vinorelbine(Navelbine) | ✓ | ✓ |
| 71 | VX-680 | ✓ | ✓ |
| 72 | YM155 | ✓ | ✓ |

**Table A.3. Elesclomol Induced Significant Differentially Expressed Genes in sensitive cell lines** (OAC-P4C and SK-GT-4).

| Gene ID | Log2 fold change | P-value | Adjusted P-value | Gene set annotation |
|---|---|---|---|---|
| HSPA6 | 9.8 | 1.17E-08 | 0.000128 | MAPK |
| HMOX1 | 5.12 | 1.44E-07 | 0.000332 | Cytokine & Chemokine Signaling, KEAP1NRF2 Pathway, Reactive Oxygen Response |
| HSPA1A | 4.39 | 1.97E-07 | 0.000359 | MAPK |
| IL11 | 3.62 | 4.79E-06 | 0.00187 | JAK-STAT |
| NR4A1 | 3 | 0.000171 | 0.0143 | MAPK, PI3K |
| GADD45G | 2.95 | 0.00033 | 0.0218 | Cell Cycle - Apoptosis, MAPK |
| DDIT3 | 2.91 | 4.09E-05 | 0.00574 | MAPK, Transcriptional Misregulation |
| PGF | 2.68 | 7.60E-06 | 0.00244 | PI3K, Ras |
| HSPB1 | 2.67 | 1.06E-05 | 0.00281 | MAPK |
| NR4A3 | 2.42 | 0.00016 | 0.0136 | Transcriptional Misregulation |
| IL8 | 2.36 | 0.000111 | 0.0107 | Transcriptional Misregulation |
| GADD45A | 2.34 | 8.47E-05 | 0.00899 | Cell Cycle - Apoptosis, MAPK |
| GADD45B | 2.32 | 0.000528 | 0.0309 | Cell Cycle - Apoptosis, MAPK |

| | | | | |
|---|---|---|---|---|
| FOS | 2.28 | 0.00217 | 0.0871 | MAPK |
| CDKN1A | 2.25 | 4.39E-05 | 0.00599 | Cell Cycle - Apoptosis, PI3K, Transcriptional Misregulation |
| INHBA | 2.15 | 0.000483 | 0.0293 | TGF-beta |
| MAP1LC3B | 2.12 | 5.25E-07 | 0.000723 | Autophagy |

**Table A.4. Rank order for core enriched genes in the MYC 1 Hallmark Geneset.**

| Number | Gene Symbol | Gene Name | Rank |
|---|---|---|---|
| 1 | UBA2 | ubiquitin like modifier activating enzyme 2 | 87 |
| 2 | PSMD8 | proteasome 26S subunit, non-ATPase 8 | 212 |
| 3 | SRM | spermidine synthase | 215 |
| 4 | LSM2 | LSM2 homolog, U6 small nuclear RNA and mRNA degradation associated | 267 |
| 5 | RACK1 | receptor for activated C kinase 1 | 287 |
| 6 | PSMC4 | proteasome 26S subunit, ATPase 4 | 375 |
| 7 | PWP1 | PWP1 homolog, endonuclein | 376 |
| 8 | UBE2E1 | ubiquitin conjugating enzyme E2 E1 | 479 |
| 9 | SF3A1 | splicing factor 3a subunit 1 | 507 |
| 10 | SF3B3 | splicing factor 3b subunit 3 | 575 |
| 11 | PABPC4 | poly(A) binding protein cytoplasmic 4 | 725 |
| 12 | PHB2 | prohibitin 2 | 751 |
| 13 | SLC25A3 | solute carrier family 25 member 3 | 860 |
| 14 | UBE2L3 | ubiquitin conjugating enzyme E2 L3 | 875 |
| 15 | USP1 | ubiquitin specific peptidase 1 | 887 |
| 16 | VDAC3 | voltage dependent anion channel 3 | 954 |
| 17 | SNRPA | small nuclear ribonucleoprotein polypeptide A | 1055 |
| 18 | XPOT | exportin for tRNA | 1133 |
| 19 | MRPL9 | mitochondrial ribosomal protein L9 | 1155 |

| | | | |
|---|---|---|---|
| 20 | SMARCC1 | SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 1 | 1161 |
| 21 | NAP1L1 | nucleosome assembly protein 1 like 1 | 1171 |
| 22 | RANBP1 | RAN binding protein 1 | 1353 |
| 23 | G3BP1 | G3BP stress granule assembly factor 1 | 1366 |
| 24 | PTGES3 | prostaglandin E synthase 3 | 1381 |
| 25 | HNRNPC | heterogeneous nuclear ribonucleoprotein C | 1384 |
| 26 | MRPS18B | mitochondrial ribosomal protein S18B | 1403 |
| 27 | CCT2 | chaperonin containing TCP1 subunit 2 | 1429 |
| 28 | SRPK1 | SRSF protein kinase 1 | 1490 |
| 29 | NOLC1 | nucleolar and coiled-body phosphoprotein 1 | 1512 |
| 30 | RAN | RAN, member RAS oncogene family | 1551 |
| 31 | HNRNPR | heterogeneous nuclear ribonucleoprotein R | 1565 |
| 32 | MCM5 | minichromosome maintenance complex component 5 | 1610 |
| 33 | CDK4 | cyclin dependent kinase 4 | 1636 |
| 34 | LSM7 | LSM7 homolog, U6 small nuclear RNA and mRNA degradation associated | 1652 |
| 35 | VBP1 | VHL binding protein 1 | 1661 |
| 36 | GNL3 | G protein nucleolar 3 | 1724 |
| 37 | CDC45 | cell division cycle 45 | 1769 |
| 38 | CSTF2 | cleavage stimulation factor subunit 2 | 1802 |
| 39 | CDC20 | cell division cycle 20 | 1807 |
| 40 | RSL1D1 | ribosomal L1 domain containing 1 | 1820 |
| 41 | KARS1 | lysyl-tRNA synthetase 1 | 1885 |
| 42 | U2AF1 | U2 small nuclear RNA auxiliary factor 1 | 1925 |
| 43 | PSMA4 | proteasome 20S subunit alpha 4 | 1931 |
| 44 | EIF3B | eukaryotic translation initiation factor 3 subunit B | 1978 |
| 45 | CAD | carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase | 2065 |

| | | | |
|---|---|---|---|
| 46 | YWHAE | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon | 2083 |
| 47 | FBL | fibrillarin | 2138 |
| 48 | TRA2B | transformer 2 beta homolog | 2140 |
| 49 | XPO1 | exportin 1 | 2151 |
| 50 | PPM1G | protein phosphatase, Mg2+/Mn2+ dependent 1G | 2423 |
| 51 | ETF1 | eukaryotic translation termination factor 1 | 2500 |
| 52 | TXNL4A | thioredoxin like 4A | 2578 |
| 53 | EIF4E | eukaryotic translation initiation factor 4E | 2646 |
| 54 | POLD2 | DNA polymerase delta 2, accessory subunit | 2664 |
| 55 | EIF3J | eukaryotic translation initiation factor 3 subunit J | 2667 |
| 56 | TUFM | Tu translation elongation factor, mitochondrial | 2700 |
| 57 | IMPDH2 | inosine monophosphate dehydrogenase 2 | 2734 |
| 58 | PSMB2 | proteasome 20S subunit beta 2 | 2808 |
| 59 | PSMD14 | proteasome 26S subunit, non-ATPase 14 | 2847 |
| 60 | EIF1AX | eukaryotic translation initiation factor 1A X-linked | 2855 |
| 61 | NPM1 | nucleophosmin 1 | 2877 |
| 62 | C1QBP | complement C1q binding protein | 2934 |
| 63 | BUB3 | BUB3 mitotic checkpoint protein | 3008 |
| 64 | POLE3 | DNA polymerase epsilon 3, accessory subunit | 3093 |
| 65 | RRP9 | ribosomal RNA processing 9, U3 small nucleolar RNA binding protein | 3205 |
| 66 | NHP2 | NHP2 ribonucleoprotein | 3209 |
| 67 | LDHA | lactate dehydrogenase A | 3235 |
| 68 | NOP16 | NOP16 nucleolar protein | 3266 |
| 69 | SNRPD3 | small nuclear ribonucleoprotein D3 polypeptide | 3334 |
| 70 | CCT4 | chaperonin containing TCP1 subunit 4 | 3422 |
| 71 | DEK | DEK proto-oncogene | 3531 |
| 72 | PA2G4 | proliferation-associated 2G4 | 3591 |
| 73 | CCT7 | chaperonin containing TCP1 subunit 7 | 3647 |

| | | | |
|---|---|---|---|
| 74 | NDUFAB1 | NADH:ubiquinone oxidoreductase subunit AB1 | 3668 |
| 75 | TCP1 | t-complex 1 | 3695 |
| 76 | PSMD7 | proteasome 26S subunit, non-ATPase 7 | 3716 |
| 77 | GLO1 | glyoxalase I | 3749 |
| 78 | CTPS1 | CTP synthase 1 | 3838 |
| 79 | PSMD1 | proteasome 26S subunit, non-ATPase 1 | 3883 |
| 80 | SSB | small RNA binding exonuclease protection factor La | 3961 |
| 81 | PCBP1 | poly(rC) binding protein 1 | 4034 |
| 82 | RPL14 | ribosomal protein L14 | 4035 |
| 83 | HPRT1 | hypoxanthine phosphoribosyltransferase 1 | 4109 |
| 84 | EPRS1 | glutamyl-prolyl-tRNA synthetase 1 | 4111 |
| 85 | TARDBP | TAR DNA binding protein | 4122 |
| 86 | EIF3D | eukaryotic translation initiation factor 3 subunit D | 4147 |
| 87 | MRPL23 | mitochondrial ribosomal protein L23 | 4192 |
| 88 | ODC1 | ornithine decarboxylase 1 | 4203 |
| 89 | RFC4 | replication factor C subunit 4 | 4218 |
| 90 | AIMP2 | aminoacyl tRNA synthetase complex interacting multifunctional protein 2 | 4249 |
| 91 | TYMS | thymidylate synthetase | 4371 |
| 92 | CYC1 | cytochrome c1 | 4420 |
| 93 | PRPS2 | phosphoribosyl pyrophosphate synthetase 2 | 4436 |
| 94 | APEX1 | apurinic/apyrimidinic endodeoxyribonuclease 1 | 4457 |

**Table A.5. Rank order for core enriched genes in the MYC 2 Hallmark Geneset.**

| Number | Gene Symbol | Gene Name | Rank |
|---|---|---|---|
| 1 | SRM | spermidine synthase | 215 |
| 2 | MYBBP1A | MYB binding protein 1a | 495 |
| 3 | NOP2 | NOP2 nucleolar protein | 503 |

| | | | |
|---|---|---|---|
| 4 | SLC19A1 | solute carrier family 19 member 1 | 837 |
| 5 | PLK1 | polo like kinase 1 | 977 |
| 6 | MRTO4 | MRT4 homolog, ribosome maturation factor | 1085 |
| 7 | NIP7 | nucleolar pre-rRNA processing protein NIP7 | 1333 |
| 8 | NOLC1 | nucleolar and coiled-body phosphoprotein 1 | 1512 |
| 9 | MCM5 | minichromosome maintenance complex component 5 | 1610 |
| 10 | CDK4 | cyclin dependent kinase 4 | 1636 |
| 11 | UTP20 | UTP20 small subunit processome component | 1656 |
| 12 | GNL3 | G protein nucleolar 3 | 1724 |
| 13 | NOC4L | nucleolar complex associated 4 homolog | 1908 |
| 14 | PPRC1 | PPARG related coactivator 1 | 2002 |
| 15 | LAS1L | LAS1 like ribosome biogenesis factor | 2132 |
| 16 | TBRG4 | transforming growth factor beta regulator 4 | 2354 |
| 17 | PES1 | pescadillo ribosomal biogenesis factor 1 | 2676 |
| 18 | RRP12 | ribosomal RNA processing 12 homolog | 2716 |
| 19 | NPM1 | nucleophosmin 1 | 2877 |
| 20 | FARSA | phenylalanyl-tRNA synthetase subunit alpha | 2951 |
| 21 | RRP9 | ribosomal RNA processing 9, U3 small nucleolar RNA binding protein | 3205 |
| 22 | WDR74 | WD repeat domain 74 | 3225 |
| 23 | NOP16 | NOP16 nucleolar protein | 3266 |
| 24 | TCOF1 | treacle ribosome biogenesis factor 1 | 3544 |
| 25 | PA2G4 | proliferation-associated 2G4 | 3591 |
| 26 | IMP4 | IMP U3 small nucleolar ribonucleoprotein 4 | 3623 |
| 27 | UNG | uracil DNA glycosylase | 4244 |
| 28 | AIMP2 | aminoacyl tRNA synthetase complex interacting multifunctional protein 2 | 4249 |
| 29 | BYSL | bystin like | 4293 |
| 30 | MAP3K6 | mitogen-activated protein kinase kinase kinase 6 | 4596 |
| 31 | PUS1 | pseudouridine synthase 1 | 4623 |

**Table A.6. Rank order for core enriched genes in the G2M checkpoint Hallmark Geneset.**

| Number | Gene Symbol | Gene Name | Rank |
|---|---|---|---|
| 1 | ESPL1 | extra spindle pole bodies like 1, separase | 41 |
| 2 | TPX2 | TPX2 microtubule nucleation factor | 96 |
| 3 | H2AZ2 | H2A.Z variant histone 2 | 230 |
| 4 | CENPA | centromere protein A | 263 |
| 5 | ATF5 | activating transcription factor 5 | 358 |
| 6 | KATNA1 | katanin catalytic subunit A1 | 456 |
| 7 | HUS1 | HUS1 checkpoint clamp component | 475 |
| 8 | RACGAP1 | Rac GTPase activating protein 1 | 642 |
| 9 | SS18 | SS18 subunit of BAF chromatin remodeling complex | 694 |
| 10 | PRMT5 | protein arginine methyltransferase 5 | 712 |
| 11 | EZH2 | enhancer of zeste 2 polycomb repressive complex 2 subunit | 810 |
| 12 | UBE2C | ubiquitin conjugating enzyme E2 C | 836 |
| 13 | PBK | PDZ binding kinase | 845 |
| 14 | CDKN3 | cyclin dependent kinase inhibitor 3 | 894 |
| 15 | NUP98 | nucleoporin 98 and 96 precursor | 931 |
| 16 | PLK1 | polo like kinase 1 | 977 |
| 17 | PTTG3P | pituitary tumor-transforming 3, pseudogene | 1067 |
| 18 | KNL1 | kinetochore scaffold 1 | 1135 |
| 19 | UCK2 | uridine-cytidine kinase 2 | 1139 |
| 20 | SMARCC1 | SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 1 | 1161 |
| 21 | BUB1 | BUB1 mitotic checkpoint serine/threonine kinase | 1170 |
| 22 | SAP30 | Sin3A associated protein 30 | 1174 |
| 23 | KIF23 | kinesin family member 23 | 1201 |

| | | | |
|---|---|---|---|
| 24 | RPA2 | replication protein A2 | 1217 |
| 25 | NUSAP1 | nucleolar and spindle associated protein 1 | 1308 |
| 26 | G3BP1 | G3BP stress granule assembly factor 1 | 1366 |
| 27 | FBXO5 | F-box protein 5 | 1382 |
| 28 | STMN1 | stathmin 1 | 1484 |
| 29 | KIF4A | kinesin family member 4A | 1502 |
| 30 | NOLC1 | nucleolar and coiled-body phosphoprotein 1 | 1512 |
| 31 | AURKB | aurora kinase B | 1528 |
| 32 | CDKN2C | cyclin dependent kinase inhibitor 2C | 1545 |
| 33 | KIF11 | kinesin family member 11 | 1555 |
| 34 | MCM5 | minichromosome maintenance complex component 5 | 1610 |
| 35 | TMPO | thymopoietin | 1628 |
| 36 | CDK4 | cyclin dependent kinase 4 | 1636 |
| 37 | E2F1 | E2F transcription factor 1 | 1644 |
| 38 | CENPF | centromere protein F | 1720 |
| 39 | BIRC5 | baculoviral IAP repeat containing 5 | 1733 |
| 40 | CDC45 | cell division cycle 45 | 1769 |
| 41 | CDC20 | cell division cycle 20 | 1807 |
| 42 | SMC4 | structural maintenance of chromosomes 4 | 1860 |
| 43 | CCNF | cyclin F | 1906 |
| 44 | E2F2 | E2F transcription factor 2 | 2023 |
| 45 | CHMP1A | charged multivesicular body protein 1A | 2046 |
| 46 | POLE | DNA polymerase epsilon, catalytic subunit | 2137 |
| 47 | CDC25B | cell division cycle 25B | 2139 |
| 48 | TRA2B | transformer 2 beta homolog | 2140 |
| 49 | XPO1 | exportin 1 | 2151 |
| 50 | PTTG1 | PTTG1 regulator of sister chromatid separation, securin | 2182 |
| 51 | SRSF10 | serine and arginine rich splicing factor 10 | 2270 |
| 52 | PRIM2 | DNA primase subunit 2 | 2277 |
| 53 | KIF20B | kinesin family member 20B | 2342 |

| | | | |
|---|---|---|---|
| 54 | PRPF4B | pre-mRNA processing factor 4B | 2393 |
| 55 | ORC6 | origin recognition complex subunit 6 | 2508 |
| 56 | ODF2 | outer dense fiber of sperm tails 2 | 2558 |
| 57 | KIF15 | kinesin family member 15 | 2563 |
| 58 | E2F3 | E2F transcription factor 3 | 2781 |
| 59 | MKI67 | marker of proliferation Ki-67 | 2801 |
| 60 | SMAD3 | SMAD family member 3 | 2962 |
| 61 | KIF2C | kinesin family member 2C | 2974 |
| 62 | NEK2 | NIMA related kinase 2 | 2984 |
| 63 | BUB3 | BUB3 mitotic checkpoint protein | 3008 |
| 64 | DDX39A | DExD-box helicase 39A | 3029 |
| 65 | MYBL2 | MYB proto-oncogene like 2 | 3052 |
| 66 | TRAIP | TRAF interacting protein | 3068 |
| 67 | WRN | WRN RecQ like helicase | 3116 |
| 68 | HMGA1 | high mobility group AT-hook 1 | 3120 |
| 69 | KIF5B | kinesin family member 5B | 3215 |
| 70 | PURA | purine rich element binding protein A | 3238 |
| 71 | CENPE | centromere protein E | 3241 |
| 72 | INCENP | inner centromere protein | 3399 |
| 73 | CDC7 | cell division cycle 7 | 3405 |
| 74 | CHEK1 | checkpoint kinase 1 | 3566 |
| 75 | TENT4A | terminal nucleotidyltransferase 4A | 3616 |
| 76 | SUV39H1 | suppressor of variegation 3-9 homolog 1 | 3674 |
| 77 | MT2A | metallothionein 2A | 3880 |
| 78 | NOTCH2 | notch receptor 2 | 3926 |
| 79 | CHAF1A | chromatin assembly factor 1 subunit A | 3956 |
| 80 | ILF3 | interleukin enhancer binding factor 3 | 4176 |
| 81 | ODC1 | ornithine decarboxylase 1 | 4203 |
| 82 | PAFAH1B1 | platelet activating factor acetylhydrolase 1b regulatory subunit 1 | 4242 |
| 83 | E2F4 | E2F transcription factor 4 | 4304 |

| | | | |
|---|---|---|---|
| 84 | NUP50 | nucleoporin 50 | 4311 |

**Table A.7. Rank order for core enriched genes in the E2F Targets Hallmark Geneset.**

| Number | Gene Symbol | Gene Name | Rank |
|---|---|---|---|
| 1 | TP53 | tumor protein p53 | 3 |
| 2 | ESPL1 | extra spindle pole bodies like 1, separase | 41 |
| 3 | PPP1R8 | protein phosphatase 1 regulatory subunit 8 | 74 |
| 4 | HUS1 | HUS1 checkpoint clamp component | 475 |
| 5 | CIT | citron rho-interacting serine/threonine kinase | 629 |
| 6 | RACGAP1 | Rac GTPase activating protein 1 | 642 |
| 7 | TUBB | tubulin beta class I | 762 |
| 8 | EZH2 | enhancer of zeste 2 polycomb repressive complex 2 subunit | 810 |
| 9 | USP1 | ubiquitin specific peptidase 1 | 887 |
| 10 | CDKN3 | cyclin dependent kinase inhibitor 3 | 894 |
| 11 | PLK1 | polo like kinase 1 | 977 |
| 12 | DLGAP5 | DLG associated protein 5 | 1162 |
| 13 | NAP1L1 | nucleosome assembly protein 1 like 1 | 1171 |
| 14 | RPA2 | replication protein A2 | 1217 |
| 15 | SMC3 | structural maintenance of chromosomes 3 | 1223 |
| 16 | RANBP1 | RAN binding protein 1 | 1353 |
| 17 | TIMELESS | timeless circadian regulator | 1461 |
| 18 | STMN1 | stathmin 1 | 1484 |
| 19 | KIF4A | kinesin family member 4A | 1502 |
| 20 | NOLC1 | nucleolar and coiled-body phosphoprotein 1 | 1512 |
| 21 | AURKB | aurora kinase B | 1528 |
| 22 | CCP110 | centriolar coiled-coil protein 110 | 1542 |
| 23 | CDKN2C | cyclin dependent kinase inhibitor 2C | 1545 |
| 24 | RAN | RAN, member RAS oncogene family | 1551 |

| | | | |
|---|---|---|---|
| 25 | MCM5 | minichromosome maintenance complex component 5 | 1610 |
| 26 | TMPO | thymopoietin | 1628 |
| 27 | CDK4 | cyclin dependent kinase 4 | 1636 |
| 28 | BIRC5 | baculoviral IAP repeat containing 5 | 1733 |
| 29 | CDC20 | cell division cycle 20 | 1807 |
| 30 | SMC4 | structural maintenance of chromosomes 4 | 1860 |
| 31 | SPC25 | SPC25 component of NDC80 kinetochore complex | 1966 |
| 32 | CCNE1 | cyclin E1 | 2101 |
| 33 | PSIP1 | PC4 and SFRS1 interacting protein 1 | 2136 |
| 34 | POLE | DNA polymerase epsilon, catalytic subunit | 2137 |
| 35 | CDC25B | cell division cycle 25B | 2139 |
| 36 | TRA2B | transformer 2 beta homolog | 2140 |
| 37 | XPO1 | exportin 1 | 2151 |
| 38 | PTTG1 | PTTG1 regulator of sister chromatid separation, securin | 2182 |
| 39 | CENPM | centromere protein M | 2205 |
| 40 | PRIM2 | DNA primase subunit 2 | 2277 |
| 41 | TBRG4 | transforming growth factor beta regulator 4 | 2354 |
| 42 | NBN | nibrin | 2355 |
| 43 | SSRP1 | structure specific recognition protein 1 | 2411 |
| 44 | DNMT1 | DNA methyltransferase 1 | 2470 |
| 45 | NUP107 | nucleoporin 107 | 2483 |
| 46 | DCLRE1B | DNA cross-link repair 1B | 2493 |
| 47 | PRPS1 | phosphoribosyl pyrophosphate synthetase 1 | 2501 |
| 48 | ORC6 | origin recognition complex subunit 6 | 2508 |
| 49 | TIPIN | TIMELESS interacting protein | 2585 |
| 50 | RAD50 | RAD50 double strand break repair protein | 2645 |
| 51 | POLD2 | DNA polymerase delta 2, accessory subunit | 2664 |
| 52 | CDKN2A | cyclin dependent kinase inhibitor 2A | 2725 |

| | | | |
|---|---|---|---|
| 53 | RBBP7 | RB binding protein 7, chromatin remodeling factor | 2741 |
| 54 | MKI67 | marker of proliferation Ki-67 | 2801 |
| 55 | TCF19 | transcription factor 19 | 2832 |
| 56 | ASF1B | anti-silencing function 1B histone chaperone | 2930 |
| 57 | IPO7 | importin 7 | 2949 |
| 58 | KIF2C | kinesin family member 2C | 2974 |
| 59 | DDX39A | DExD-box helicase 39A | 3029 |
| 60 | RFC2 | replication factor C subunit 2 | 3045 |
| 61 | RNASEH2A | ribonuclease H2 subunit A | 3048 |
| 62 | MYBL2 | MYB proto-oncogene like 2 | 3052 |
| 63 | ANP32E | acidic nuclear phosphoprotein 32 family member E | 3063 |
| 64 | HMGA1 | high mobility group AT-hook 1 | 3120 |
| 65 | PRKDC | protein kinase, DNA-activated, catalytic subunit | 3170 |
| 66 | CENPE | centromere protein E | 3241 |
| 67 | AK2 | adenylate kinase 2 | 3254 |
| 68 | CDCA8 | cell division cycle associated 8 | 3427 |
| 69 | MLH1 | mutL homolog 1 | 3471 |
| 70 | DEK | DEK proto-oncogene | 3531 |
| 71 | CHEK1 | checkpoint kinase 1 | 3566 |
| 72 | PA2G4 | proliferation-associated 2G4 | 3591 |
| 73 | LYAR | Ly1 antibody reactive | 3655 |
| 74 | SUV39H1 | suppressor of variegation 3-9 homolog 1 | 3674 |
| 75 | CTPS1 | CTP synthase 1 | 3838 |
| 76 | GINS3 | GINS complex subunit 3 | 3884 |
| 77 | NUP153 | nucleoporin 153 | 3910 |
| 78 | ILF3 | interleukin enhancer binding factor 3 | 4176 |
| 79 | UNG | uracil DNA glycosylase | 4244 |
| 80 | NUDT21 | nudix hydrolase 21 | 4271 |

| | | | |
|---|---|---|---|
| 81 | UBR7 | ubiquitin protein ligase E3 component n-recognin 7 | 4358 |
| 82 | NCAPD2 | non-SMC condensin I complex subunit D2 | 4369 |
| 83 | GINS4 | GINS complex subunit 4 | 4456 |
| 84 | SPC24 | SPC24 component of NDC80 kinetochore complex | 4629 |
| 85 | UBE2T | ubiquitin conjugating enzyme E2 T | 4785 |
| 86 | DUT | deoxyuridine triphosphatase | 4902 |
| 87 | TRIP13 | thyroid hormone receptor interactor 13 | 4935 |
| 88 | POLA2 | DNA polymerase alpha 2, accessory subunit | 4941 |
| 89 | DONSON | DNA replication fork stabilization factor DONSON | 4950 |
| 90 | RPA1 | replication protein A1 | 5046 |
| 91 | RRM2 | ribonucleotide reductase regulatory subunit M2 | 5086 |
| 92 | MELK | maternal embryonic leucine zipper kinase | 5244 |
| 93 | MRE11 | MRE11 homolog, double strand break repair nuclease | 5294 |
| 94 | DEPDC1 | DEP domain containing 1 | 5356 |
| 95 | RAD1 | RAD1 checkpoint DNA exonuclease | 5359 |
| 96 | SMC6 | structural maintenance of chromosomes 6 | 5425 |
| 97 | H2AX | H2A.X variant histone | 5449 |
| 98 | E2F8 | E2F transcription factor 8 | 5506 |

**Table A.8. Rank order for core enriched genes in the Proteasome KEGG Geneset.**

| Number | Gene Symbol | Gene Name | Rank |
|---|---|---|---|
| 1 | PSMD8 | proteasome 26S subunit, non-ATPase 8 | 212 |
| 2 | PSMB5 | proteasome 20S subunit beta 5 | 302 |
| 3 | PSMC4 | proteasome 26S subunit, ATPase 4 | 375 |

| | | | |
|---|---|---|---|
| 4 | PSMD13 | proteasome 26S subunit, non-ATPase 13 | 625 |
| 5 | PSMB1 | proteasome 20S subunit beta 1 | 681 |
| 6 | PSMB6 | proteasome 20S subunit beta 6 | 740 |
| 7 | PSMD6 | proteasome 26S subunit, non-ATPase 6 | 864 |
| 8 | PSME1 | proteasome activator subunit 1 | 1425 |
| 9 | PSMA3 | proteasome 20S subunit alpha 3 | 1692 |
| 10 | PSMA4 | proteasome 20S subunit alpha 4 | 1931 |
| 11 | PSMB4 | proteasome 20S subunit beta 4 | 1935 |
| 12 | PSMB7 | proteasome 20S subunit beta 7 | 2085 |
| 13 | PSMA5 | proteasome 20S subunit alpha 5 | 2370 |
| 14 | PSMB9 | proteasome 20S subunit beta 9 | 2695 |
| 15 | PSMB2 | proteasome 20S subunit beta 2 | 2808 |
| 16 | PSMD14 | proteasome 26S subunit, non-ATPase 14 | 2847 |
| 17 | PSMC5 | proteasome 26S subunit, ATPase 5 | 3258 |
| 18 | PSMD7 | proteasome 26S subunit, non-ATPase 7 | 3716 |
| 19 | PSMD11 | proteasome 26S subunit, non-ATPase 11 | 3833 |
| 20 | PSMD1 | proteasome 26S subunit, non-ATPase 1 | 3883 |
| 21 | PSMD4 | proteasome 26S subunit, non-ATPase 4 | 4215 |

**Table A. 9. Connectivity Map compounds with similar gene expression signatures.**

| Score | Compound | Description |
|---|---|---|
| 99.98 | Elesclomol | apoptosis stimulant, HSP agonist, HSP inducer, oxidative stress inducer, topoisomerase |

| | | |
|---|---|---|
| | | inhibitor |
| 88.65 | Pyrrolidinedithiocarbamate | NFkB pathway inhibitor |
| 98.63 | Quinoclamine | algicide |
| 98.52 | Phenethyl-isothiocyanate | cancer cell growth inhibitor, unidentified pharmacological activity |
| 98.24 | CA-074-Me | cathepsin inhibitor, antiamyloidogenic agent |
| 98.24 | RO-25-6981 | ionotropic glutamate receptor antagonist, monamine transporter modulator |
| 98.03 | Sphingosine | ceramidase inhibitor |
| 97.71 | BMS-641988 | androgen receptor antagonist |
| 97.67 | Quinine | cytochrome P450 inhibitor, hemozoin biocrystallization inhibitor, P glycoprotein inhibitor |
| 97.62 | GBR-12783 | dopamine uptake inhibitor |

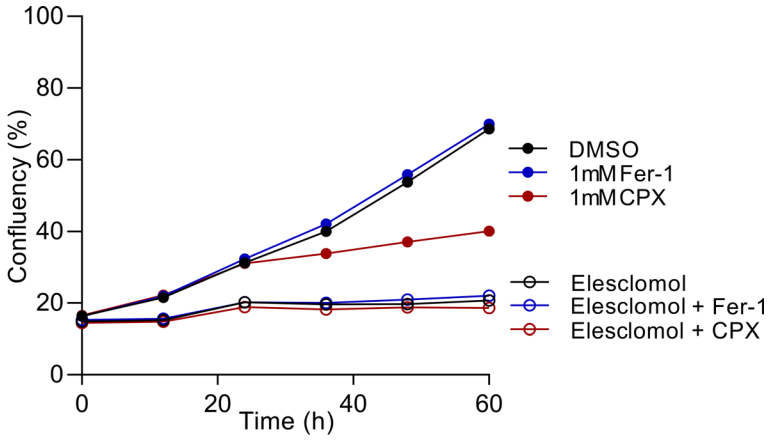



**Figure A.1. Role of iron.** OAC-P4C cell confluency after treatment with elesclomol (10 μM) in the presence or absence of the iron chelator ciclopirox olamine (CPX),

and ferroptosis inhibitor Ferrostatin-1 (Fer-1). Experiment carried out by Kenji Fujihara, Peter MacCallum Cancer Centre.

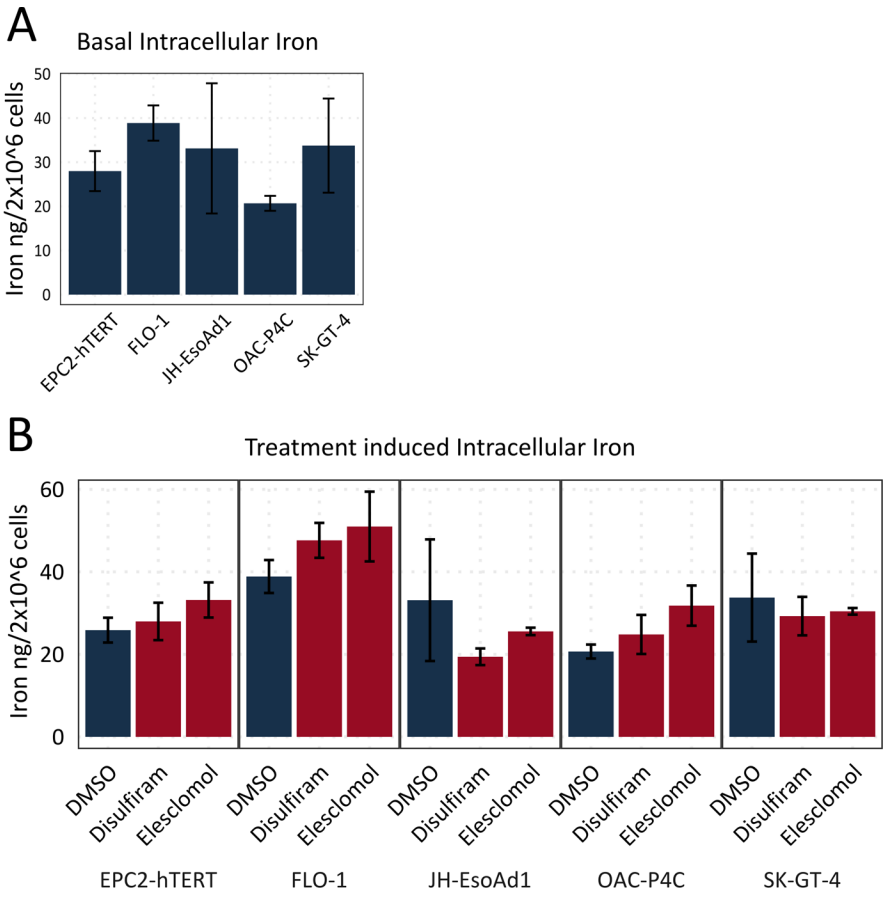



**Figure A.2. Iron ICP-MS.** A) Basal and B) Treatment induced intracellular iron levels determined by inductively coupled plasma mass spectrometry. Error bars indicate SE. n= 3.

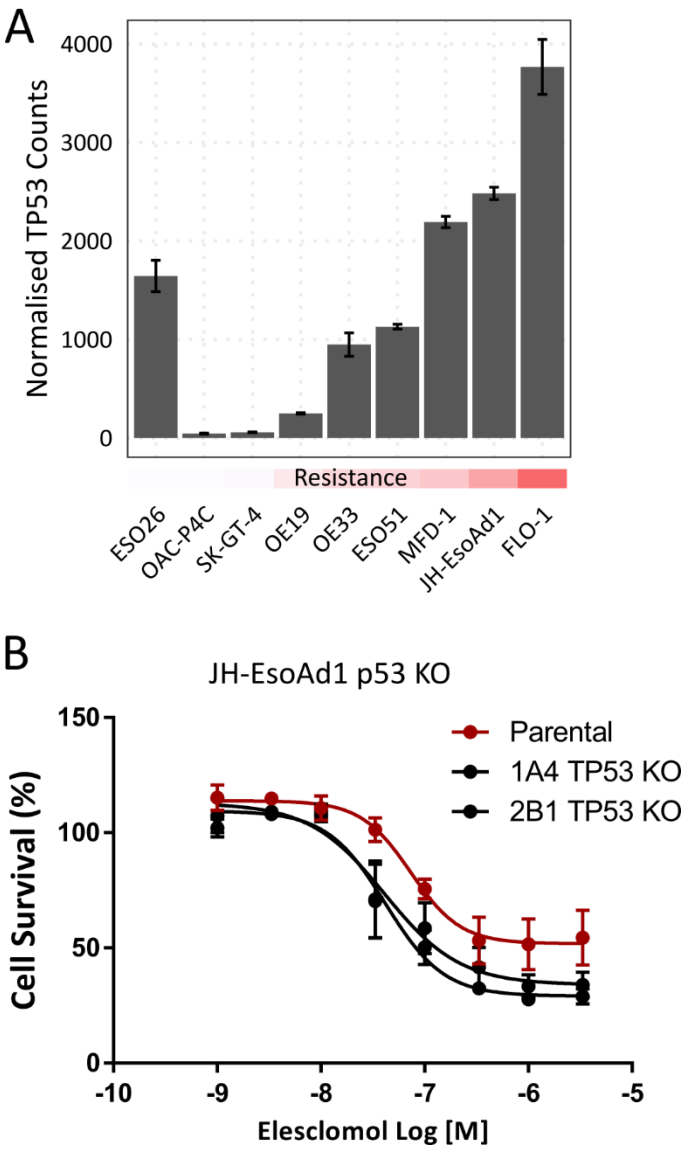



**Figure A.3. TP53 expression and knockout.** A) Normalised TP53 expression counts acorss OAC panel ranked by resistance to disulfiram and elesclomol. B) Elesclomol dose response in TP53 knockout JH-EsoAd1. JHesoAD1 parental cells (mut-p53 G266E) and two p53 KO clones (1A4 and 2B1) 48hrs after dosing, cell viability was accesses by AlamarBlue (~2hr incubation) N=3-4. Experiment carried out by Kenji Fujihara, Peter MacCallum Cancer Centre.

## 2. Published Papers

This appendix includes the first author publications accepted for publication during the course of my PhD studies. The first paper is a methods paper forming the basis of the material for chapters 3 and 4, whereas the second paper is a review on the topic of high content and pathway profiling to inform drug mechanism of action. I am about to submit a third paper for publication relating to the data forming chapters 5 and 6. Not included here, there are several other papers published during my PhD where I am not the first author.

---

1. Hughes, R. E. *et al.* (2020) 'High-Content Phenotypic Profiling in Esophageal Adenocarcinoma Identifies Selectively Active Pharmacological Classes of Drugs for Repurposing and Chemical Starting Points for Novel Drug Discovery', *SLAS Discovery*. SAGE Publications Inc., 25(7), pp. 770–782. doi: 10.1177/2472555220917115.

2. Hughes, R. E. *et al*. (2021). 'High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need', Cell Chemical Biology. CellPress, doi: 10.1016/j.chembiol.2021.02.015

Check for updates

Original Research

SLAS Discovery
2020, Vol. 25(7) 770–782
© 2020 The Author(s)

DOI: 10.1177/2472555220917115
journals.sagepub.com/home/jbx
SAGE

# High-Content Phenotypic Profiling in Esophageal Adenocarcinoma Identifies Selectively Active Pharmacological Classes of Drugs for Repurposing and Chemical Starting Points for Novel Drug Discovery

**Rebecca E. Hughes[1], Richard J. R. Elliott[1], Alison F. Munro[1], Ashraff Makda[1], J. Robert O'Neill[2], Ted Hupp[1], and Neil O. Carragher[1]**

## Abstract

Esophageal adenocarcinoma (EAC) is a highly heterogeneous disease, dominated by large-scale genomic rearrangements and copy number alterations. Such characteristics have hampered conventional target-directed drug discovery and personalized medicine strategies, contributing to poor outcomes for patients. We describe the application of a high-content Cell Painting assay to profile the phenotypic response of 19,555 compounds across a panel of six EAC cell lines and two tissue-matched control lines. We built an automated high-content image analysis pipeline to identify compounds that selectively modified the phenotype of EAC cell lines. We further trained a machine-learning model to predict the mechanism of action of EAC selective compounds using phenotypic fingerprints from a library of reference compounds. We identified a number of phenotypic clusters enriched with similar pharmacological classes, including methotrexate and three other antimetabolites that are highly selective for EAC cell lines. We further identify a small number of hits from our diverse chemical library that show potent and selective activity for EAC cell lines and that do not cluster with the reference library of compounds, indicating they may be selectively targeting novel esophageal cancer biology. Overall, our results demonstrate that our EAC phenotypic screening platform can identify existing pharmacologic classes and novel compounds with selective activity for EAC cell phenotypes.

## Keywords

esophageal adenocarcinoma, phenotypic, high content, mechanism of action, machine learning

## Introduction

Combined, the two major histological subtypes of esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma represent the sixth leading cause of cancer deaths worldwide, with fewer than one in five patients surviving 5 y from diagnosis.[1] A shift in epidemiology over the past 50 y has meant the incidence of EAC now vastly exceeds that of esophageal squamous cell carcinoma in Western countries,[2] accounting for more than 80% of esophageal cancers in the United States.[3] Defining the optimal neoadjuvant treatment regime is an area of active investigation,[4] as current treatments carry a significant risk of systemic toxicity, histologic response rates remain poor,[5] and only a limited subgroup of patients experience any survival benefit over surgery alone.[6,7]

EAC is a highly heterogeneous disease, dominated by large-scale genomic rearrangements and copy number alterations.[8] This has made clinically meaningful subgroups and well-validated therapeutic targets difficult to define. Clinical trials with new molecular targeted agents have predominantly been

[1]MRC Institute of Genetics & Molecular Medicine, The University of Edinburgh, Western General Hospital, Edinburgh, UK
[2]Cambridge Oesophagogastric Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, Cambridgeshire, UK

Received Dec 18, 2019, and in revised form Feb 26, 2020. Accepted for publication Mar 16, 2020.

Supplemental material is available online with this article.

**Corresponding Author:**
Neil O. Carragher, Cancer Research UK Edinburgh Centre, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh, EH4 2XR, UK.
Email: n.carragher@ed.ac.uk

directed toward epidermal growth factor receptor and human epidermal growth factor receptor 2 (HER2) receptors[9–12] but thus far have proven unsuccessful. A potential explanation is the almost ubiquitous coamplification of alternative receptor tyrosine kinases and downstream pathways leading to redundancy and drug resistance.[8,13,14] An alternative to target-based drug discovery, and increasing in popularity with technological advances, is phenotypic drug discovery, defined as the identification of novel compounds or other types of therapeutic agents with no prior knowledge of the drug target. Recent advances in phenotypic screening include automated high-content profiling.[15,16] This approach involves quantifying a large number of morphological features from cell or small-model organism assays in an unbiased way to identify changes and phenotypes of interest. One benefit to this method is that a target does not need to be predefined, but the mechanism of action (MoA) of hit compounds can be inferred by reference to known compound sets using multivariate statistics and machine-learning approaches. Thus, this may prove a beneficial strategy for complex, heterogeneous diseases in which target biology is poorly understood and modern, target directed drug discovery strategies have made little impact on patient care, as exemplified by EAC.

Taking an unbiased, profiling approach to phenotypic screening, we chose to apply the Cell Painting assay to capture large amounts of information on cellular and subcellular morphology to quantify the cellular state across a panel of genetically distinct EAC cell lines. Cell Painting is an assay developed to capture as many biologically relevant morphological features in a single assay so as not to constrain discovery to what we think we already know.[17,18] Therefore, upon chemical perturbation, we can detect changes in a subset of profiled features, allowing a phenotypic fingerprint to be assigned to a particular perturbation or compound.[15,19–21] These fingerprints can then be used to identify specific phenotypic changes of interest, identify compounds that cause strong alterations in cell morphology suggesting changes in cellular state or stress, or predict MoA by similarity comparison to reference libraries of well-annotated compound mechanisms.[17,21] However, this type of analysis is typically performed in a single "model" cell line, chosen for its suitability for image analysis. As a proof of principle that high-content phenotypic profiling could be applied to a panel of morphologically distinct EAC and tissue-matched control cell lines, we iteratively optimized cell culture conditions, cell-plating densities, and the Cell Painting assay staining protocol across our cell panel. Assay performance in terms of distinguishing distinct compound MoA for each cell type was evaluated by testing a small reference set of well-annotated compounds representing eight distinct mechanistic classes and performing principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) to visualize clustering of distinct mechanistic classes. We further developed a machine-learning model capable of predicting MoA across the panel of heterogeneous EAC cell lines. Following assay validation, we subsequently screened a library of 19,555 small molecules comprising target annotated probe compounds, approved drug libraries, and two diverse chemical sets with unknown MoA. PCA clustering of compound fingerprints distinguished a number of phenotypic clusters composed of similar pharmacologic classes active in the EAC cell lines. We also applied a Mahalanobis distance threshold and differential Z-score on our phenotypic data to identify compounds from our screen that were selectively active in EAC versus tissue-matched control cells. For prioritized hits, we have selected a subset and validated EAC selectivity with follow-up dose-response testing and performed transcriptomic pathway analysis pre- and posttreatment on sensitive and insensitive cell lines to further elucidate MoA. We further applied PCA and machine-learning analysis to phenotypic fingerprints from our diverse chemical set to identify compounds that exhibit selective activity on EAC cell phenotypes by a mechanism distinct from our reference set, indicating they may exhibit novel MoA.

Herein we describe the development and validation of a high-content phenotypic profiling assay and associated image informatics and machine-learning toolbox to classify the MoA of phenotypic screening hits across a panel of EAC and tissue-matched control cell lines. This approach has enabled the identification of chemical and target classes, including histone deacetylase (HDAC) inhibitors, which consistently cause the same cellular response across the panel of EAC lines, demonstrating efficacy against the heterogeneity of the disease. In addition, we identify pharmacologic classes such as the antimetabolites and new chemical entities with high selectivity for some EAC cell lines relative to tissue-matched controls. We propose that applying high-content multiparametric phenotypic profiling to a panel of genetically annotated EAC cell lines may stimulate new drug discovery and drug development programs for EAC through the identification of drug-repurposing opportunities and novel chemical starting points with selective activity for specific EAC genotypes.

## Materials and Methods

### Cell Culture

EPC2-hTERT cells were a kind donation from Anil Rustgis's Lab, University of Pennsylvania.[22]

### Cell Line Authentication

Cell line identification (not carried out for the EPC2-hTERT line, as there is no reference sequence) was confirmed by short tandem repeat genotyping (Cell Line Authentication, Public Health England).

**Table 1.** Cell Painting Reagents, Concentrations, Excitation/Emission Wavelengths of the Filters Used for Imaging, and Suppliers.

| Stain | Structure | Wavelength, ex/em (nm) | Channel | Concentration | Original Concentration[a] | Catalog No.; Supplier |
|---|---|---|---|---|---|---|
| Hoescht 33342 | Nuclei | 387/447 | DAPI | 4 µg/mL | 5 µg/mL | H1399; Molecular Probes, Eugene, OR |
| SYTO 14 | Nucleoli | 531/593 | CY3 | 3 µM | 3 µM | S7576; Invitrogen, Carlsbad, CA |
| Phalloidin 594 | F-actin | 562/624 | TxRED | 0.14X | 5 µL/mL | ab176757; Abcam, Cambridge, UK |
| Wheat germ agglutinin Alexa Fluor 594 | Golgi and plasma membrane | 562/624 | TxRED | 1 µg/mL | 1.5 µg/mL | W11262; Invitrogen |
| Concanavalin A Alexa Fluor 488 | Endoplasmic reticulum | 462/520 | FITC | 20 µg/mL | 100 µg/mL | C11252; Invitrogen |
| MitoTracker DeepRed | Mitochondria | 628/692 | CY5 | 600 nM | 500 nM | M22426; Invitrogen |

ex, excitation; em, emission.
[a]We also provide a comparison of reagent concentrations used in this study with the original Cell Painting protocol.[18]

The cell lines were confirmed to be mycoplasma negative using the VenorGeM mycoplasma detection PCR kit (MP0025; Sigma, St. Louis, MO).

## Cell Subculture

EAC lines were grown in RPMI (#31870025, Life Technologies, Carlsbad, CA) supplemented with fetal bovine serum (10%) and L-glutamine (2 mM) and incubated under standard tissue culture conditions (37 °C and 5% $CO_2$). The Barrett's esophagus line (CP-A) and the esophageal epithelial line (EPC2-hTERT) were grown in KSFM (#17005075, Gibco, Carlsbad, CA) supplemented with human recombinant epidermal growth factor (5 g/L) and bovine pituitary extract (50 mg/L). Soybean trypsin inhibitor (250 mg/L, 5 mL) was used to neutralize trypsin.

## High-Content EAC Cell Painting Assay

Cells were seeded (50 µL per well) into 384-well, CELLSTAR Cell Culture Microplates (#781091, Greiner Bio-One, Kremsmünster, Austria) and incubated under standard tissue culture conditions for 24 h before the addition of compounds. CP-A cells were seeded at 800 cells per well, SK-GT-4 cells were seeded at 1000 cells per well, and the remaining cell lines were all seeded at 1500 cells per well.

Compound source plates were made at 1000-fold assay concentration and added to the cells with an overall dilution in media of 1:1000 from source to assay plate. Library concentrations are shown in **Supplementary Table S1**.

The primary screen was carried out as a single replicate, and the validation dose-response study was in triplicate.

After 48 h of incubation in the presence of the compounds, cells were fixed by the addition of an equal volume of formaldehyde (8%, 50 µL; #28908, Thermo Scientific, Waltham, MA) to the existing media, incubated at room temperature (20 min), and washed twice in phosphate-buffered saline (PBS). Cells were then permeabilized in Triton-X100 (0.1%, 50 µL) and incubated at room temperature (20 min) followed by two more washes with PBS.

The staining solution (**Table 1**) was prepared in bovine serum albumin solution (1%). Staining solution was added to each well (25 µL) and incubated in the dark at room temperature (30 min), followed by three washes with PBS and no final aspiration. Plates were foil sealed.

## Image Acquisition

Plates were imaged on an ImageXpress micro XLS (Molecular Devices, Eugene, OR) equipped with a robotic plate loader (Scara4, PAA, UK). Four fields of view were captured per well using a 20× objective and five filters (**Table 1**). Each field of view typically contained 300 cells.

## Image Analysis

*CellProfiler 2D image analysis.* CellProfiler v3.0.0[23] image analysis software was used to segment the cells and extract 733 features per cell per image. First, the pipeline identified the nuclei from the DAPI channel and used these as seeds to aid a segmentation algorithm to identify the cell boundaries from the TxRed channel, and finally these two masks were subtracted to provide the cytoplasm. These three masks marking the cellular boundaries were then used to measure morphological features including size, shape, texture, and intensity per object across the five image channels.

*Image preprocessing.* The cell-level data were aggregated to the image level by taking the median for each measured

feature per image. Low-quality images and image artifacts were then identified and removed using image quality metrics extracted by CellProfiler. Images with fewer than 20 cells were also removed from the final analysis. For the remaining images, features were normalized on a plate-by-plate basis by dividing each feature by the median DMSO response for that feature. Features with NA values were removed, as were features with zero or near-zero variance, using the findCorrelation and nearZero functions in the R package Caret. All remaining features were scaled and centered globally by dividing by the standard deviation of each feature and subtracting the feature mean respectively. The pairwise correlations were calculated for all remaining features, and highly correlated features ($>0.95$) were removed. Finally, the image-level data were aggregated to the well (compound) level, and this was used in the analysis.

*Random forest classifier.* The random forest classifier was implemented using R's Random Forest package with the following specified parameters: ntree = 500, data stratified by class, and sample size set to the smallest class size for balance. The images from three concentrations for each compound were pooled and treated as a single class. Two different analyses were run: first, MoA prediction was implemented for each cell line individually, and second, using leave-one-out cross-validation, one EAC cell line was left out of the training set at a time and that line was run as a test set.

PCA and t-SNE were implemented using the built-in R functions prcomp and RTSNE, respectively, to visualize the clustering of the compounds for each cell line.

*Hierarchical clustering.* Z-scores and Mahalanobis scores were centered and scaled for each compound across the panel of cell lines. Spearman correlation was then used to generate a distance matrix, and hierarchical clustering was determined using complete linkage.

*NanoString transcriptomic analysis.* Cells were seeded in six-well plates and incubated for 24 h. Media were then removed and replaced with DMSO (0.1%) or methotrexate (5 μM) in DMSO and incubated for 6 h. Cells were scraped and lysed using QIAshredders (#79654, Qiagen, Hilden, Germany), and RNA was extracted by means of the Qiagen RNeasy Mini kit (#74104, Qiagen; with β-mercaptoethanol) according to the manufacturer's instructions and included a DNase digestion step (#79254, Qiagen).

Of the purified RNA, 100 ng was used as input for amplification-free RNA quantification by the NanoString nCounter Analysis System with the Human PanCancer Pathways and Metabolic Pathways panels. Raw counts were normalized to the internal positive controls and housekeeping genes using the nSolver 4.0 software.

## Results

### Assay Development

Because EAC is such a heterogeneous disease, we chose to develop a high-content phenotypic screening assay composed of a panel of EAC and tissue-matched nontransformed cell lines that captured this heterogeneity and thus provides a discovery platform for identification of novel targets and drug MoA that selectively target EAC. We assessed the amenability of 12 cell lines to high-content profiling, 10 EAC lines (JH-EsoAD1, FLO-1, MFD-1, OE33, OACM5.1, OAC-P4C, SK-GT-4, ESO51, ESO26, and OE19), and two tissue-matched nontransformed lines; a Barrett's esophagus line CP-A, and a normal esophageal squamous line immortalized by expression of telomerase EPC2-hTERT. We assessed each cell line against a list of criteria that indicated high performance for high-content screening, including cell adhesion quality, cellular morphology, proliferation in 384-well plates, image segmentation, and MoA prediction accuracy. These criteria ensure image quality/information content, high-throughput screening compatibility, and image segmentation accuracy for downstream analysis pipelines. Based on suitable cell adhesion and morphological properties, we took forward the following eight cell lines for high-content assay development, including image segmentation and machine learning analysis: CP-A, EPC2-hTERT, FLO-1, JH-EsoAD1, MFD-1, OAC-P4C, OE33, and SK-GT-4 (**Fig. 1**).

The published Cell Painting protocol[17,18] was adapted for our cell lines specifically as follows: the MitoTracker DeepRed was originally added before the cells were fixed; however, morphological changes have been seen in certain cell lines upon the addition of MitoTracker.[24] Therefore, we opted to fix the cells first and add all of the Cell Painting reagents together after fixation to prevent artifactual morphological changes due to cell staining and to reduce complexity for robotic handling in a high-throughput setting. This also necessitated that we reoptimize the dye concentrations across our cell panel. Here, we increased the MitoTracker DeepRed concentration and reduced the concentration of Hoechst, Concanavalin A, and Wheat Germ Agglutinin and switched to a different phalloidin supply (**Table 1**).

### Machine Learning

Standard assay quality control metrics such as Z′Factor are unsuitable for multiparametric assays, particularly cell-based phenotypic profiling assays in which a desired phenotype is unknown and/or there is a lack of positive controls.[25–27] To assess assay quality from a compound MoA profiling perspective, we used MoA prediction accuracy on a small well-annotated reference library of compounds with

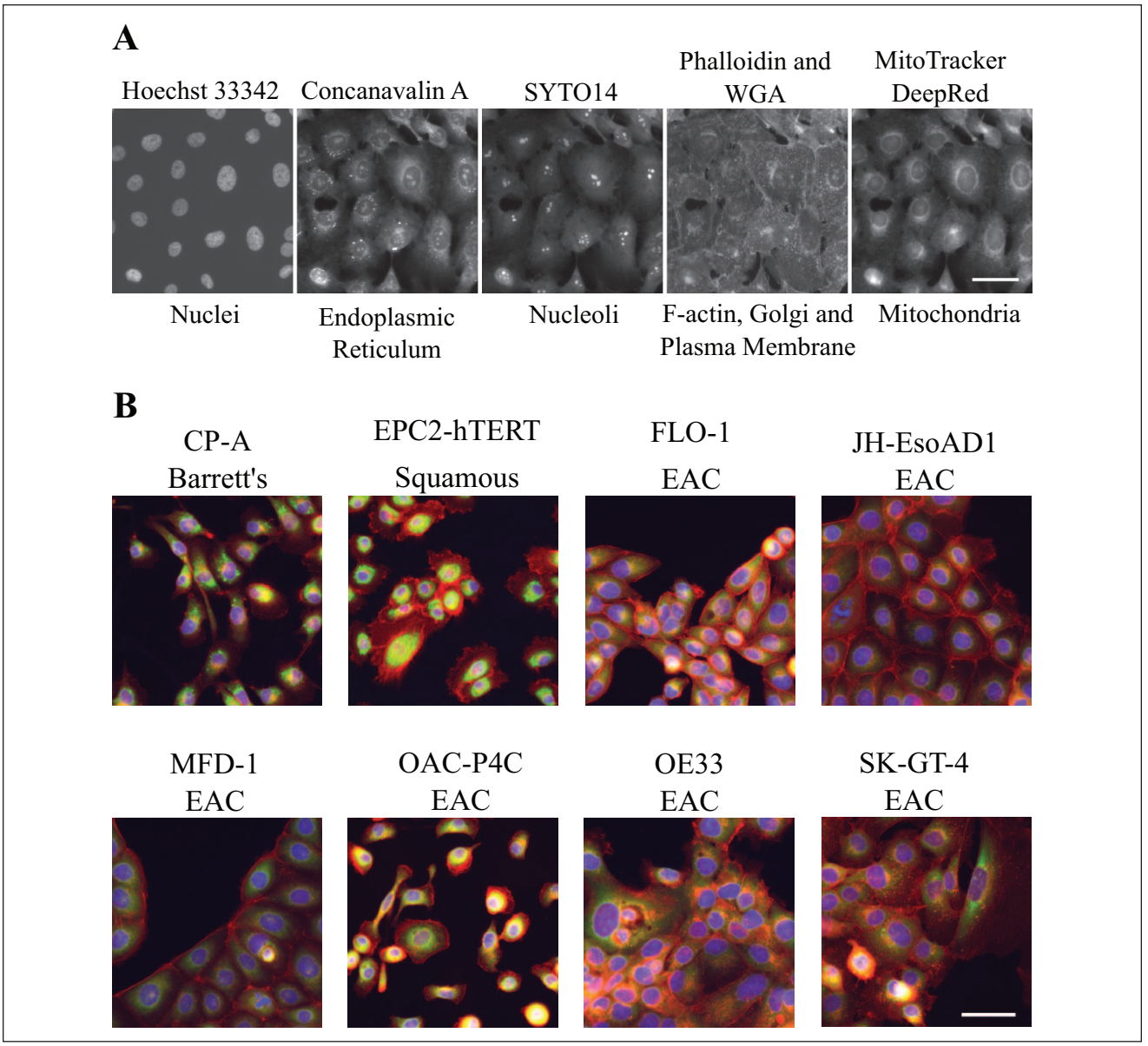



**Figure 1.** Cell Painting assay and the cell panel. (**A**) The five channels imaged in the Cell Painting assay for the representative cell line SK-GT-4, with dyes and cellular structures labeled. Scale bar is 50 µm. (**B**) Color combined representative control (DMSO) images of the eight cell lines in the cell panel: DAPI (blue), TxRED (red), FITC (green). Scale bar is 50 µm. See **Table 1** for additional details about the stains and channels imaged.

well-defined, known MoA (**Suppl. Table S2**). For this, we trained a random forest classifier using the CellProfiler extracted phenotypic information from the images of cells treated with the reference set of compounds.

Accuracy in the ability to predict MoA was used to assess whether the EAC and tissue-matched control cell lines were amenable to the phenotypic profiling assay, further validate whether image segmentation was accurate, and ensure that the phenotypic information extracted was relevant and broad enough to allow accurate prediction of MoA. To robustly evaluate compound selectivity and MoA across our heterogeneous panel of genetically distinct EAC cells, it was particularly important to assess the performance of each individual cell line and ensure that one cell line did not perform significantly better or worse than the others. A characteristic of EAC cell lines (OE33, MFD-1, and SK-GT-4 in particular) is the migration and formation of cell clumps, which are challenging to segment accurately by automated image analysis. Here we wanted to confirm that they were equal to the rest of the panel and suitable for the assay pipeline. OAC-P4C is a particularly morphologically heterogeneous line, so it was also important to ensure

that image-level data can be used for phenotypic compound profiling in these types of cell lines.

To visualize the phenotypic information extracted, we performed two data reduction methods, PCA and T-SNE, on the well-level data for the small reference library of compounds and plotted the first two components, colored by mechanistic class. PCA is a linear feature extraction technique, projecting the data in a lower-dimensional space while preserving the global structure of the higher-dimensional data.[28,29] t-SNE is a nonlinear technique and is capable of capturing local as well as global structure.[30] Input feature importance can also be examined when using PCA but not t-SNE. Given the complexity of the relationship between the features extracted from the images using both techniques allowed us to gauge feature importance, assess both a linear and nonlinear technique as there were no prior assumptions about the relationships between features, and look at both local and global trends in the data set. The results demonstrate that distinct compound classes generally cluster together. However, the strength of the phenotypic response varied across compound classes with some, such as the statins, producing a less distinct response than others. PCA clustering shows the statins are much closer to the DMSO controls (**Fig. 2**; **Suppl. Fig. S1**).

We next optimized a random forest classifier to test the MoA prediction on our reference library of well-annotated compounds. The extracted features from three concentrations of each compound were pooled and used to train the classifier. We chose 0.1, 1, and 10 μM, because using a broad range of concentrations means that each compound does not need to be optimized individually across each cell line.

When trained and tested on each individual cell line the average out-of-bag error[31,32] was 20.38% across the entire panel of cell lines, ranging from 12% to 27%. The overall prediction accuracy for each cell line ranged from 73% to 88% across all eight compound classes, demonstrating the assay was well optimized across the panel. The weakest cell line was the SK-GT-4.

To confirm that the classifier was not overfitting, we used leave-one-out cross-validation.[33,34] We implemented leave-one-cell-line-out and trained it on five of the EAC lines, testing on the remaining line. Here, as expected, it performed less well overall. However, the accuracy for each cell line ranged from 58% to 71% (**Suppl. Fig. S2**), indicating the ability of this classifier to be transferred to new cell lines despite having no prior training on them and thus the potential for the application of the classifier across a broader panel of cell lines without the need to train each cell line individually.

Overall, the accuracy of the machine learning demonstrates that the phenotypic profiling assay is of high quality across all eight cell lines, including morphologically heterogeneous cells, and feature extraction produces meaningful data for phenotypic analysis. The phenotypic profiling assay can therefore be applied to provide an initial evaluation of MoA of hit compounds influencing EAC cell proliferation, survival, and morphology. As such, our multiparametric high-content phenotypic profiling assay may prove useful in the prioritization of compound hits, which represent novel MoAs, and the deprioritization of compounds, which represent undesirable MoAs for subsequent medicinal chemistry and target deconvolution investments. We therefore prioritized the full panel of eight lines that passed our quality control criteria (six EAC lines with diverse genetic backgrounds, a Barrett's esophagus line, and a nontransformed squamous esophageal line) for a phenotypic screen of 19,555 small molecules.

## *Small-Molecule Screen*

A total of 19,555 small molecules, including approved drugs, were profiled against our panel of eight cell lines using the ImageXpress microXL high-content imaging platform. Cells were treated with the commercially available Prestwick Chemical Library of 1280 mostly off-patent drugs, the LOPAC library of pharmacologically active compounds (1280 compounds), a proprietary diverse chemical library provided by CRUK Therapeutics Discovery Laboratories (Cambridge; 13,408 compounds), the BioAscent library of 3200 compounds, and bespoke libraries of 387 target-annotated compounds and chemical probes. The primary phenotypic screen across all eight cell lines encompassed 512 × 384 well plates, 3.9 million images, and 36 TB of data in total. Image analysis was performed using CellProfiler across a computer cluster.

Using a panel of cell lines better represents a heterogeneous disease and allowed us to identify compounds that demonstrated selective activity across multiple EAC lines and not in the tissue-matched control. We ran two parallel analyses for primary hit selection against the EAC lines: one based on broad, morphological, phenotypic changes and the other on cell growth and survival using nuclei count. At cytotoxic concentrations, there are few attached cells, and these are often rounded up, leading to a lack of information in the images. Therefore, images with 20 or fewer cells were removed from the morphological analysis.

We began our analysis with a subset of 3000 annotated compounds (excluding the CRUK Therapeutics Discovery Laboratories and BioAscent lead-like molecules).

To identify compounds inducing strong phenotypic changes, we used PCA on the feature data to reduce the dimensions and then calculated the Mahalanobis distance to the DMSO controls for the first 15 principal components, which explain approximately 90% of the variation in the data across each cell line. The Mahalanobis distance measures the distance of each point from the data distribution

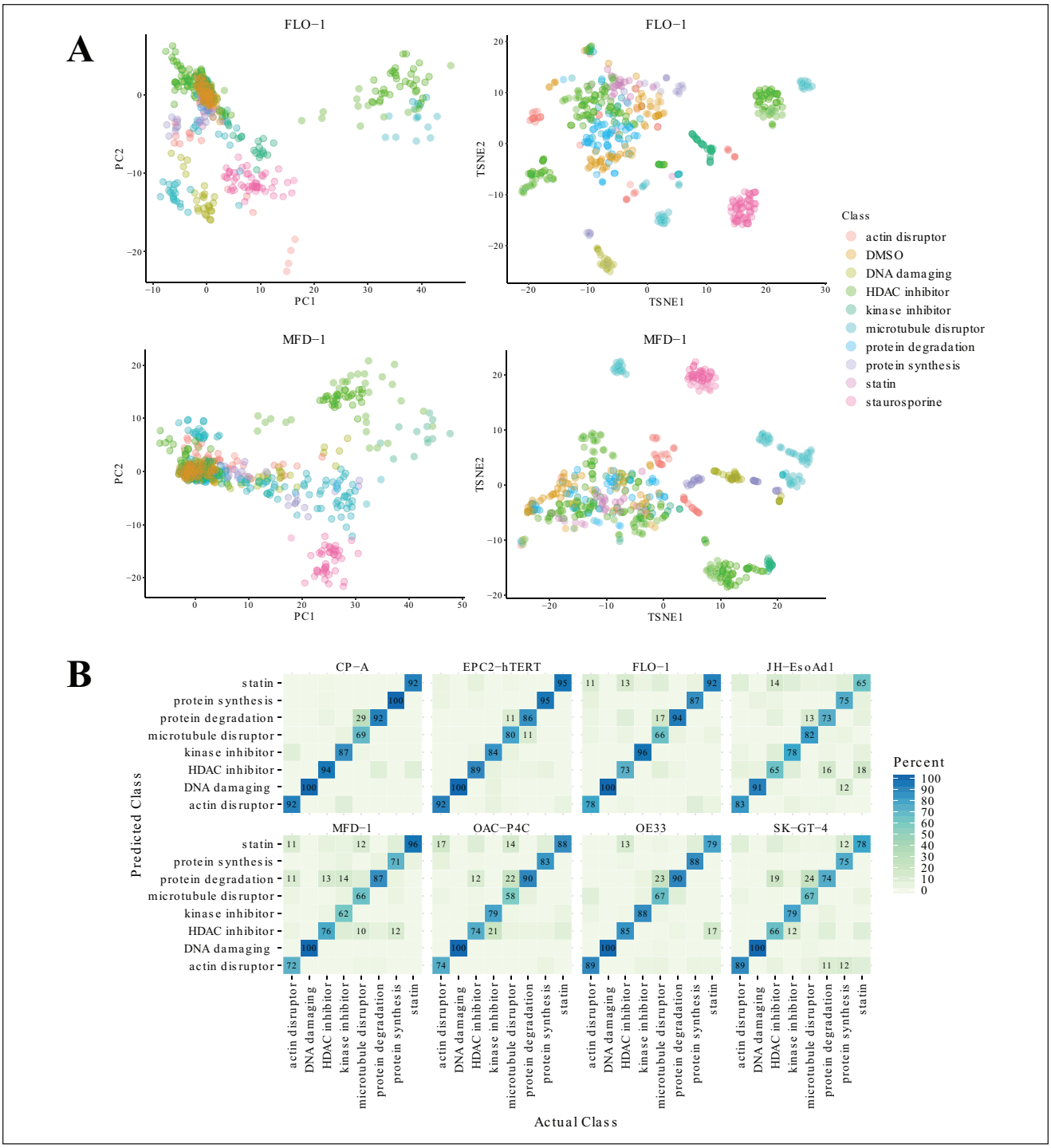



**Figure 2.** Reference library clustering and machine learning. (**A**) The first two components of principal component analysis (PCA) and t-distributed stochastic neighbor embedding (T-SNE) for the reference library compound treatments for the esophageal adenocarcinoma lines FLO-1 and MFD-1 (see **Suppl. Fig. S1** for remaining cell lines). Points are colored by mechanistic class, and multiple compounds concentrations are plotted. (**B**) Random forest classifier: confusion matrices of prediction accuracies per cell line in the cell panel for the reference library of compounds. Diagonal values show class sensitivities.

(in this case, the DMSO controls). The data distribution takes into account the mean and the spread of the data points using the covariance matrix as a normalization factor.[35] It therefore addresses problems of both scale and correlation of the variables and is particularly beneficial for large multivariate data sets. This leads to elliptic rather than circular

decision boundaries, as is the case for Euclidean. We therefore chose to use it as an unbiased metric of compound activity upon each cell line in the screen.

Phenotypic analysis identified 62 compounds that selectively target two or more of the EAC lines over the non-transformed esophageal cells. Clustering the cell panel's responses to these molecules showed a number of phenotypic clusters enriched with similar pharmacologic classes, including HDAC inhibitors, microtubule disruptors, and antimetabolites, suggesting that hits have clustered mechanistically (**Fig. 3A**; **Suppl. Fig. S3**).

Based on cell growth and survival (i.e., nuclei count), we identified 27 compounds that were selectively active in two or more of our EAC lines. Here, hits were defined as having a z-score of −3 or greater in the EAC lines and a difference of at least 2 in one or both of the control cell lines (e.g., for a hit with a z-score of −3 in an EAC line, the z-score in the EPC-2 would have to be greater than or equal to -1). This comparison was made between each EAC line and the control lines to define hits and then selected if they were selectively active in at least two EAC lines across the panel (**Fig. 3B**).

Compounds from the growth and survival analysis cluster into several therapeutic classes, suggesting mechanistic pathways that may be selective for EAC cell growth and survival. Classes include antimetabolites and HDAC inhibitors. These classes were also identified in the morphometric phenotypic analysis (**Suppl. Fig. S3**).

We performed hierarchical clustering of cell-line responses to the compounds, as determined by the Mahalanobis metric (morphometric phenotypic analysis) and the z-scores (nuclei count; **Fig. 3C, D**), enabling pharmacologic discrimination of cell lines. These results show that the control cell lines, EPC2-hTERT and CP-A, can be discriminated from the EAC panel based on global drug screening data, providing confidence that our high-content Cell Painting assay can identify compounds with selectivity for EAC over the tissue-matched control lines.

## *Antimetabolites Are Selectively Lethal to EAC Cells*

From the subset of 3000 annotated compounds, we identified the drug methotrexate and three other structurally related antimetabolites, pemetrexed, raltitrexed, and aminopterin, as highly selective for EAC cell lines relative to tissue-matched control CP-A and EPC2-hTERT cells in both the nuclei count and morphological phenotypic analyses. We therefore validated this class of compound for dose-dependent activity. Aminopterin was removed from further analysis because of its toxicity profile in the clinic[36]; however, it showed potent activity in an initial dose response in the EAC lines, validating it as a hit from our screen (results not shown).

Nuclei count dose responses for methotrexate, pemetrexed, and raltitrexed demonstrated strong selectivity against the EAC lines and showed minimal cytotoxic or phenotypic activity in either the CP-A or the EPC2-hTERT line even at 10 μM (**Fig. 4A**; **Suppl. Table S3**), validating our hit selection criteria.

Multiparametric phenotypic dose-response profiles of the antimetabolites overlaid on the reference library of annotated compounds (**Suppl. Table S2**) showed strong dose-dependent phenotypic changes, moving from phenotypically inactive (clustering with DMSO controls) to clustering with the DNA-damaging agents at active concentrations (**Fig. 4B**; **Suppl. Fig S4**) in all but the JH-EsoAD1 and MFD-1 lines. All three compounds also showed little or no effect in the control lines EPC2-hTERT and CP-A, clustering closely with the DMSO controls at all concentrations tested.

Class probabilities from the pretrained machine-learning model for each of the compounds predicted that they belong to the DNA damage class for all but the MFD-1 and JH-EsoAD1 lines (**Fig. 4C**), consistent with the clustering above. Probabilities also increased in a dose-dependent manner, indicating that cellular phenotypic activity follows a linear on-target dose-response relationship. These results further confirm the ability of the Cell Painting assay to accurately predict the MoA of validated hit compounds.

NanoString differential expression analysis[37] revealed methotrexate treatment caused a significant reduction in the expression of Histone H3 subunits (HIST1H3B, HIST1H3G, HISTH3H; **Fig. 4D**) in the sensitive cell lines only, with no effect in either of the tissue-matched controls (**Suppl. Table S4)**. Several other genes changed with methotrexate treatment, but none were significant. Further mechanistic studies are required to further elucidate how and if such expression changes confer selectivity to methotrexate.

## *Toward Novel Therapies and Targets for EAC*

From a subset of 13,000 small-molecule compounds with unknown targets, we further identified a small number of compound hits from our diverse chemical library that showed potent and selective activity for the EAC cell lines. Compound 1 was selective for the OAC-P4C and MFD-1 cells (**Fig. 5A**; **Suppl. Fig. S5**), and machine-learning probabilities for all classes were low (**Fig. 5B**). Compound 2 induced a strong phenotypic dose response in the OAC-P4C and OE33 cell lines only and did not cluster with the reference library of known MoA (**Fig. 5A**; **Suppl. Fig. S6**). Machine learning predicted it to be DNA damaging (91% probability) in the OAC-P4C cells; however, its clustering was distinct, and the machine-learning probability that it is DNA damaging in the OE33 cell line was only 52% (**Fig. 5B**). Therefore. it may in fact represent a novel MoA or be acting to cause DNA damage in a novel way. This indicates that these compounds may be selectively targeting novel esophageal cancer biology. Subsequent transcriptomic and

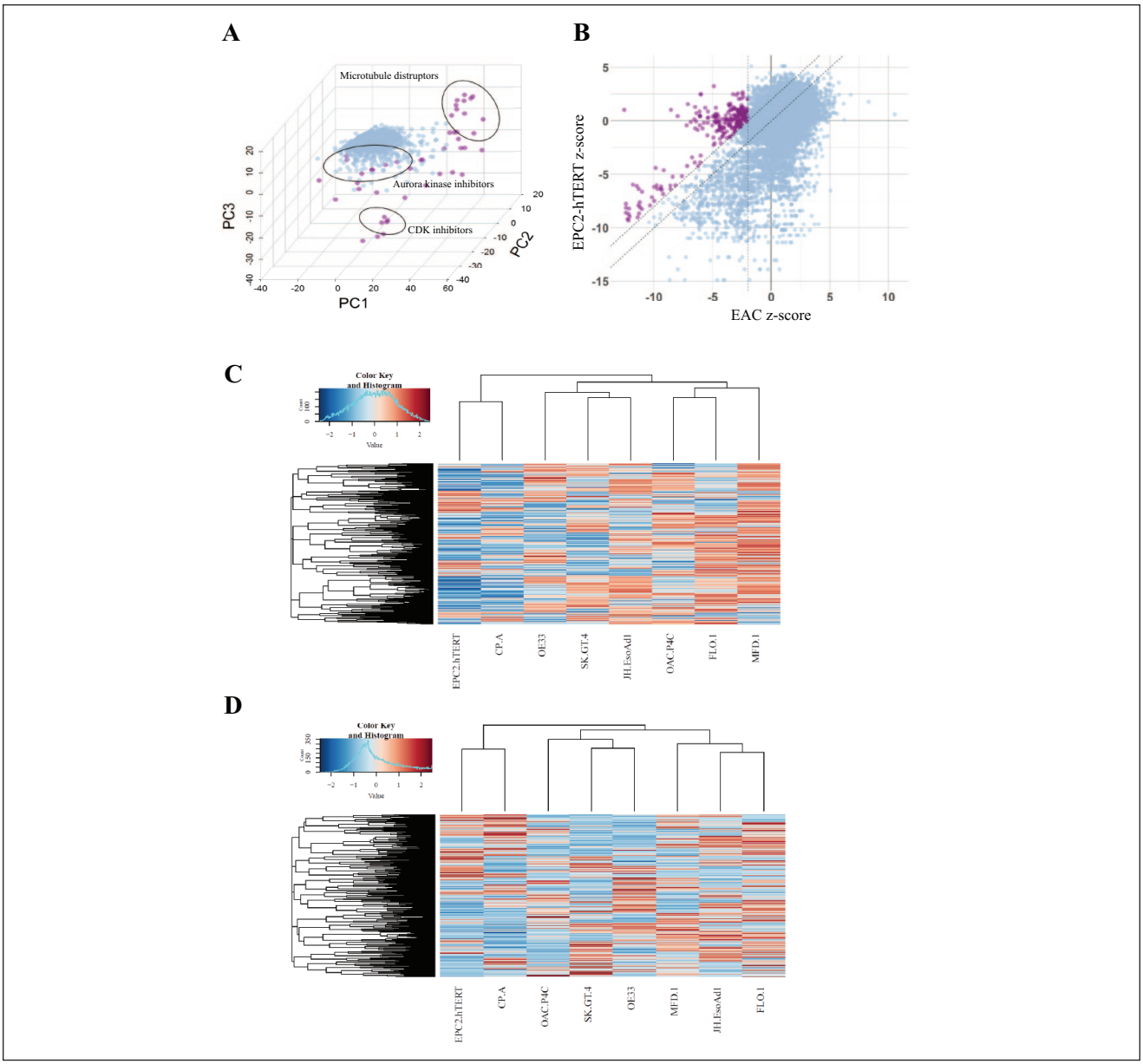



**Figure 3.** Hit analysis. (**A**) The first three components of principal component analysis (PCA) for exemplar data from the esophageal adenocarcinoma (EAC) cell line; JH-EsoAD1. Hits (purple) are defined as having a Mahalanobis distance of greater than 1500 from the DMSO controls. (**B**) Z-score plot for all EAC lines overlaid versus the EPC2-hTERT esophageal squamous control line. Hits (purple) are defined as having a z-score of −3 or greater in the EAC lines and showing selectivity of at least 2 z-scores compared with the EPC2-hTERT line. (**C**) Z-score hierarchical clustering of the cell panels' response to compounds. (**D**) Mahalanobis distance clustering of phenotypic response to compound treatments across cell lines.

proteomic pathway analysis and target deconvolution studies may reveal the mechanistic pathways involved.

## Discussion

Conventional target-directed drug discovery strategies remain to make any impact on the discovery and translation of effective new treatments for esophageal cancer patients. Key challenges in esophageal cancer include a highly heterogeneous genetic landscape with few mutations in oncogenic drivers, thereby confounding the identification of a clear drug-target hypothesis and modern personalized medicine strategies. In this study, we sought to adapt and evaluate the utility of an advanced high-content phenotypic

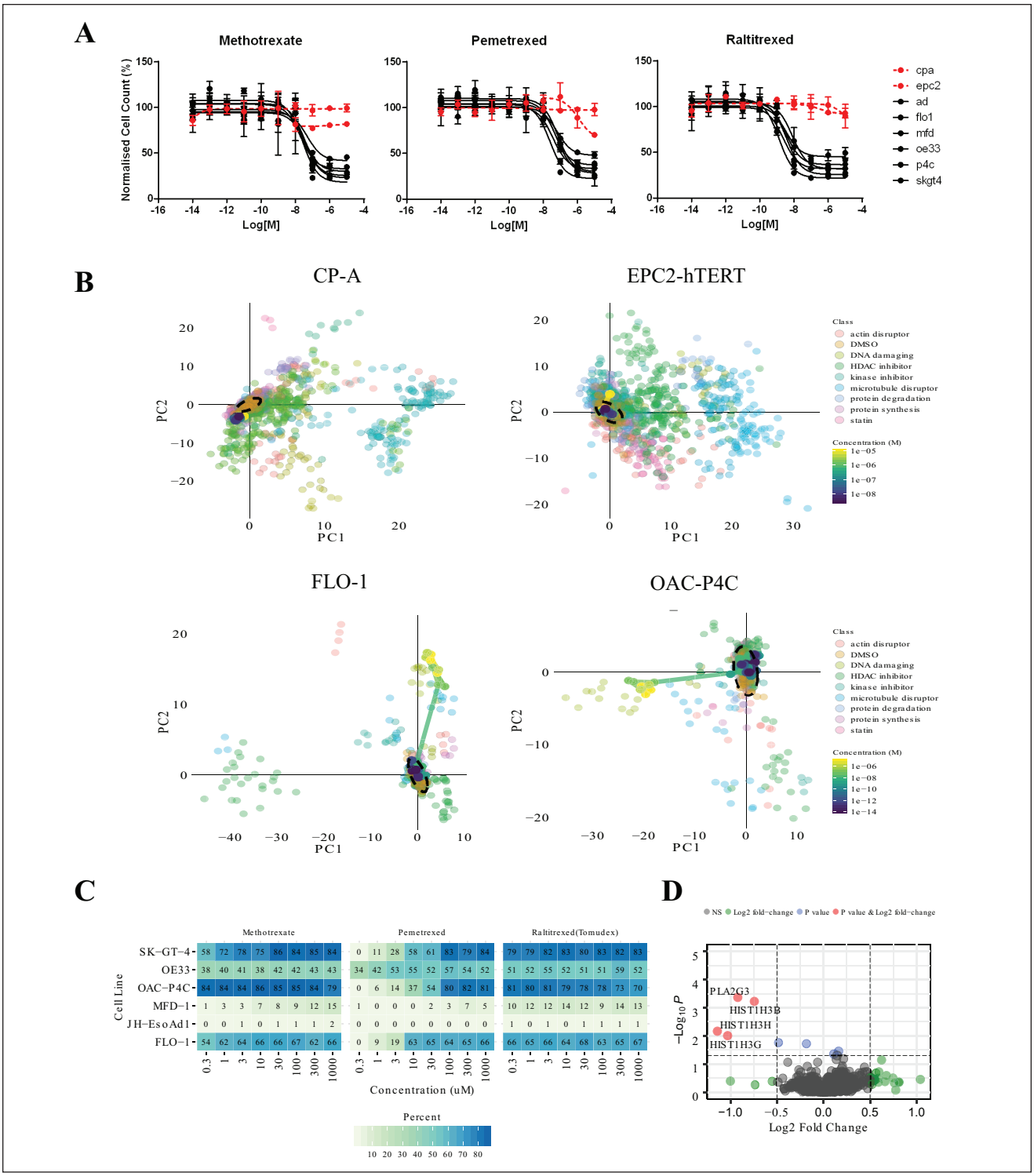



**Figure 4.** Antimetabolite evaluation. (**A**) Dose responses for methotrexate, pemetrexed, and raltitrexed across a panel of cell lines. (**B**) Principal component analysis of dose responses overlaid on the reference library for methotrexate in two resistant lines (CP-A and EPC2-hTERT) and two sensitive lines (FLO-1 and OAC-P4C). (**C**) Probabilities expressed as percentages for DNA damaging class for each cell line and each of methotrexate, pemetrexed, and raltitrexed. (**D**) Differential expression analysis for methotrexate treatment (5 μM, 6 h) for FlO-1, SK-GT-4, and OE33 cell lines. Red indicates genes reaching both the *p*-value and fold-change threshold, blue indicates genes that reached the *p*-value threshold, and green indicates genes that reached the fold-change threshold. $p = 0.05$, $\log_2$-fold change $= 0.5$.

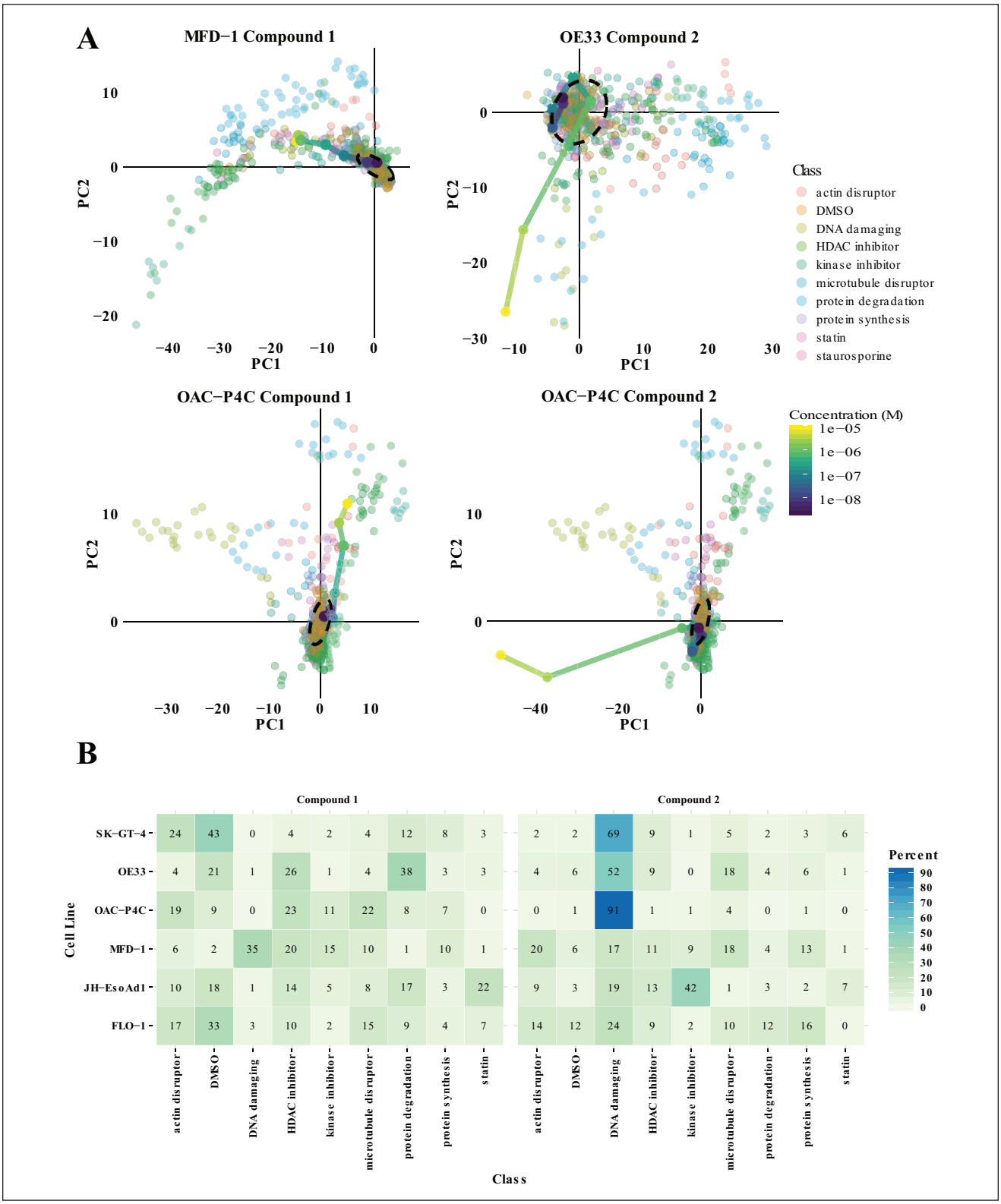



**Figure 5.** Phenotypic analysis of novel compounds. (**A**) Principal component analysis of compound 1 and compound 2 dose responses overlaid on the reference library for the two most sensitive cell lines for each compound. (**B**) Probabilities expressed as percentages for compound 1 and compound 2 (10 µM) belonging to each class in the reference library for each cell line.

screening method as an empirical strategy for identifying novel drug targets, MoAs, and pharmacologic classes that target EAC.

Here we have shown that combining high-content screening and image informatics with machine learning can be effective for the identification and mechanistic characterization of hit compounds with selective activity on EAC cell phenotypes. Most multiparametric high-content screening assays and associated machine-learning methods used to predict drug MoA are typically performed on a single cell line. In this study, we have further shown that this format can be applied to heterogeneous panels of cancer cell lines and normal tissue-matched control cells for the identification and prioritization of hit compounds and MoA, which demonstrate selective activity for EAC cells.

Machine learning can be implemented as a tool for multiparametric phenotypic assay quality control (e.g., confirming if the assay is suitable as a discovery platform to classify specific cell phenotypes and elucidate MoA) as well as a tool for MoA deconvolution of hit compounds. Our results demonstrate that this can be standardized across heterogeneous panels of cells.

Following one class of compounds identified in our primary phenotypic screen of 19,555 small molecules tested across all eight esophageal cell lines, we validated antimetabolites as selectively lethal to the EAC lines in vitro following dose-response studies. Using the multiparametric phenotypic information to generate phenotypic dose responses, combined with a reference library of compounds, machine learning, and clustering techniques, we demonstrated the ability to study/predict the MoA of hits from the screen. Here we validated this technique using the antimetabolite hit compounds (methotrexate, pemetrexed, and raltitrexed), showing DNA damage as a likely MoA for the selectivity of these compounds, which is consistent with the literature.[38,39] These results, together with our identification of hit compounds from our diverse chemical set, which are not classified by our reference set of known MoAs, demonstrates the impact of phenotypic screening in combination with machine learning for MoA studies. This strategy will be used to assess and prioritize novel small-molecule hits from the diverse chemical library screen for further mechanistic studies. From our primary phenotypic screen, we have identified in total 75 compounds that match our hit selection criteria for selective activity across the EAC panel. These 75 hits are an accumulation of the 62 compounds defined by cell morphometric phenotypic analysis and 27 compounds defined by cell proliferation and survival (nuclei count) analysis, with 14 compounds overlapping. The 75 hits shall be further progressed through dose-response studies and secondary assays to confirm and prioritize classes of selective compounds for subsequent drug repurposing and or drug discovery studies.

In addition, using bioinformatic approaches, we hope that integration of phenotypic data with genetic data across our panel of diverse cell lines may provide insight into the selective activity of phenotypic hits and generate the basis for future genetic biomarker–based clinical trials in EAC.

Overall, our high-content EAC assay has proven effective in the identification and mechanistic characterization of hit compounds, demonstrating its utility as a novel empirical strategy for the discovery of new therapeutic targets, chemical starting points, and repurposing of existing drug classes to reignite drug discovery and development in EAC.

## Acknowledgments

We thank Fabrice Turlais and Mathew Calder from CRUK-Therapeutics Discovery Labs for provision of compound libraries and Anil Rustig, University of Pennsylvania, for provision of EPC2-hTERT cells.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by an MRC-Institute of Genetics and Molecular Medicine PhD studentship award to R.E.H. and the Anne Forrest Fund for Oesophageal Cancer Research and a CRUK Small Molecule Drug Discovery project award to N.O.C.

## ORCID iD

Rebecca E. Hughes https://orcid.org/0000-0002-0590-4135

## References

1. Herszényi, L.; Tulassay, Z. Epidemiology of Gastrointestinal and Liver Tumors. *Eur. Rev. Med. Pharmacol. Sci*. **2010**, *14*, 249–258.
2. Pennathur, A.; Gibson, M. K.; Jobe, B. A.; et al. Oesophageal Carcinoma. *Lancet* **2013**, *381*, 400–412.
3. Napier, K. J.; Scheerer, M.; Misra, S. Esophageal Cancer: A Review of Epidemiology, Pathogenesis, Staging Workup and Treatment Modalities. *World J. Gastrointest. Oncol*. **2014**, *6*, 112–120.
4. Reynolds, J.; Preston, S.; O'Neill, B.; et al. ICORG 10-14: NEOadjuvant Trial in Adenocarcinoma of the OEsophagus and OesophagoGastric Junction International Study (NEO-AEGIS). *BMC Cancer* **2017**, *17*, 401–401.
5. Alderson, D.; Cunningham, D.; Nankivell, M.; et al. Neoadjuvant Cisplatin and Fluorouracil versus Epirubicin, Cisplatin, and Capecitabine followed by Resection in Patients with Oesophageal Adenocarcinoma (UK MRC OE05): An Open-Label, Randomised Phase 3 Trial. *Lancet Oncol*. **2017**, *18*, 1249–1260.
6. Noble, F.; Lloyd, M. A.; Turkington, R.; et al. Multicentre Cohort Study to Define and Validate Pathological Assessment

of Response to Neoadjuvant Therapy in Oesophagogastric Adenocarcinoma. *Br. J. Surg*. **2017**, *104*, 1816–1828.
7. O'Neill, J. R.; Kennedy, E. D.; Save, V.; et al. Patients Unfit for Neoadjuvant Therapy May Still Undergo Resection of Locally Advanced Esophageal or Esophagogastric Junctional Cancer with Acceptable Oncological Results. *Int. J. Surg. Oncol*. **2017**, *2*, e09.
8. Secrier, M.; Li, X.; de Silva, N.; et al. Mutational Signatures in Esophageal Adenocarcinoma Define Etiologically Distinct Subgroups with Therapeutic Relevance. *Nat. Genet*. **2016**, *48*, 1131–1141.
9. Okines, A. F. C.; Ashley, S. E.; Cunningham, D.; et al. Epirubicin, Oxaliplatin, and Capecitabine With or Without Panitumumab for Advanced Esophagogastric Cancer: Dose-Finding Study for the Prospective Multicenter, Randomized, Phase II/III REAL-3 Trial. *J. Clin. Oncol*. **2010**, *28*, 3945–3950.
10. Bang, Y.-J.; Van Cutsem, E.; Feyereislova, A.; et al. Trastuzumab in Combination with Chemotherapy versus Chemotherapy Alone for Treatment of HER2-Positive Advanced Gastric or Gastro-Oesophageal Junction Cancer (ToGA): A Phase 3, Open-Label, Randomised Controlled Trial. *Lancet* **2010**, *376*, 687–697.
11. Dutton, S. J.; Ferry, D. R.; Blazeby, J. M.; et al. Gefitinib for Oesophageal Cancer Progressing after Chemotherapy (COG): A Phase 3, Multicentre, Double-Blind, Placebo-Controlled Randomised Trial. *Lancet Oncol*. **2014**, *15*, 894–904.
12. Hecht, J. R.; Bang, Y.-J.; Qin, S. K.; et al. Lapatinib in Combination with Capecitabine Plus Oxaliplatin in Human Epidermal Growth Factor Receptor 2–Positive Advanced or Metastatic Gastric, Esophageal, or Gastroesophageal Adenocarcinoma: TRIO-013/LOGiC—A Randomized Phase III Trial. *J. Clin. Oncol*. **2016**, *34*, 443–451.
13. Dulak, A. M.; Schumacher, S. E.; van Lieshout, J.; et al. Gastrointestinal Adenocarcinomas of the Esophagus, Stomach, and Colon Exhibit Distinct Patterns of Genome Instability and Oncogenesis. *Cancer Res*. **2012**, *72*, 4383–4393.
14. Paterson, A. L.; O'Donovan, M.; Provenzano, E.; et al. Characterization of the Timing and Prevalence of Receptor Tyrosine Kinase Expression Changes in Oesophageal Carcinogenesis. *J. Pathol*. **2013**, *230*, 118–128.
15. Perlman, Z. E.; Slack, M. D.; Feng, Y.; et al. Multidimensional Drug Profiling By Automated Microscopy. *Science*. **2004**, *306*, 1194–1198.
16. Bickle, M. The Beautiful Cell: High-Content Screening in Drug Discovery. *Anal. Bioanal. Chem*. **2010**, *398*, 219–226.
17. Gustafsdottir, S. M.; Ljosa, V.; Sokolnicki, K. L.; et al. Multiplex Cytological Profiling Assay to Measure Diverse Cellular States. *PLoS One* **2013**, *8*, e80999.
18. Bray, M.-A.; Singh, S.; Han, H.; et al. Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes. *Nat. Protoc*. **2016**, *11*, 1757–1774.
19. Caie, P. D.; Walls, R. E.; Ingleston-Orme, A.; et al. High-Content Phenotypic Profiling of Drug Response Signatures across Distinct Cancer Cells. *Mol. Cancer Ther*. **2010**, *9*, 1913–1926.
20. Ljosa, V.; Caie, P. D.; ter Horst, R.; et al. Comparison of Methods for Image-Based Profiling of Cellular Morphological Responses to Small-Molecule Treatment. *J. Biomol. Screen*. **2013**, *18*, 1321–1329.
21. Young, D. W.; Bender, A.; Hoyt, J.; et al. Integrating High-Content Screening and Ligand-Target Prediction to Identify Mechanism of Action. *Nat. Chem. Biol*. **2008**, *4*, 59–68.
22. Harada, H.; Nakagawa, H.; Oyama, K.; et al. Telomerase Induces Immortalization of Human Esophageal Keratinocytes without P16INK4a Inactivation. *Mol. Cancer Res*. **2003**, *1*, 729–738.
23. Carpenter, A. E.; Jones, T. R.; Lamprecht, M. R.; et al. CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes. *Genome Biol*. **2006**, *7*, R100.
24. Warchal, S. J.; Dawson, J. C.; Carragher, N. O. Development of the Theta Comparative Cell Scoring Method to Quantify Diverse Phenotypic Responses between Distinct Cell Types. *Assay Drug Dev. Technol*. **2016**, *14*, 395–406.
25. Kummel, A.; Gubler, H.; Gehin, P.; et al. Integration of Multiple Readouts into the Z′ Factor for Assay Quality Assessment. *J. Biomol. Screen*. **2010**, *15*, 95–101.
26. Birmingham, A.; Selfors, L. M.; Forster, T.; et al. Statistical Methods for Analysis of High-Throughput RNA Interference Screens. *Nat. Methods* **2009**, *6*, 569–575.
27. Sui, Y.; Wu, Z. Alternative Statistical Parameter for High-Throughput Screening Assay Quality Assessment. *J. Biomol. Screen*. **2007**, *12*, 229–234.
28. Pearson, K. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *London Edinburgh Dublin Philos. Mag. J. Sci*. **1901**, *2*, 559–572.
29. Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol*. **1933**, *24*, 417–441.
30. Van Der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J Machine Learn. Res*. **2008**, *9*, 2579–2605.
31. Breiman, L. Out-of-Bag Estimation, 1996. Statistics Department University of California Berkeley, CA. https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf (accessed April 21, 2020).
32. Janitza, S.; Hornung, R. On the Overestimation of Random Forest's out-of-Bag Error. *PLoS One* **2018**, *13*, e0201904.
33. Larson, S. C. The Shrinkage of the Coefficient of Multiple Correlation. *J. Educ. Psychol*. **1931**, *22*, 45–55.
34. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc*. **1974**, *36*, 111–147.
35. Mahalanobis, P. C. On the Generalized Distance in Statistics. *Proc. Natl. Inst. Sci. India* **1936**, *2*, 49–55.
36. Bertino, J. R.; Göker, E.; Gorlick, R.; et al. Resistance Mechanisms to Methotrexate in Tumors. *Oncologist* **1996**, *1*, 223–226.
37. Wang, H.; Horbinski, C.; Wu, H.; et al. NanoStringDiff: A Novel Statistical Method for Differential Expression Analysis Based on NanoString NCounter Data. *Nucleic Acids Res*. **2016**, *44*, e151.
38. Martin, S. A.; Mccarthy, A.; Barber, L. J.; et al. Methotrexate Induces Oxidative DNA Damage and Is Selectively Lethal to Tumour Cells with Defects in the DNA Mismatch Repair Gene MSH2. *EMBO Mol. Med.* **2009**, *1*, 323–327.
39. Xie, L.; Zhao, T.; Cai, J.; et al. Methotrexate Induces DNA Damage and Inhibits Homologous Recombination Repair in Choriocarcinoma Cells. *Onco. Targets Ther*. **2016**, *9*, 7115–7122.

CellPress

Cell Chemical Biology

Review

# High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need

Rebecca E. Hughes,[1] Richard J.R. Elliott,[1] John C. Dawson,[1] and Neil O. Carragher[1,*]
[1]Cancer Research UK Edinburgh Centre, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XR, UK
*Correspondence: n.carragher@ed.ac.uk
https://doi.org/10.1016/j.chembiol.2021.02.015

## SUMMARY

Conventional thinking in modern drug discovery postulates that the design of highly selective molecules which act on a single disease-associated target will yield safer and more effective drugs. However, high clinical attrition rates and the lack of progress in developing new effective treatments for many important diseases of unmet therapeutic need challenge this hypothesis. This assumption also impinges upon the efficiency of target agnostic phenotypic drug discovery strategies, where early target deconvolution is seen as a critical step to progress phenotypic hits. In this review we provide an overview of how emerging phenotypic and pathway-profiling technologies integrate to deconvolute the mechanism-of-action of phenotypic hits. We propose that such in-depth mechanistic profiling may support more efficient phenotypic drug discovery strategies that are designed to more appropriately address complex heterogeneous diseases of unmet need.

## INTRODUCTION

Despite the many remarkable successes of modern target-based drug discovery (TDD) technologies, recent analysis of drug approval rates between 2003 and 2011 indicates that attrition rates during clinical development continues to increase (Hay et al., 2014). Across all human therapeutic areas, approximately 10% of drug candidates that undergo initial clinical evaluation are successful in obtaining regulatory approval (Hay et al., 2014), highlighting the inefficiency of preclinical drug discovery to predict clinical success. These limitations of contemporary drug discovery strategies are brought into sharper focus when specifically considering diseases of unmet therapeutic need. The list of unmet therapeutic areas include many of the most common human neurological disorders and several cancers (Scavone et al., 2019). The failure to develop any effective new treatment options for these diseases over the past 30 years further highlights current limitations of TDD and the urgent need to consider alternative strategies and a less-conventional drug discovery philosophy to tackle diseases of unmet therapeutic need.

Advances in next-generation sequencing (NGS) have revealed remarkable molecular heterogeneity in underlying disease mechanisms within and between patients for many disease types. Thus, the conventional ''one disease, one target, one drug'' approach grossly oversimplifies disease mechanisms, which are in fact driven by complex signaling networks within the biological interactome (Aguirre-Plans et al., 2018; Hopkins, 2008). Understanding target biology and drug mechanism-of-action (MOA) within a disease-relevant context is becoming realized with new advances in human cell-based phenotypic assay technologies integrated with advances in molecular biology, genomics, proteomics, and computational biology to advance the fields of network pharmacology and emerging phenomic drug discovery (Hopkins, 2008; Wang et al., 2013; Warchal et al., 2016; Zhao and Iyengar, 2012). These advances are revealing phenotypic and pathway network robustness and plasticity, indicating that highly selective compounds may exhibit weak clinical efficacy for complex diseases relative to multi-target drugs or drug combinations. Indeed, many effective drugs act via modulation of multiple proteins resulting from polypharmacology rather than via their primary targets (Ericson et al., 2008; Lim et al., 2016; Lin et al., 2019).

TDD is complemented by phenotypic drug discovery (PDD) strategies, which have recently re-emerged as a valuable drug discovery approach (Moffat et al., 2014; Swinney and Anthony, 2011). PDD describes the screening and selection of hit or lead compounds based on quantifiable phenotypic endpoints from cell-based assays or model organisms without prior knowledge of the drug target. PDD offers some advantages over target-based methods, mainly because it allows the discovery of hits and leads in complex biological systems, with intact pathway signaling mechanisms, designed to more closely model distinct disease states. The renewed interest in phenotypic screening may be attributed to several factors, including: increased demand to expand novel drug target space to feed TDD platforms; ongoing unmet therapeutic need in many complex human conditions where target biology is poorly understood; demonstrative success of PDD in the development of novel, first-in-class, small-molecule drugs (Eder et al., 2014; Moffat et al., 2014; Ratni et al., 2018; Swinney and Anthony, 2011) and the convergence of new advances in cell-based screening, target deconvolution, and computational analysis.

However, a potential disadvantage of PDD is the challenge of target deconvolution. Target deconvolution describes the conclusive association of a pharmacologically perturbed

phenotype to modulation of a specific cellular or extracellular molecular target. While target deconvolution is not a pre-requisite for progressing candidate drugs from phenotypic screens into late-stage preclinical or even clinical development (Han et al., 2018; Moffat et al., 2017; Vincent et al., 2020), the preference of many drug discovery groups, funders, and investors is to identify the target. Advances in the field of chemical proteomics and development of new experimental tools has significantly advanced our ability to determine protein targets underlying hits from phenotypic screens (Warchal et al., 2016) (Comess et al., 2018). However, many target deconvolution methods assume that a phenotypic hit exerts its phenotypic outcome through modulation of a single target, yet it is most probable that small molecules interact with multiple cellular and/or extracellular targets or non-protein targets (Lin et al., 2012; Schneidewind et al., 2020). Furthermore, the identification of a single target responsible for drug MOA or indeed a disease phenotype is in many cases based on flawed logic given previous observations of contributing off-target drug activity and the requirement for multi-targeted therapies to address complex multifactorial disease mechanisms (Lu et al., 2012; Rena et al., 2017). Perhaps a more holistic understanding of drug MOA at phenotypic and pathway levels may support more effective progression of phenotypic hits via in-depth mechanistic deconvolution at the integrated pathway network level rather than at the individual target level.

Network pharmacology embraces the older traditional drug discovery ideal that understanding the biological MOA of a drug across multiple layers of biological complexity is more important, and in some instances may substitute for, identification of individual targets. This advancement, in turn, is well placed to shift the paradigm of target deconvolution and drug MOA profiling from a ''one-target, one-drug'' mode to a ''network, multi-target therapeutic'' mode. This approach therefore embraces the complexity of disease and supports the development of rational drug combinations and polypharmacology strategies at earlier phases in the drug discovery process.

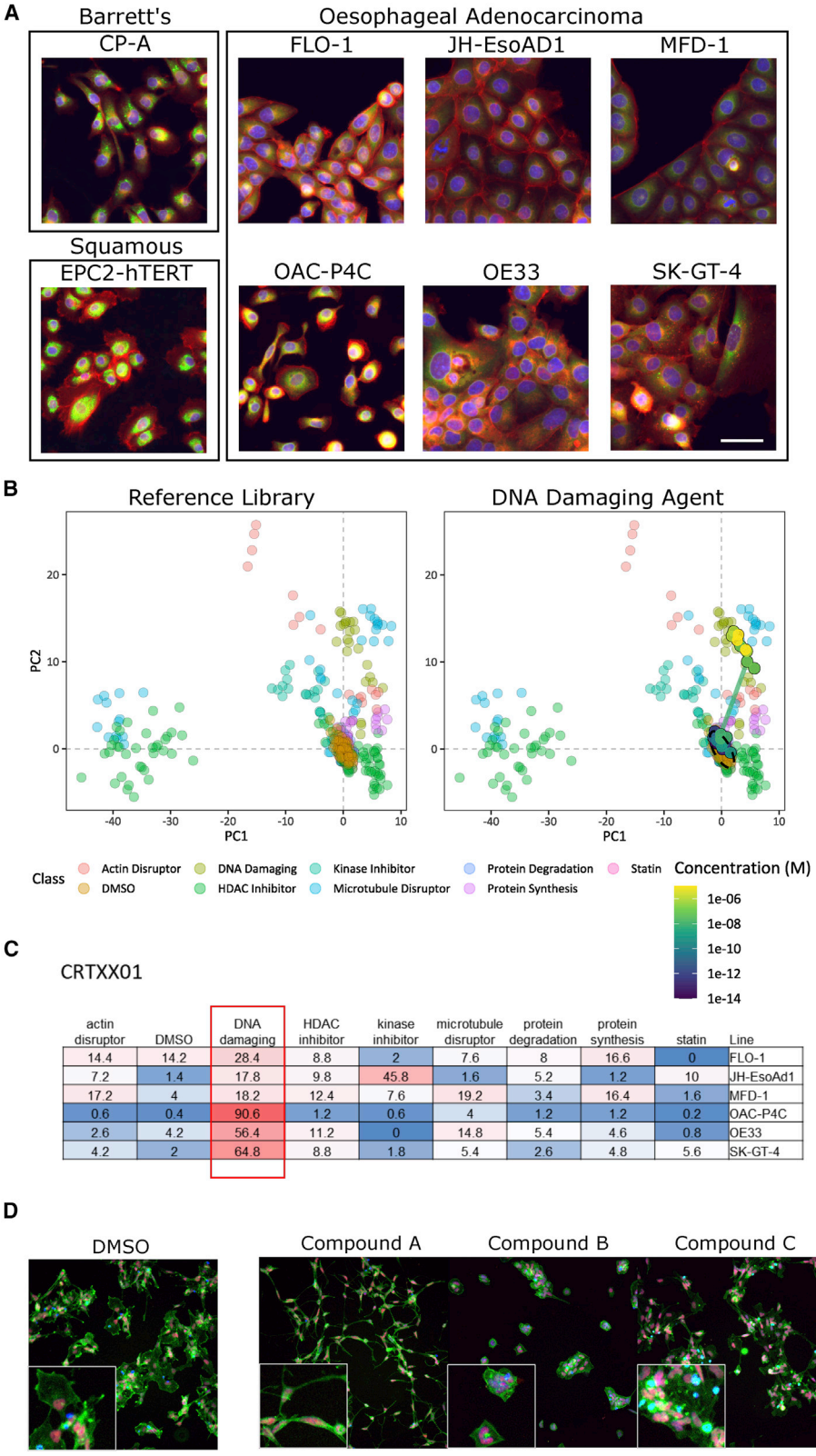
This manuscript intends to provide an extensive review of how recent advances in phenotypic, transcriptomic, and proteomic technologies can integrate with evolving chemical and biological databases and new computational methods to efficiently and comprehensively elucidate the MOA of phenotypic screening hits. Here, we provide several case studies describing the application and deployment of these methods across panels of genetically distinct human cell models, including glioblastoma and esophageal cancer, which represent the heterogeneity of complex diseases of unmet need. We illustrate how empirical phenotypic and pathway-profiling studies integrate with cheminformatics, *in silico* target prediction, and pathway network analysis tools to support the deconvolution and triage of phenotypic hits at the pathway network level. We propose that such pathway network MOA analysis can effectively link complex disease association with drug MOA and uncover pharmacodynamic pathway biomarkers to support more productive PDD and clinical translation.

## Multiparametric high-content profiling

PDD encompasses a large variety of assays and endpoints, from univariate cell line assays, to multivariate whole-organism assays. High-content multivariate profiling assays have been exploited for many applications, including target identification and functional genomic studies, but one major benefit is the ability to elucidate MOA from the primary assay, owing to the depth of information in the data. High-content screening involves capturing biological information relating to the cellular state that is contained in multiple cellular markers, including those associated with morphology. Upon perturbation, cells undergo detectable changes in a subset of profiled features, allowing quantitative data to be extracted from microscopy images and a phenotypic fingerprint to be assigned to a perturbation (Perlman et al., 2004; Young et al., 2008; Caie et al., 2010; Ljosa et al., 2013). These morphological changes are a result of the pathways and/or networks affected. Consequently, high-content profiling is capable of detecting changes that arise from modulation at the pathway/network level making it suitable for studying complex diseases, multi-target drugs and drug combinations.

Because the cell morphological changes associated with a compound treatment are a consequence of its effects on specific cellular pathways, compounds acting on a given pathway elicit more similar morphological responses than those acting on different pathways. This means that the morphological information in images can be used to identify phenotype-altering compounds and elucidate MOA simultaneously. This was first demonstrated in 2004 when Perlman et al. showed that compound-induced morphologies could be grouped by MOA using similarity measures and hierarchical clustering. Since then, numerous research groups have implemented unsupervised clustering methods to group treatments based on phenotypic response (Young et al., 2008; Ljosa et al., 2013; Gustafsdottir et al., 2013). More recently, MOA has been inferred by similarity comparison to reference compound sets using multivariate statistics and a variety of supervised machine learning techniques. One of the earliest studies using reference libraries set out to compare phenotypic profiles and draw mechanistic inferences from several reference compounds using k-nearest neighbor proximity labeling to predict the mechanism of hits (Caie et al., 2010). More recently, this has been achieved through use of trained classifiers and we and others have shown that these are transferable to new cell lines (Dawson et al., 2020; Hughes et al., 2020; Warchal et al., 2019). Finally, artificial intelligence (AI)-inspired deep learning is an area of growing interest in the field of biological image analysis and is being applied to a number of problems, including classification of small-molecule MOA. Several AI methods have been developed to predict compound MOA, including supervised networks trained on segmented or whole images (Godinez et al., 2017; Kraus et al., 2016), self-supervised methods using pseudo-labels (Ando et al., 2017; Godinez et al., 2017), and even unsupervised methods using just image intensity values, making segmentation and feature extraction unnecessary steps (Janssens et al., 2020).

In theory, high-content screening can utilize any image-based assay endpoint. However, to predict mechanism, broad morphological information is required. Cell Painting was developed in 2013 as an unbiased, cell-based, morphological profiling assay capable of capturing subtle changes in morphology (Gustafsdottir et al., 2013). It combines six fluorescent dyes, imaged in five channels, to visualize seven cellular components and organelles; nuclei, F-actin, endoplasmic reticulum, mitochondria,




| actin disruptor | DMSO | DNA damaging | HDAC inhibitor | kinase inhibitor | microtubule disruptor | protein degradation | protein synthesis | statin | Line |
|---|---|---|---|---|---|---|---|---|---|
| 14.4 | 14.2 | 28.4 | 8.8 | 2 | 7.6 | 8 | 16.6 | 0 | FLO-1 |
| 7.2 | 1.4 | 17.8 | 9.8 | 45.8 | 1.6 | 5.2 | 1.2 | 10 | JH-EsoAd1 |
| 17.2 | 4 | 18.2 | 12.4 | 7.6 | 19.2 | 3.4 | 16.4 | 1.6 | MFD-1 |
| 0.6 | 0.4 | 90.6 | 1.2 | 0.6 | 4 | 1.2 | 1.2 | 0.2 | OAC-P4C |
| 2.6 | 4.2 | 56.4 | 11.2 | 0 | 14.8 | 5.4 | 4.6 | 0.8 | OE33 |
| 4.2 | 2 | 64.8 | 8.8 | 1.8 | 5.4 | 2.6 | 4.8 | 5.6 | SK-GT-4 |

**Figure 1. High-content profiling assays in heterogeneous cell panels**

(A) Color combined representative images for a panel of eight esophageal cell lines using the Cell Painting assay: DAPI, blue; TxRED, red; FITC, green. Scale bar, 50 μm.

*(legend continued on next page)*

Golgi, plasma membrane, and nucleoli (Bray et al., 2016; Gustafsdottir et al., 2013). Due to the broad morphological information contained in the images, Cell Painting data can be exploited for compound MOA prediction. A dataset of images and morphological profiles of 30,000 small-molecule treatments using the Cell Painting assay has been published and made publicly available (Bray et al., 2017). While this current dataset is limited to a small-molecule compound library and a single cell line (U2OS cells), integration of this dataset with data from other types of perturbations (e.g., genetic) and inclusion of additional patient-derived cells lines will increase the utility of such high-content datasets in identifying gene-related phenotypes and defining small-molecule MOA, including target identification.

To specifically address the challenges in elucidating novel therapeutic targets and chemical starting points in complex heterogeneous cancers, we describe the application of the high-content Cell Painting assay across well-characterized panels of both patient-derived glioblastoma and esophageal cancer cell lines. Esophageal cancer and glioblastoma multiforme (GBM) represent two cancers of unmet therapeutic need and exemplify how complex disease heterogeneity confounds modern TDD approaches, with molecularly targeted therapies having little impact in clinical studies to date (Bang et al., 2010; Davis, 2016). To address this complexity, a panel of eight esophageal adenocarcinoma (OAC) and tissue-matched non-transformed control cells was used to profile a total of 19,555 small molecules consisting of both target-annotated or off-patent drugs (2,947 compounds) and diverse chemical and lead-like compounds (16,608 compounds) (Figure 1). Here, we demonstrated the utility of combining the Cell Painting assay with machine learning to classify drug MOA across heterogeneous cell lines (Hughes et al., 2020) (Figure 1B). We provide an example of a compound hit (CRTXX01) from the diverse chemical library, which displayed potent and selective activity upon OAC cell morphology and cytotoxicity relative to non-transformed control cells (Figure 1C). The Cell Painting profile for this compound was imputed into a machine learning model created using a random forest classifier to predict MOA based on similarity to a reference library of MOA annotated compounds. As shown, the compound is most highly predicted to be a DNA-damaging agent (Figure 1C). Prediction accuracy varies across the heterogeneous OAC cell panel, indicating that this compound may be a specific type of DNA-damaging agent that exhibits cell line selectivity. In addition to exhibiting a high degree of heterogeneity within and between patients tumors (Lee et al., 2018), GBM hijacks mechanisms of neural development to produce subcompartments of glioblastoma stem cells that are thought to exhibit resistance to radiotherapy and chemotherapies (Bao et al., 2006; Chen et al., 2012; Lathia et al., 2015; Parada et al., 2017). To address such heterogeneity in GBM we have also initiated the development of a multiparametric high-content phenotypic profiling assay to classify different glioma cell phenotypes across a panel of patient-derived glioma stem cell lines, which represent distinct GBM clinical subtypes. Initial studies using the Prestwick Chemical library of FDA/EMA-approved drugs before further assay optimization and screening against larger chemical libraries has revealed several drug classes that perturb glioma stem cell phenotypes (Figure 1D). It is anticipated that these high-content phenotypic profiling studies will reinvigorate drug discovery in these complex heterogeneous cancers by identifying multiple chemical starting points, in addition to drug-repurposing and drug-combination opportunities.

High-content profiling is evolving and, as computational power increases, so does throughput. Expansion of high-content studies across heterogeneous panels of cell lines (Hughes et al., 2020; Warchal et al., 2019; Willis et al., 2020) has allowed associations to be made linking phenotypes with genotypes utilizing basal genomic, transcriptomic, and proteomic data. This in turn, further supports compound MOA studies, biomarker discovery, and target identification, as part of a larger personalized medicine strategy. Very recently, a new initiative was announced to create an academic-industry cell imaging consortium; The Joint Undertaking in Morphological Profiling with Cell Painting (JUMP-CP), which aims to create the largest Cell Painting dataset of cells imaged under more than 140,000 different conditions, including both small-molecule and genetic perturbations, acting as a morphological atlas to match diseases, drugs, and genes. This resource will provide a huge reference dataset to advance the field of cell-based phenotypic screening and enhance applications in predicting MOA, cellular toxicities, and drug-repurposing opportunities. However, the problem remains that compounds and genes are often distilled down to single compound—target interactions to link phenotypes with genotypes and such perturbations are not viewed within the context of integrated pathway signaling networks leading to lack of insight for complex diseases and compounds with multiple targets.

Despite the wide applications of high-content profiling and recent advances in high-content instruments and parallel computer processing that have overcome many of the throughput and capacity issues, there remain several limitations, predominantly with assay complexity (Table 1). High-content profiling as yet is not suitable for complex 3D or non-adherent cell cultures at sufficient scale and robustness, limiting its use in assays that may better recapitulate the complexity of many diseases. High-content profiling has previously been used to infer target prediction based on the phenotypic similarity of a compound with phenotypic clusters of target-annotated compounds enriched in a specific target class (Reisen et al., 2015). However, recent studies have shown that compounds with different annotated cellular targets can induce a shared phenotypic response (e.g., modulation of cell-cycle S to G2 phase transition) (Schneidewind et al., 2020). Thus, such phenotypic clusters are based on a shared mode of action and not on a shared target. There are also barriers to adoption across many laboratories owing

(B) The first two components of principal-component analysis (PCA) for a reference library of compound treatments with and without overlay of phenotypic dose-response to a known DNA-damaging agent (points highlighted in bold on the right hand plot). Reference library points are colored by mechanistic class and DNA-damaging agent treatments are colored by concentration. This example is obtained from the FLO-1 OAC cell line.
(C) Application of a random forest classifier to a novel uncharacterized hit compound (CRTXX01) from the Cell Painting indicating high prediction accuracy as a DNA-damaging MOA.
(D) Representative images of patient-derived glioma cells treated with compounds, color combined image of cytoskeletal (green) and nuclear markers (blue and red). Negative control, DMSO (dimethyl sulfoxide).

**Table 1. Advantages and disadvantages of different biological and chemoinformatic profiling methods for elucidating MOA and target prediction of phenotypic screening hits**

| Method | Advantages | Disadvantages |
|---|---|---|
| High-content profiling | high-throughput and suitable for incorporating multiple time points very cost-effective at scale broad mechanistic coverage, including non-protein targets | compounds within the same MOA cluster may have distinct targets requires sophisticated image analysis and computational resources not yet optimized for more complex 3D assays |
| L1000 | medium high throughput reasonably cost-effective at scale broad mechanistic coverage | limited coverage of the transcriptome, dependence on computational inference of whole transcriptome may be prone to false negatives cost may prevent profiling across multiple time points no direct information on post-translational MOA |
| DRUG-seq | high throughput reasonably cost-effective at scale broad mechanistic coverage of the majority of the transcriptome | cost may prevent profiling across multiple time points no direct information on post-translational MOA |
| NanoString | high-precision quantification of transcriptomic data from small samples suitable for complex and heterogeneous *in vitro* and *in vivo* samples (GeoMX spatial profiling) can multiplex transcriptomic and candidate protein markers | low throughput coverage limited to known canonical and noncanonical signaling pathways |
| RPPA | high precision quantification of post-translational pathway markers from small samples medium throughput and costs suitable for multiple dose-response and time-series studies | mechanistic coverage limited by the availability of high-quality and validated monospecific antibodies |
| Chemin formatics (structural similarity ligand-based target prediction) | high throughput highly cost-effective rapid | multiple target predictions require validating contributing to potential bottlenecks not optimal for truly novel chemical structures exploring novel biological target space not suitable for non-protein targets |

to the complexity of implementing image analysis pipelines, but these are being lowered as standard practices are introduced and generic software for multiparametric high-content profiling are launched. While broad MOA prediction at scale is now possible and enabled by high-content profiling, specifically linking phenotype to genotype for personalized medicine approaches still presents a bottleneck in the discovery process and requires integration with other profiling platforms and technologies.

Recent large-scale initiatives in cancer cell line panel screening have produced several resources to help connect phenotype to genotype through the identification of genetic dependencies and small-molecule sensitivities. The Cancer Dependency Map Consortium is an extension of the Broad Institutes DepMap project, an initiative to systematically catalog and identify biomarkers of genetic vulnerabilities and drug sensitivities in hundreds of cancer cell lines. DepMap combines the Cancer Cell Line Encyclopedia with Project Achilles; a catalog of genetic vulnerabilities using genome-wide RNAi and CRISPR loss of function to identify genes essential for cancers, and a comprehensive cell line drug sensitivity resource (Tsherniak et al., 2017). Another project has led to the Genomics of Drug Sensitivity in Cancer, a resource for therapeutic biomarker discovery in cancer cells, combining large-scale drug screening and genomic analyses (Yang et al., 2013). These databases have been used to aid PDD and elucidate MOA, via correlation of basal gene expression with chemical sensitivity. For example, the sensitivity of a small-molecule compound ML239, originally identified from a phenotypic screen to exhibit selectivity for breast cancer stem cells, was profiled across a large cell panel and correlation with basal gene expression and follow-up studies indicate that fatty acid desaturase 2 (FADS2) activity is required for ML239 sensitivity (Rees et al., 2016). Additional examples and resources for correlating gene expression signature to predict drug MOA have been described in further detail by Keenan et al. (2019). However, resources linking gene expression signatures to cell sensitivity are typically limited to simple cell viability and survival phenotypic endpoints and thus are unsuited to disease areas and high-content datasets studying more complex phenotypic cellular readouts. The evolution of multiparametric high-content profiling and Cell Painting assays across larger panels of genetically distinct cell lines promise to advance significantly this field of functional genomics.

## Transcriptomic and pathway screening technologies

Recent technical advances in transcriptomic and post-translational pathway-profiling approaches have substantially improved sensitivity and throughput and thus have become applicable to drug MOA studies across complex human cell-based assays, *in vivo* models, and clinical samples. The increase in high-throughput technologies available for generating large-scale genomic, transcriptomic, epigenetic, and proteomic measurements has also accelerated biomarker discovery and development of more sophisticated multiparameter biomarker signatures (Ghosh and Poisson, 2009). The emergence of new technologies supporting the identification of clinical biomarkers encompassing multiparametric gene and pathway signatures provide new opportunities to translate phenotypic hits that have been deconvoluted at the pathway network level. We briefly

outline four emerging and complementary technologies that provide sufficient throughput, sensitivity, and robustness to deconvolute phenotypic hits at the transcriptomic and post-translational pathway level; L1000, DRUG-seq, NanoString, and Reverse Phase Protein Array (RPPA).

The Connectivity Map project and public repository of transcription profiles was developed to enable a systematic comparative bioinformatics analysis of gene expression profiles (Lamb et al., 2006). The Connectivity Map combines a catalog of gene expression profiles from large panels of pharmacologically perturbed samples with computational and statistical methods to support similarity profiling of gene expression patterns to infer compound MOA (Lamb et al., 2006; Subramanian et al., 2017). Published examples of this application include, identifying that the activity of two novel androgen receptor signaling inhibitors, celastrol and gedunin, is mediated through upstream HSP90 inhibition based on the similarity of these compound-induced gene signatures to those of known HSP90 inhibitors (Hieronymus et al., 2006). The previously uncharacterized novel agent, epoxy anthraquinone derivative, identified in a phenotypic screen against neuroblastoma cells, was found to exhibit strong similarity in gene expression signature to those induced by a class of DNA targeted drugs, such as topoisomerase inhibitors, DNA intercalators, and DNA alkylation agents (Gheeya et al., 2010). The previous unclassified small-molecule, b-AP15, induced a gene expression profile similar to several well-characterized proteasome inhibitors culminating in the hypothesis of inhibition of proteasome deubiquitinating activity as a potential new anticancer therapy (D'Arcy et al., 2011). These and other applications of Connectivity Map are reviewed in further depth by Qu and Rajpal (2012). The Connectivity Map concept was originally based on data provided from gene expression microarrays. However, new advances in high-throughput (e.g., L1000 and DRUG-seq), high-precision, and spatial transcriptomics (e.g., NanoString) promise to extend the application of the Connectivity Map concept across larger compound libraries and more complex biological samples.

### L1000

Technical advances in gene expression profiling include the development of higher-throughput and low-cost gene expression methods, such as the L1000 platform. L1000 expression profiling is based upon the quantification of several landmark transcripts in 384-well plate format and a computational model to predict expression across the wider genome (Liu et al., 2015). The L1000 technology underpins the National Institute of Health Library of Integrated Cellular Signatures program to generate a comprehensive library of perturbed gene expression profiles across multiple cells, including small-molecule pharmacological perturbations supporting drug MOA profiling at scale (Liu et al., 2015). From 978 landmark gene expression profiles, modulation of the entire transcriptome is modeled from computational processing of thousands of gene expression datasets from the Gene Expression Omnibus (Barrett et al., 2007). An analysis of L1000 compound profiling data by De Wolf et al. (2018) revealed that compound characterization based on three cell lines at two time points resulted in the identification of more pharmacologically perturbed gene expression changes, than using six cell lines at a single time point. This result highlights how temporal gene expression profiling may be important to fully elucidate the MOA of small-molecule perturbations at the transcriptomic level. In a study profiling 31,000 compounds in the MCF7 breast cancer cell line using the L1000 platform, the authors identified 22 new compounds with predicted activity against NR3C1 (nuclear receptor subfamily 3, group C, member 1) and HSP90 (heat shock protein) by machine learning, which were subsequently confirmed in cell-based target confirmation assays (De Wolf et al., 2018).

### DRUG-Seq

DRUG-seq represents a massively parallelized, low-cost NGS-based method to profile whole transcriptome changes in cells exposed to chemical and genetic perturbations in high-throughput 384- and 1,536-well assay formats (Ye et al., 2018). Ye et al., used the DRUG-seq workflow to profile a collection of 433 compounds with predominantly known targets. t-Distributed stochastic neighbor embedding was used to cluster compounds with similar transcription profiles, and the MOA of compounds with unknown targets were then inferred from their neighbors in the cluster. For example, brusatol, a natural product-derived small molecule with potent tumor-suppressing effects closely clusters with several compounds, including homoharringtonine and cycloheximide, which target components of the translation machinery (EIF4E, EIF2AK1, and RPL6) suggesting that brusatol may also share a similar target mechanism (Ye et al., 2018). In addition to significant cost reduction, another advantage of the DRUG-seq platform compared with L1000 is the ability to directly measure >10,000 genes without computational inference. Further recent advances in single-cell genomics and phenotypic screening in pooled cell cultures provides new opportunities to resolve drug MOA at the single-cell transcriptomic level (Klein et al., 2015). In a recent development, multiplexed interrogation of gene expression through single-cell RNA sequencing combines the ability to pool hundreds of cancer cell lines with single-cell RNA sequencing to simultaneously monitor the response of the cells to pharmacological perturbation and resolve the identity of each cell based on single-nucleotide polymorphisms (McFarland et al., 2020). Such advances in NGS technologies promise to revolutionize the study of drug MOA across large panels of distinct cell lines at the single-cell level.

### NanoString

The NanoString nCounter gene expression system, utilizes innovative technology to capture and quantify individual mRNA transcripts (Geiss et al., 2008). Advantages over existing platforms include direct measurement of mRNA expression levels in solubilized or intact (GeoMx Digital Spatial Profiling) biological samples without enzymatic reactions, high sensitivity coupled with high multiplex capability, and precise digital readouts. The NanoString technology utilizes a multiplexed probe library comprised of two sequence-specific probes, a capture probe and a reporter probe with a color-coded tag for each gene of interest. The reporter probes are mixed together with total RNA in a single hybridization reaction. Comparative studies have demonstrated that the NanoString nCounter gene expression platform is more sensitive than microarrays and similar in sensitivity and accuracy to real-time PCR (Geiss et al., 2008). However, the NanoString system is more scalable and amenable to small-sample or miniaturized assay formats than real-time PCR or microarrays. Requiring just nanograms of RNA, NanosString is

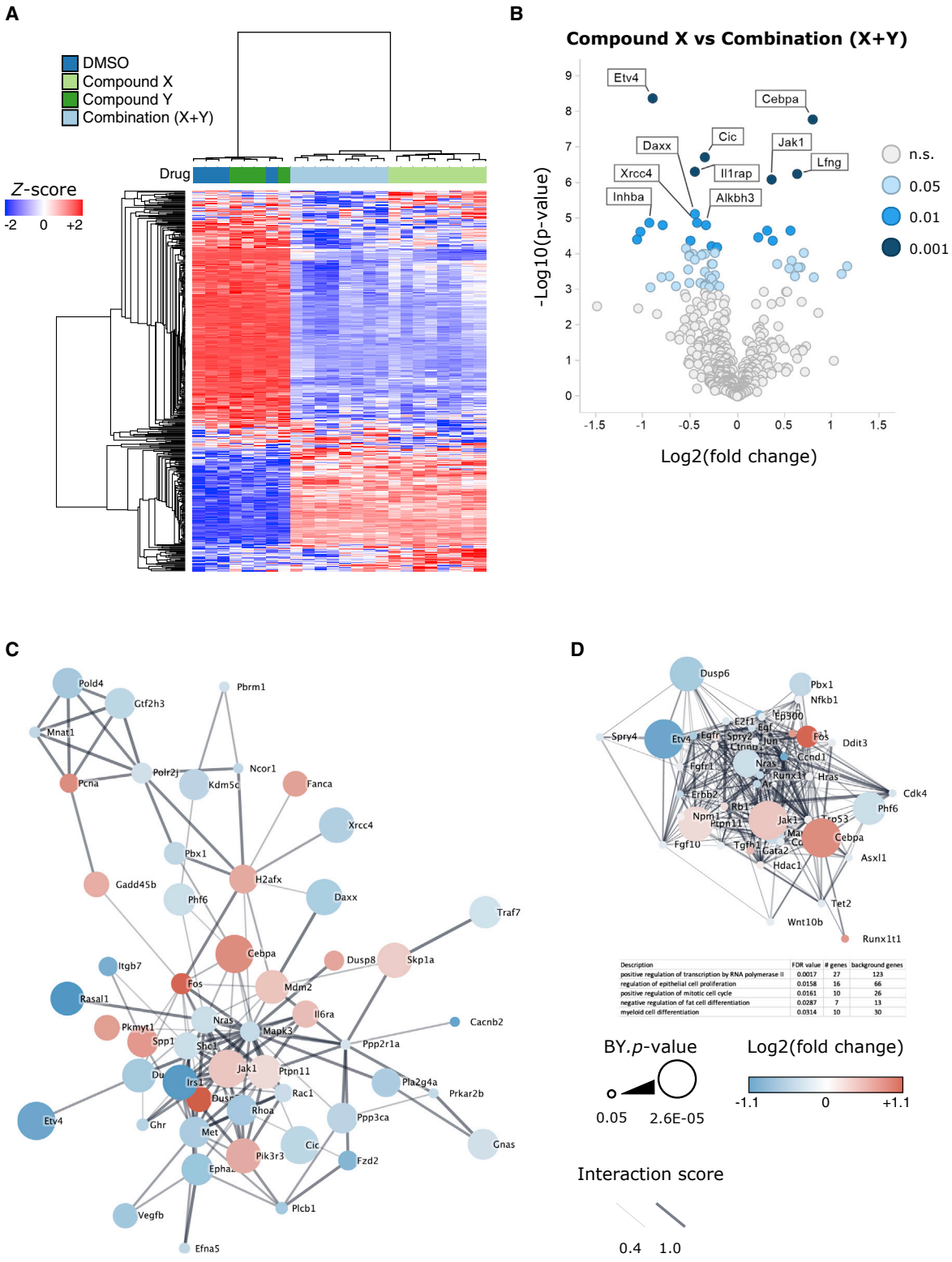



**Figure 2. NanoString analysis of combination compound treatments in squamous cell carcinoma cells**

(A) Heatmap of mRNA expression (y axis) using the NanoString PanCancer Pathways Panel in squamous cell carcinoma cells treated with (x axis) compound X, Y, or a combination of the two compounds (X + Y) previously identified by phenotypic screening. Hierarchical clustering is average linkage using one minus Pearson correlation. Values are log2 transformed and displayed as a *Z* score.

*(legend continued on next page)*

ideally suited to drug MOA studies in complex preclinical or clinical tissue samples with limited amounts of RNA, such as tissue biopsies, micro-dissected or laser-captured samples, and cells sorted by flow cytometry.

Thus, NanoString technology can fill an immediate niche in high-precision expression analysis of hundreds of genes across complex biological samples, including those from dose-response and time-series studies of phenotypic hits to elucidate downstream MOA at the transcriptomic level. We demonstrate the application of NanoString and integrated pathway network analysis to profile the MOA of phenotypic hits previously identified by phenotypic screening in squamous carcinoma cells (Figure 2). However, a limitation of the NanoString platform for deconvoluting phenotypic hits at the pathway level is sample throughput; the current NanoString nCounter system is designed to run hybridization reactions across only 12 samples. Increased throughput at reasonable cost would be beneficial for deconvoluting and triaging increased numbers of phenotypic hits.

### *RPPA*

While transcriptomic profiling has proven valuable in the identification of new therapeutic targets and understanding of drug MOA at the transcriptomic level, the majority of small-molecule and biologic drugs target proteins. Mass spectrometry represents the standard technology for global analysis of the proteome but is limited by throughput, cost, and sensitivity with regard to detection of low-abundance proteins and post-translational modifications. In contrast, antibody-based proteomic technologies provide the combined advantages of high sensitivity, high-throughput, and precise quantification at a relatively low cost, albeit with limited coverage across the proteome. RPPA is an antibody-based microarray methodology allowing the sensitive and precise quantitation of the abundance of total protein and post-translational modifications in medium throughput, making it well placed to support MOA profiling of phenotypic hits across dose-response and time-series studies (Akbani et al., 2014; Charboneau et al., 2002; Lee et al., 2012; van Oostrum et al., 2009). RPPA has been successfully implemented in a variety of formats by multiple laboratories and a wide range of microarray slides are commercially available for the immobilization of protein samples before the addition of labeled antibodies for protein detection. The number of samples and protein analytes that can be detected on a single microarray slide is therefore flexible and dictated by the number and size of each sub-array coupled onto each RPPA slide, which physically separates the antibody-detection reagents. RPPA formats range from 1 to 64 subarrays/analyte, which can be accommodated on a single slide; additional analytes and sample numbers can be studied by printing protein samples across additional slides (Akbani et al., 2014). A typical RPPA study would incorporate several hundred protein analytes evaluated across hundreds of protein samples. Dependent on the specific RPPA technical set up the analytical sensitivity of RPPA and lower limits of detection have been reported in the fg/mL range, with linearity in the sub-pg/mL range (Pawlak et al., 2002; Rapkiewicz et al., 2007).

RPPA has been utilized to help elucidate the MOA of several natural product compounds which first displayed activity in phenotypic assays (Rummun et al., 2019). For example, RPPA analysis of a panel of ovarian cancer cell lines displaying sensitivity to Corilagin, a member of the ellagitannin class of hydrolyzable tannins demonstrated suppression of canonical Smad and noncanonical ERK/AKT pathways, which correlated with inhibition of transforming growth factor β (TGF-β) secretion and TGF-β pathway activation (Jia et al., 2013). In a novel chemical-genetic phenotypic screen to identify compounds that synergize with the catalytic activity of focal adhesion kinase (FAK) we identified synergy between the histone deacetylase inhibitor, vorinostat, and the FAK inhibitor, VS4178, across several cancer cell lines, including OAC lines (Dawson et al., 2020). To elucidate the MOA of the synergy between vorinostat and VS4718, protein extracts were prepared from two OAC cell lines (FLO1 and SKGT4) after treatment with each compound alone and in combination (Figure 3). This study revealed several pathways modulated by each compound individually and cooperative downregulation of pathway markers, such as EphA2, and autophosphorylation of FAK397 upon combination treatments relative to monotherapy (Figure 3). While many published examples describing the application of RPPA in drug MOA studies have been performed on compounds derived from target-directed drug discovery programs, the success of this approach in revealing unanticipated drug MOA indicates that this will also be a useful method for novel compounds derived from phenotypic screens too. Furthermore, these studies demonstrate the utility of RPPA to complement transcriptomic profiling by identifying downstream mechanisms at the post-translational pathway level and potential pharmacodynamic biomarkers for phenotypic hits.

### *In silico* target deconvolution of phenotypic hits

A major limitation in small-molecule phenotypic screening is the lack of suitably selective compounds annotated for activity against specific protein targets. For example, given the approximately 1.3 billion compounds available for virtual screening (Gorgulla et al., 2020), from which approximately 16 million are physically available from commercial vendors (Volochnyuk et al., 2019), less than a million compounds have been reported active in a target-specific assay according to public databases, such as ChEMBL or other annotated collections indexed by ZINC (Irwin et al., 2018). Several initiatives aim to provide high-quality chemical probes for target validation, such as the Structural Genomics Consortium (https://www.thesgc.org), the Chemical Probes Portal (http://www.chemicalprobes.org/), Target 2035 (Carter et al., 2019), and a program led by Novartis representing the largest disclosed collection to date of 4,185 compounds with their primary annotated gene targets (https://github.com/Novartis/MoaBox) (Canham et al., 2020). However, the development of highly selective probe compounds and experimental validation of protein targets is a slow and expensive process.

(B) Volcano plot showing differential expression analysis. Top 10 most significant genes are labeled. Colors denote p value cutoffs.
(C) STRING analysis of significantly altered genes. Benjmini-Yekutieli (BY) adjusted p-value.
(D) STRING analysis focusing on the top 2 most significantly altered genes (Etv4 and Cebpa). Pathway enrichment was performed using STRING. False Discovery Rate (FDR) adjusted p-value.

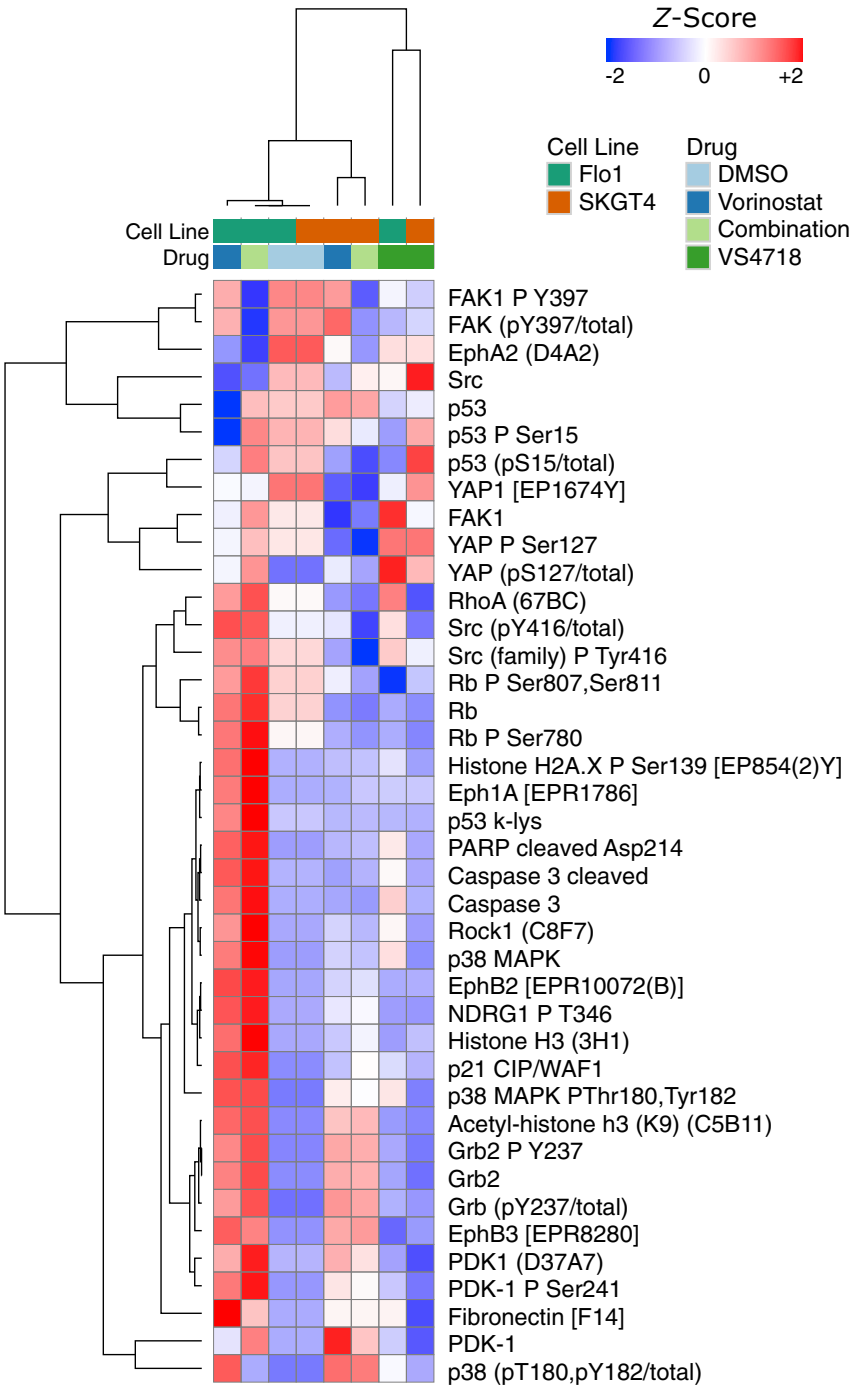



**Figure 3. Application of RPPA to elucidate the mechanism-of-action of phenotypic activity**

Lysates from Flo1 and SKGT4 cell lines were subjected to RPPA analysis using 120 antibodies targeting canonical cancer pathways and post-translational modifications. Selected relative fluorescence intensity for each antibody were hierarchically clustered and displayed as a heatmap.

In contrast to experimental target deconvolution methods, recently developed computational methods, including ligand- and structure-based prediction approaches, allow rapid prediction of targets for large numbers of compounds enabling the triage of phenotypic hits with only desired biological target hypotheses (Schenone et al., 2013). Such ligand-based prediction methods and database resources include: Similarity Ensemble Approach (SEA) (Keiser et al., 2007), SwissTargetPrediction (Gfeller et al., 2014), PASS (Poroikov et al., 2007), SOM-based prediction of drug equivalence relationships (SPiDER) (Reker et al., 2014), multiple-category Bayesian models (Nidhi et al., 2006), support vector machines (Wale and Karypis, 2009), and DePick (Liu et al., 2016). Several of these target prediction methods have been applied to hits from phenotypic screens for anti-bacterial (Martinez-Jimenez et al., 2013), antimalarial (Plouffe et al., 2008; Spitzmuller and Mestres, 2013), and anticancer activity (Flachner et al., 2012; Liggi et al., 2014; Lo et al., 2015) to propose links between pharmacological targets and phenotypes. For example, Liu et al. (2016) have developed a computational target deconvolution tool tailored toward phenotypic screening hits called DePick. This tool identifies targets specifically linked to chemical phenotypic screens of mammalian organisms. DePick first identifies the compounds with specific activity in a screening assay relative to control/counter screens and predicts their targets with the HitPick *in silico* target prediction method (Liu et al., 2013). The authors applied DePick to eight mammalian chemical phenotypic screens stored in the ChemBank repository, where control/counter screening data was available to determine the specificity of the active compounds for each phenotype. They identified 59 target-phenotype relationships (Liu et al., 2016). To evaluate the reliability of these drug target-phenotype predictions, the authors performed a manual curation of the scientific literature to source experimentally validated evidence on the relationship of drug targets and phenotypes. They found experimental evidence that directly related a chemical or genetic modulation with the phenotype for 27% of the DePick predictions, and indirect links mediated via an intermediate protein or metabolite common to both the drug target and the phenotype for 34% of the predictions. For the remaining predictions, no clear links between drug target and phenotype could be demonstrated (Liu et al., 2016). However, literature searches suggesting novel connections between drug targets and phenotypes

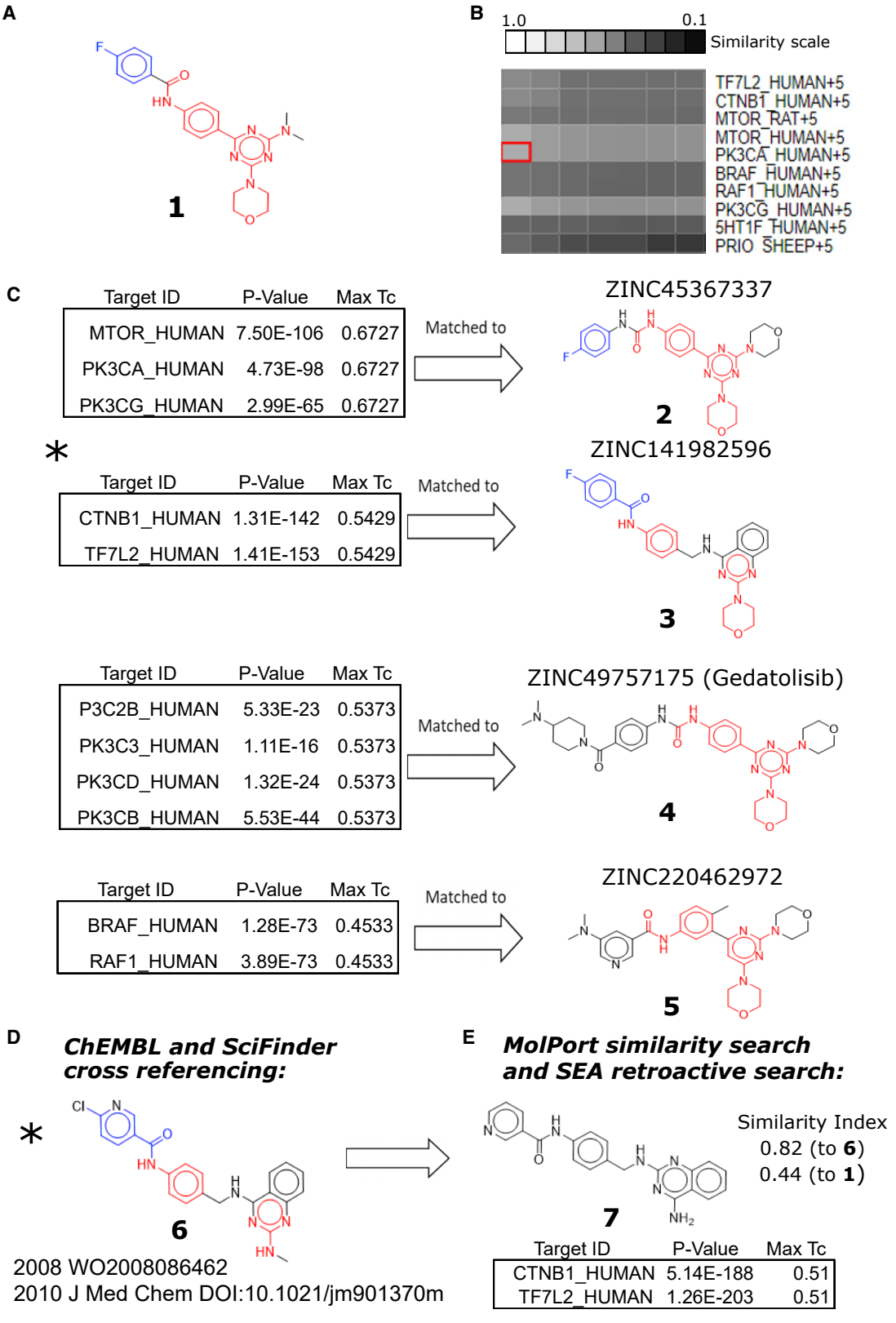



**Figure 4. Application of SEA and other chemoresources to elucidate target activity**

(A) A compound with antiproliferative properties against patient-derived glioma stem cells for query in SEA. Common structural features to compounds in (C) indicated in red and blue.

*(legend continued on next page)*

warrant further experimental investigation. In another study, Kremer et al. (2017) used two separate chemoinformatic software tools, SPiDER (Reker et al., 2014) and TIGER (Schneider and Schneider, 2017), to apply structural similarity-based target prediction methods in parallel with cell-based bioactivity assays to accelerate target identification of a novel hedgehog (Hh) pathway inhibitor. In this study, the investigators used purmorphamine, a smoothened (SMO) agonist, to activate Hh signaling in a high-throughput osteoblast differentiation assay. After screening of a 336,639 compound library the most active compounds were validated in an orthogonal glioma-associated oncogene homolog (GLI)-dependent Hh pathway reporter assay. The SPiDER analysis of the top hit suggested the transmembrane protein SMO as a potential target with the highest confidence followed by members of the ligase and tyrosine kinase target families. To provide a second opinion, they applied TIGER software, which relies on pharmacophore patterns for target inference. Of the three targets suggested by SPiDER, only SMO was reported by the TIGER tool (Kremer et al., 2017). SMO is known to activate the Hh pathway via GLI-dependent transcription, thus in this case the *in silico* target prediction is corroborated by two separate structural similarity target prediction tools and both the Hh phenotypic and cell reporter assays.

As an example of how the SEA (http://sea.bkslab.org) and other chemoresources can also be used to inform target prediction of phenotypic hits, we demonstrate how a compound which indicated antiproliferative properties against glioma stem cells was assessed (compound **1**, Figure 4). A SMILES string was generated using Chemdraw Prime (PerkinElmer) and pasted into the SEA query box. The SEA output includes an interactive similarity matrix that is useful for comparing chemical structures directly. Lighter-shaded squares indicate that the query molecule has a closer similarity (white = max 1.0 for exact match) to a known molecule and its proposed/published target, listed on the right (Figure 4B). The most significant putative target IDs are listed by SEA p values and by maximum Tanimoto coefficient (Max Tc), with the most highly significant thresholds giving $p \leq 10^{-40}$ (or negative log SEA $\geq 40$) and Max Tc $\geq 0.40$ (Figure 4C) (Irwin et al., 2018). The SEA output in our example suggests that compound **1** potentially targets several signal transduction pathways that are particularly relevant to cancer, specifically phosphatidylinositol 3-kinase (PI3K)/AKT/mTOR (Figure 4C, compounds **2** and **4**), Wnt (compound **3**), and MAPK signaling (compound **5**). In fact, compound **1** had the highest Max Tc (0.67) compared with compound **2**, a mTOR/PI3K inhibitor that was cross-referenced via ZINC and ChEMBL databases to compound **4** (gedatolisib) (Table 2). From the literature, gedatolisib is a more potent inhibitor of PI3K$\alpha$ and is clearly an analog derived from compound **2**. The ZINC database indicates that compound **2** is not commercially available, whereas it provides a list of commercial sources for gedatolisib, affording a convenient route to validation of mTOR/PI3K inhibitors (of a similar structural class to compound **1**) on our glioma stem cell lines, alongside other well-known (and more selective) mTOR/PI3K inhibitors. This chemical-led approach would complement parallel target validation of mTOR or PIK3CA via the use of short interfering RNA or CRISPR approaches. Of further interest, the SEA output attributes the highest p value to compound **1**, targeting $\beta$-catenin/TCF transcriptional signaling, i.e., the Wnt pathway (Figure 4C, compound **3**). Unfortunately, compound **3** is not commercially available and cross-referencing with ChEMBL and SciFinder indicates that the compound was part of a patent submitted in 2008 (WO2008086462). The lead compound from this series (compound **6**) is also not commercially available, but use of online resources, such as MolPort, found a compound chemically similar (compound **7**, similarity index 0.82 to compound **6**). The accumulated Tc coefficient from compound **1** to **7** is 0.44 and a retroactive SEA search of compound **7** also indicated highly probable targeting of Wnt signaling, justifying the potential purchase of this compound for further validation (Figure 4). In summary, the increase of commercially available chemical structures and compound/drug target information in the public domain is accelerating the creation of efficient computational methods for the prediction of targets (Irwin et al., 2018; Keiser et al., 2007; Liu et al., 2013, 2016; Nidhi et al., 2006). These resources facilitate the development of *in silico* drug target deconvolution approaches to triage hits from small-molecule phenotypic screens, identify common targets and pathways associated with discrete phenotypes and provide target hypotheses for subsequent network pharmacology studies.

While these approaches provide a rapid and inexpensive means to identify target predictions, they come with a number of caveats and limitations (Table 1). One limitation of cheminformatic-based target prediction is the restriction to structure and sequence similarity among protein targets. Such methods will not detect new emerging target classes, such as lipids, DNA, RNA, or essential metals involved in a broad range of cellular processes (Schneidewind et al., 2020). Another key limitation of *in-silico*-based target prediction is that these are simply target predictions requiring validation and thus when applied in isolation they simply shift the bottleneck of target confirmation further downstream. As no one single target deconvolution method is conclusive it is key to apply multiple methods and software tools to corroborate target hypotheses, such as described by Kremer et al. (2017), to aid prioritization of the most compelling target hypotheses for further validation.

## Convergence of phenotypic profiling and network pharmacology

Network pharmacology remains an evolving drug discovery discipline characterized by the integration of computational

(B) A similarity matrix comparing chemical structures versus protein targets from the SEA analysis.
(C) Putative target IDs listed by SEA p values and by maximum Tanimoto coefficient (Max Tc), with the most highly significant thresholds giving p values $\leq 10^{-40}$ (or negative log SEA $\geq$40) and Max Tc $\geq$ 0.40.
(D) Cross-referencing with ChEMBL and SciFinder indicates the similar compound **3** identified by SEA was part of a patent submitted in 2008 (WO2008086462), which included lead compound **6**.
(E) MolPort was used to find a commercially available compound chemically similar to compound **6** (compound **7**, similarity index 0.82).

**Table 2. Similarity Ensemble Approach target predictions for phenotypic hits—ZINC database cross-reference**

| ZINC ID | Ortholog_name | pKi (affinity) | Pubmed_id | Chembl_id | Journal | Year |
|---|---|---|---|---|---|---|
| ZINC000045367337 | MTOR_HUMAN | 8.41 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| (Compound **2**) | PK3CA_HUMAN | 8.19 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| | PK3CG_HUMAN | 7.28 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| | | | | | | |
| ZINC000049757175 | PK3CG_HUMAN | 8.27 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| Gedatolisib | PK3CD_HUMAN | 8.22 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| (Compound **4**) | PK3CB_HUMAN | 8.22 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| | PK3CA_HUMAN | 9.4 | 21763134 | CHEMBL1806467 | Bioorg. Med. Chem. Lett. | 2011 |
| | PK3CA_HUMAN | 9.4 | | CHEMBL3217752 | MedChemComm | 2012 |
| | MTOR_HUMAN | 9.4 | 21763134 | CHEMBL1806467 | Bioorg. Med. Chem. Lett. | 2011 |
| | PK3CA_HUMAN | 9.4 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| | MTOR_HUMAN | 8.8 | 20166697 | CHEMBL1157610 | J. Med. Chem. | 2010 |
| | MTOR_HUMAN | 8.8 | | CHEMBL3217752 | MedChemComm | 2012 |
| | PK3CG_HUMAN | 8.27 | | CHEMBL3217752 | MedChemComm | 2012 |
| | PK3CG_HUMAN | 7.96 | 21763134 | CHEMBL1806467 | Bioorg. Med. Chem. Lett. | 2011 |

Similarity Ensemble Approach (SEA) output links to ZINC/ChEMBL databases for further compound information, including potency, commercial availability, and literature sources. The SEA predictions for compound **1**, leading to compounds **2** and **4**, likely arise from the same data source (J. Med. Chem., 2010) and the ZINC database for compound **4** (gedatolisib) provides further literature evidence of this molecule as a PI3Kα/mTOR inhibitor.

tools with genomic and proteomic profiling technologies, drug target databases, and graph theory mathematics to model pairwise relationships between proteins and compounds within intact signaling networks (Hopkins, 2008) (Pavlopoulos et al., 2011). In the last 20 years, a variety of related databases and software tools have been developed and evolved to support network pharmacology research. Commonly used databases for network pharmacology studies include Drugbank (Wishart et al., 2006), STITCH (Kuhn et al., 2014), ChEMBL (Gaulton et al., 2012), PubChem (Wang et al., 2009), target interaction databases STRING (Szklarczyk et al., 2015), HPRD (Peri et al., 2003), MINT (Zanzoni et al., 2002), IntAct (Kerrien et al., 2012), Reactome (Matthews et al., 2009), and HAPPI (Chen et al., 2009), and gene-disease association databases OMIM (Hamosh et al., 2002) and GAD (Becker et al., 2004). Such databases integrate closely with open-source and commercial software, such as Cytoscape (Shannon et al., 2003), which is suitable for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations and gene expression profiles (Figure 2).

To enable a system-wide understanding of cellular function and small-molecule MOA at the pathway level requires knowledge of all functional interactions between expressed proteins and small-molecule compounds or drugs. The STRING database aims to collect and integrate pathway networks by consolidating known and predicted protein-protein association data for a large number of organisms. The associations in STRING include direct (physical) interactions, as well as indirect (functional) interactions. In addition to collecting and imputing experimental data on protein-protein interactions, and importing known pathways and protein complexes from curated databases, interaction predictions are derived from the following sources: (1) systematic co-expression analysis, (2) publicly accessible high-throughput lab experiments, (3) detection of shared selective signals across genomes, (4) automated text-mining of the scientific literature, and (5) computational transfer of interaction knowledge between organisms based on gene orthology (Szklarczyk et al., 2017). The latest version of STRING (11.0) connects over 24.6 million protein annotations across 5,090 organisms incorporating >2,000 protein-protein interactions (Szklarczyk et al., 2019) (https://string-db.org/). Knowledge of interactions between proteins and small molecules, within the context of integrated pathway signaling networks operational in live cells, is also essential for understanding the impact of small-molecule therapeutic interventions upon cellular phenotypes and tissue functions. However, information on such interactions is widely dispersed across multiple data repositories and the published literature. To facilitate access to and interpretations of such data, STITCH ("search tool for interactions of chemicals") integrates information about interactions from metabolic pathways, crystal structures, binding experiments, and drug-target relationships (http://stitch.embl.de/).

By using SEA and STITCH-STRING databases in tandem we better inform the MOA of phenotypic hits from their known and hypothesized protein targets and expand to their "first neighbors" represented by direct protein-protein interactions, and ultimately predict upstream and downstream regulators or "influencers." This strategy allows both the expansion of the drug target search space within the context of protein networks and identifies existing compounds or drugs to validate target hypothesis. Elucidated targets, including influencers, can subsequently be queried against the Human Protein Atlas (Uhlen et al., 2015) and the Open Targets databases (Carvalho-Silva et al., 2019; Koscielny et al., 2017) to determine tissue distribution and disease association. Applications of pathway network tools for phenotypic screening

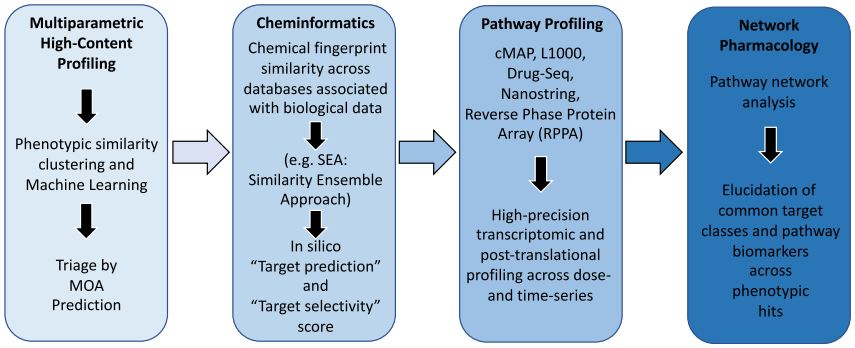



**Figure 5. Pathway to rapid mechanistic deconvolution of phenotypic hits**

Multiparametric high-content profiling enables initial MOA predictions and triage of phenotypic hits. Chemical similarity profiling of phenotypic hit structures with target-annotated chemical databases enables "target prediction" and "target selectivity" scoring. Profiling MOA at the transcriptomic and post-translational pathway by high-precision transcriptomics and antibody-based proteomics following dose-response and time-series studies of hits tested across sensitive and insensitive phenotypic assays. Pathway network analysis incorporating *in silico* target prediction and empirical transcriptomic and post-translational profiling to further elucidate target and biomarker hypotheses. Black arrows point to outputs of each approach, shaded blue arrows and text boxes indicate a sequential process. Cross-correlation between distinct methods corroborates target and pathway hypotheses to prioritize and progress phenotypic hits, independent of, or before, subsequent target deconvolution and target validation studies.

include: elucidate target hypothesis and place MOA of phenotypic hits at the protein network level, including identification of novel orthogonal targets (e.g., upstream/downstream regulators, protein-protein interactions representing target co-activators or repressors); identify common interactors (targets and pathways common to different phenotypic hits representing cross-validation); map elucidated targets to chemical library and drug databases to support further validation of target hypotheses; identify pharmacodynamic biomarkers for prioritized phenotypic hits and targets; elucidate rational drug-combination and polypharmacology hypotheses targeting redundant and cooperative signaling pathways; map compound-modulated pathway networks to disease-associated pathway networks to aid disease positioning. Such network-based drug target prioritization and identification complements reductionist target-centric drug discovery that aims to find the single most important cause of a given disease (Csermely et al., 2013). Network-based drug target prediction may highlight non-obvious hits, and "edge-targeting" of protein-protein interactions which are likely to provide additional target novelty and potentially additional target specificity compared with conventional druggable target preferences. Drug target networks allow scientists to visualize the system-wide target landscape and provide a more holistic view of drug or compound MOA which can assist with more opportunistic disease positioning and/or drug repositioning. In an example utilizing computational network analysis of public gene expression data, Dudley et al. (2011) predicted that topiramate, an anticonvulsant drug, had the potential to treat inflammatory bowel disease (IBD). This discovery was partially validated by demonstrating improvement in an IBD *in vivo* rodent model further highlighting the potential of network pharmacology approaches for discovering drug repositioning opportunities. Finally, network-based *in silico* prediction of human toxicity has been used to support pharamcovigilence studies and bridge the gap between animal toxicity studies and clinical trials. Databases of association networks between chemicals and toxicity-related proteins or processes, such as DITOP (drug-induced toxicity-related protein database; http://bioinf.xmu.edu.cn:8080/databases/DITOP/index.html [Zhang et al., 2007]), provide an opportunity for early triage of phenotypic hits which present network profiles with high prediction of human toxicity.

## OUTLOOK

An overly simplistic view of phenotypic screening is that it is an effective strategy for identification of new therapeutic targets from physiological-based assays to feed target-directed drug discovery operating models. Rather, phenotypic screens have two main deliverables: novel targets but also chemical matter that can be progressed toward the clinic, regardless of its target/MOA. As discussed in this review, target deconvolution represents a major bottleneck in PDD strategies. While drug target identification is certainly desirable whenever possible this is not a pre-requisite for candidate drug nomination or the initiation of clinical trials. What is required to secure large, multi-million dollar investment in optimizing a specific hit or series toward a clinical candidate is an accumulation of evidence that the clinical candidate will be safe and efficacious in the chosen indication. MOA information plays a critical role in building this confidence in hits and leads derived from phenotypic screens, as described in a recent article by Vincent et al. (2020). Caution should be taken to ensure that poorly designed phenotypic screening and target deconvolution strategies do not create expensive new drug discovery bottlenecks and further investment of significant chemistry resources on poorly validated targets. We therefore propose that the pathway from phenotypic screening to target deconvolution should not immediately follow one another. Rather phenotypic and pathway profiling to understand MOA within the context of complex biological systems represents a more logical and informative approach to progressing phenotypic hits (Figure 5). Initially, phenotypic hits should be carefully triaged through information-rich high-content profiling and increasingly more complex and disease-relevant secondary phenotypic assays to build further biological understanding and confidence in their translational potential. Further panel screening across *in vitro* toxicity assays, physiologically relevant assay formats, and phenotypic profiling against reference compound libraries will help select the most novel and desirable MOA for subsequent target deconvolution or pathway profiling. *In silico* target prediction, and transcriptomic and proteomic analysis of phenotypic hits, will further aid mechanistic triage and prioritization of the most appropriate phenotypic hits and/or the most appropriate target deconvolution strategies when required. Pathway profiling may also help support the design of specific pathway reporter assays to

support medicinal-chemistry optimization of potency and selectivity, identify biomarkers to progress preclinical or clinical development, and corroborate subsequent target deconvolution studies (Figure 5). This more in-depth biological investigation is well placed to shift the PDD bottleneck away from target deconvolution toward increased disease relevance, novelty, safety, and hopefully more informed and productive phenotypic-led drug discovery programs that can translate into improved efficacy and successful clinical translation.

Recent advances in AI technology, coupled with significant improvements in computational power and the evolving plethora of information-rich data resources, are promising to revolutionize how drug discovery organizations extract knowledge and interpret complex datasets. As described in this review, AI and machine learning have successfully been applied to high-content screening data to classify cell phenotypes and predict compound MOA at scale. Emerging AI applications include modeling drug MOA with large-scale gene expression profiles (Gao et al., 2021) and structure-based approaches to target deconvolution and polypharmacology profiling (Somody et al., 2017). Future applications of AI promise to surpass manual extraction and hypothesis generation from complex datasets providing significant efficiency gains in areas, such as target identification, compound design, and patient stratification (Mak and Pichika, 2019). Therefore, it is reasonable to also assume that AI will enable more advanced MOA and target deconvolution predictions by exploiting new types of integrated phenotypic, pathway network, and cheminformatic data resources, such as those described in this review. Key to the future success of AI in effectively exploiting such integrated datasets will be in the quality control and representation of the data to adequately inform AI methods to accurately predict the most promising drug target mechanisms across disease indications. Significant investment and rapid progress in the collection, integration, and curation of complex biological and chemical datasets from new academic and industrial AI groups will determine the true value of these approaches in the next few years.

Phenotypic screening across well-annotated patient-derived cell or tissue-organoid panels are well placed to help elucidate the next generation of therapies and associated biomarkers that target heterogeneous disease conditions. Such phenotypic screening will be greatly aided by the collection and curation of large biobanks of patient-derived material through initiatives such as the Human Cancer Models Initiative (Ledford, 2016). These models, combined with the latest emerging phenotypic and pathway-profiling technology, will serve as a valuable resource for the translational cancer research community. The ambition and hope are that emerging phenotypic and pathway-profiling technologies combined with appropriate biological model systems will enable further development of more effective phenotypic screening strategies and innovative medicine discovery across multiple disease areas of unmet need.

## SIGNIFICANCE

**This review describes recent developments in high-content imaging, high-throughput transcriptomic and antibody-based proteomic technologies, which enable a less reductionist approach to understanding drug mechanism-of-action in complex biological systems. We describe how these technologies can integrate with each other and emerging computational tools, including artificial intelligence and machine learning approaches to annotate small-molecule compounds with rich mechanistic information at phenotypic and pathway network levels at scale. We propose that such approaches may relieve the bottleneck in target deconvolution of hits identified from phenotypic screening assays and triage the most promising compounds for subsequent preclinical development and/or further target deconvolution studies. We further suggest that such phenotypic and pathway profiling of compound mechanism-of-action embraces the complexity of heterogeneous diseases of unmet need and represents a more appropriate approach to triaging phenotypic screening hits rather than reductionist target-centric profiling. The methods and approaches described are largely agnostic to target class and disease area and have widespread applicability to innovating drug discovery beyond common therapeutic target classes. These approaches also support the discovery and development of multi-targeted therapeutics and drug combination approaches to target multiple pathway redundancies and rapid disease evolution, which currently limit the efficacy of targeted therapies in many complex diseases. We describe how these multidisciplinary approaches converge into an effective operating model for discovery, validation, and prioritization of novel therapeutic candidates based on robust experimental evidence obtained from appropriate biological systems.**

## ACKNOWLEDGMENTS

We thank the following sources of financial support from the Medical Research Council (Institute of Genetics and Molecular Medicine) PhD studentship award to R.E.H. The Anne Forest Fund for Oesophageal Cancer Research. Cancer Research UK Small Molecule Drug Discovery project award to N.O.C. (C42454/A24892). Cancer Research UK (C42454/A28596) and Brain Tumour Charity (GN-000676) award to N.O.C. Graphical abstract was created with BioRender.com.

## AUTHOR CONTRIBUTIONS

R.E.H., R.J.R.E., and J.C.D., conducted the experiments. R.E.H., R.J.R.E., J.C.D., and N.O.C. contributed to the analysis of data and drafting of the manuscript.

## DECLARATION OF INTERESTS

N.O.C. is a scientific advisory board member and shareholder in Amplia Therapeutics Ltd (Melbourne, Australia); and founder, shareholder, and management consultant of PhenoTherapeutics Ltd (Edinburgh, UK) and Director of Ther-IP Ltd (Edinburgh).

## REFERENCES

Aguirre-Plans, J., Pinero, J., Menche, J., Sanz, F., Furlong, L.I., Schmidt, H., Oliva, B., and Guney, E. (2018). Proximal pathway enrichment analysis for targeting comorbid diseases via network endopharmacology. Pharmaceuticals (Basel) *11*, 61.

Akbani, R., Becker, K.F., Carragher, N., Goldstein, T., de Koning, L., Korf, U., Liotta, L., Mills, G.B., Nishizuka, S.S., Pawlak, M., et al. (2014). Realizing the promise of reverse phase protein arrays for clinical, translational, and basic

research: a workshop report: the RPPA (Reverse Phase Protein Array) society. Mol. Cell Proteomics *13*, 1625–1643.

Ando, D.M., McLean, C.Y., and Berndl, M. (2017). Improving phenotypic measurements in high-content imaging screens. bioRxiv, 161422. https://research.google/pubs/pub46293/.

Bang, Y.J., Van Cutsem, E., Feyereislova, A., Chung, H.C., Shen, L., Sawaki, A., Lordick, F., Ohtsu, A., Omuro, Y., Satoh, T., et al. (2010). Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet *376*, 687–697.

Bao, S., Wu, Q., McLendon, R.E., Hao, Y., Shi, Q., Hjelmeland, A.B., Dewhirst, M.W., Bigner, D.D., and Rich, J.N. (2006). Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature *444*, 756–760.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles—database and tools update. Nucleic Acids Res. *35*, D760–D765.

Becker, K.G., Barnes, K.C., Bright, T.J., and Wang, S.A. (2004). The genetic association database. Nat. Genet. *36*, 431–432.

Bray, M.A., Gustafsdottir, S.M., Rohban, M.H., Singh, S., Ljosa, V., Sokolnicki, K.L., Bittker, J.A., Bodycombe, N.E., Dancik, V., Hasaka, T.P., et al. (2017). A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. Gigascience *6*, 1–5.

Bray, M.A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., and Carpenter, A.E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat. Protoc. *11*, 1757–1774.

Caie, P.D., Walls, R.E., Ingleston-Orme, A., Daya, S., Houslay, T., Eagle, R., Roberts, M.E., and Carragher, N.O. (2010). High-content phenotypic profiling of drug response signatures across distinct cancer cells. Mol Cancer Ther *9*, 1913–1926.

Canham, S.M., Wang, Y., Cornett, A., Auld, D.S., Baeschlin, D.K., Patoor, M., Skaanderup, P.R., Honda, A., Llamas, L., Wendel, G., et al. (2020). Systematic chemogenetic library assembly. Cell Chem Biol *27*, 1124–1129.

Carter, A.J., Kraemer, O., Zwick, M., Mueller-Fahrnow, A., Arrowsmith, C.H., and Edwards, A.M. (2019). Target 2035: probing the human proteome. Drug Discov. Today *24*, 2111–2115.

Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., Carmona, M., Faulconbridge, A., Hercules, A., McAuley, E., et al. (2019). Open Targets Platform: new developments and updates two years on. Nucleic Acids Res. *47*, D1056–D1065.

Charboneau, L., Tory, H., Chen, T., Winters, M., Petricoin, E.F., 3rd, Liotta, L.A., and Paweletz, C.P. (2002). Utility of reverse phase protein arrays: applications to signalling pathways and human body arrays. Brief Funct. Genomic Proteomic *1*, 305–315.

Chen, J., Li, Y., Yu, T.S., McKay, R.M., Burns, D.K., Kernie, S.G., and Parada, L.F. (2012). A restricted cell population propagates glioblastoma growth after chemotherapy. Nature *488*, 522–526.

Chen, J.Y., Mamidipalli, S., and Huan, T. (2009). HAPPI: an online database of comprehensive human annotated and predicted protein interactions. BMC Genomics *10* (*Suppl 1*), S16.

Comess, K.M., McLoughlin, S.M., Oyer, J.A., Richardson, P.L., Stockmann, H., Vasudevan, A., and Warder, S.E. (2018). Emerging approaches for the identification of protein targets of small molecules—a practitioners' perspective. J. Med. Chem. *61*, 8504–8535.

Csermely, P., Korcsmaros, T., Kiss, H.J., London, G., and Nussinov, R. (2013). Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol. Ther. *138*, 333–408.

D'Arcy, P., Brnjic, S., Olofsson, M.H., Fryknas, M., Lindsten, K., De Cesare, M., Perego, P., Sadeghi, B., Hassan, M., Larsson, R., et al. (2011). Inhibition of proteasome deubiquitinating activity as a new cancer therapy. Nat. Med. *17*, 1636–1640.

Davis, M.E. (2016). Glioblastoma: overview of disease and treatment. Clin. J. Oncol. Nurs. *20*, S2–S8.

Dawson, J.C., Serrels, B., Byron, A., Muir, M.T., Makda, A., García-Muñoz, A., von Kriegsheim, A., Lietha, D., Carragher, N.O., and Frame, M.C. (2020). A synergistic anticancer FAK and HDAC inhibitor combination discovered by a novel chemical-genetic high-content phenotypic screen. Mol. Cancer Ther. *19*, 637–649.

De Wolf, H., Cougnaud, L., Van Hoorde, K., De Bondt, A., Wegner, J.K., Ceulemans, H., and Gohlmann, H. (2018). High-throughput gene expression profiles to define drug similarity and predict compound activity. Assay Drug Dev. Technol. *16*, 162–176.

Dudley, J.T., Sirota, M., Shenoy, M., Pai, R.K., Roedder, S., Chiang, A.P., Morgan, A.A., Sarwal, M.M., Pasricha, P.J., and Butte, A.J. (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci. Transl Med. *3*, 96ra76.

Eder, J., Sedrani, R., and Wiesmann, C. (2014). The discovery of first-in-class drugs: origins and evolution. Nat. Rev. Drug Discov. *13*, 577–587.

Ericson, E., Gebbia, M., Heisler, L.E., Wildenhain, J., Tyers, M., Giaever, G., and Nislow, C. (2008). Off-target effects of psychoactive drugs revealed by genome-wide assays in yeast. PLoS Genet. *4*, e1000151.

Flachner, B., Lorincz, Z., Carotti, A., Nicolotti, O., Kuchipudi, P., Remez, N., Sanz, F., Tovari, J., Szabo, M.J., Bertok, B., et al. (2012). A chemocentric approach to the identification of cancer targets. PLoS One *7*, e35582.

Gao, S., Han, L., Luo, D., Liu, G., Xiao, Z., Shan, G., Zhang, Y., and Zhou, W. (2021). Modeling drug mechanism of action with large scale gene-expression profiles using GPAR, an artificial intelligence platform. BMC Bioinformatics *22*, 17.

Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. *40*, D1100–D1107.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat. Biotechnol. *26*, 317–325.

Gfeller, D., Grosdidier, A., Wirth, M., Daina, A., Michielin, O., and Zoete, V. (2014). SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res. *42*, W32–W38.

Gheeya, J., Johansson, P., Chen, Q.R., Dexheimer, T., Metaferia, B., Song, Y.K., Wei, J.S., He, J., Pommier, Y., and Khan, J. (2010). Expression profiling identifies epoxy anthraquinone derivative as a DNA topoisomerase inhibitor. Cancer Lett. *293*, 124–131.

Ghosh, D., and Poisson, L.M. (2009). Omics" data and levels of evidence for biomarker discovery. Genomics *93*, 13–16.

Godinez, W.J., Hossain, I., Lazic, S.E., Davies, J.W., and Zhang, X. (2017). A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics *33*, 2010–2019.

Gorgulla, C., Boeszoermenyi, A., Wang, Z.F., Fischer, P.D., Coote, P.W., Padmanabha Das, K.M., Malets, Y.S., Radchenko, D.S., Moroz, Y.S., Scott, D.A., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. Nature *580*, 663–668.

Gustafsdottir, S.M., Ljosa, V., Sokolnicki, K.L., Anthony Wilson, J., Walpita, D., Kemp, M.M., Petri Seiler, K., Carrel, H.A., Golub, T.R., Schreiber, S.L., et al. (2013). Multiplex cytological profiling assay to measure diverse cellular states. PLoS One *8*, e80999.

Hamosh, A., Scott, A.F., Amberger, J., Bocchini, C., Valle, D., and McKusick, V.A. (2002). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. *30*, 52–55.

Han, X., Zhou, C., Jiang, M., Wang, Y., Wang, J., Cheng, Z., Wang, M., Liu, Y., Liang, C., Wang, J., et al. (2018). Discovery of RG7834: the first-in-class selective and orally available small molecule hepatitis B virus expression inhibitor with novel mechanism of action. J. Med. Chem. *61*, 10619–10634.

Hay, M., Thomas, D.W., Craighead, J.L., Economides, C., and Rosenthal, J. (2014). Clinical development success rates for investigational drugs. Nat. Biotechnol. *32*, 40–51.

Hieronymus, H., Lamb, J., Ross, K.N., Peng, X.P., Clement, C., Rodina, A., Nieto, M., Du, J., Stegmaier, K., Raj, S.M., et al. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell *10*, 321–330.

Hopkins, A.L. (2008). Network pharmacology: the next paradigm in drug discovery. Nat. Chem. Biol. *4*, 682–690.

Hughes, R.E., Elliott, R.J.R., Munro, A.F., Makda, A., O'Neill, J.R., Hupp, T., and Carragher, N.O. (2020). High-content phenotypic profiling in esophageal adenocarcinoma identifies selectively active pharmacological classes of drugs for repurposing and chemical starting points for novel drug discovery. SLAS Discov. *25*, 770–782.

Irwin, J.J., Gaskins, G., Sterling, T., Mysinger, M.M., and Keiser, M.J. (2018). Predicted biological activity of purchasable chemical space. J. Chem. Inf. Model. *58*, 148–164.

Janssens, R., Zhang, X., Kauffmann, A., de Weck, A., and Durand, E.Y. (2020). Fully unsupervised deep mode of action learning for phenotyping high-content cellular images. bioRxiv. https://doi.org/10.1101/2020.07.22.215459.

Jia, L., Jin, H., Zhou, J., Chen, L., Lu, Y., Ming, Y., and Yu, Y. (2013). A potential anti-tumor herbal medicine, Corilagin, inhibits ovarian cancer cell growth through blocking the TGF-beta signaling pathways. BMC Complement Altern. Med. *13*, 33.

Keenan, A.B., Wojciechowicz, M.L., Wang, Z., Jagodnik, K.M., Jenkins, S.L., Lachmann, A., and Ma'ayan, A. (2019). Connectivity mapping: methods and applications. Annu. Rev. Biomed. Data Sci. *2*, 69–92.

Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J., and Shoichet, B.K. (2007). Relating protein pharmacology by ligand chemistry. Nat. Biotechnol. *25*, 197–206.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., et al. (2012). The IntAct molecular interaction database in 2012. Nucleic Acids Res. *40*, D841–D846.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Koscielny, G., An, P., Carvalho-Silva, D., Cham, J.A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M., Papa, E., et al. (2017). Open Targets: a platform for therapeutic target identification and validation. Nucleic Acids Res. *45*, D985–D994.

Kraus, O.Z., Ba, J.L., and Frey, B.J. (2016). Classifying and segmenting microscopy images with deep multiple instance learning. Bioinformatics *32*, i52–i59.

Kremer, L., Schultz-Fademrecht, C., Baumann, M., Habenberger, P., Choidas, A., Klebl, B., Kordes, S., Scholer, H.R., Sterneckert, J., Ziegler, S., et al. (2017). Discovery of a novel inhibitor of the hedgehog signaling pathway through cell-based compound discovery and target prediction. Angew. Chem. Int. Ed. Engl. *56*, 13021–13025.

Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T.H., von Mering, C., Jensen, L.J., and Bork, P. (2014). Stitch 4: integration of protein-chemical interactions with user data. Nucleic Acids Res. *42*, D401–D407.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929–1935.

Lathia, J.D., Mack, S.C., Mulkearns-Hubert, E.E., Valentim, C.L., and Rich, J.N. (2015). Cancer stem cells in glioblastoma. Genes Dev. *29*, 1203–1217.

Ledford, H. (2016). Global initiative seeks 1,000 new cancer models. Nature. https://doi.org/10.1038/nature.2016.20242 https://www.scientificamerican.com/article/global-initiative-seeks-1-000-new-cancer-models/.

Lee, E., Yong, R.L., Paddison, P., and Zhu, J. (2018). Comparison of glioblastoma (GBM) molecular classification methods. Semin. Cancer Biol. *53*, 201–211.

Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012). Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. Cell *149*, 780–794.

Liggi, S., Drakakis, G., Koutsoukas, A., Cortes-Ciriano, I., Martinez-Alonso, P., Malliavin, T.E., Velazquez-Campoy, A., Brewerton, S.C., Bodkin, M.J., Evans, D.A., et al. (2014). Extending in silico mechanism-of-action analysis by annotating targets with pathways: application to cellular cytotoxicity readouts. Future Med. Chem. *6*, 2029–2056.

Lim, H., Poleksic, A., Yao, Y., Tong, H., He, D., Zhuang, L., Meng, P., and Xie, L. (2016). Large-scale off-target identification using fast and accurate dual regularized one-class collaborative filtering and its application to drug repurposing. Plos Comput. Biol. *12*, e1005135.

Lin, A., Giuliano, C.J., Palladino, A., John, K.M., Abramowicz, C., Yuan, M.L., Sausville, E.L., Lukow, D.A., Liu, L., Chait, A.R., et al. (2019). Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. Sci. Transl Med. *11*, eaaw8412.

Lin, X., Huang, X.P., Chen, G., Whaley, R., Peng, S., Wang, Y., Zhang, G., Wang, S.X., Wang, S., Roth, B.L., et al. (2012). Life beyond kinases: structure-based discovery of sorafenib as nanomolar antagonist of 5-HT receptors. J. Med. Chem. *55*, 5749–5759.

Liu, C., Su, J., Yang, F., Wei, K., Ma, J., and Zhou, X. (2015). Compound signature detection on LINCS L1000 big data. Mol. Biosyst. *11*, 714–722.

Liu, X., Baarsma, H.A., Thiam, C.H., Montrone, C., Brauner, B., Fobo, G., Heier, J.S., Duscha, S., Konigshoff, M., Angeli, V., et al. (2016). Systematic identification of pharmacological targets from small-molecule phenotypic screens. Cell Chem Biol *23*, 1302–1313.

Liu, X., Vogt, I., Haque, T., and Campillos, M. (2013). HitPick: a web server for hit identification and target prediction of chemical screenings. Bioinformatics *29*, 1910–1912.

Ljosa, V., Caie, P.D., Ter Horst, R., Sokolnicki, K.L., Jenkins, E.L., Daya, S., Roberts, M.E., Jones, T.R., Singh, S., Genovesio, A., Clemons, P.A., Carragher, N.O., Carpenter, A.E., et al. (2013). Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. J Biomol Screen *18*, 1321–1329.

Lo, Y.C., Senese, S., Li, C.M., Hu, Q., Huang, Y., Damoiseaux, R., and Torres, J.Z. (2015). Large-scale chemical similarity networks for target profiling of compounds identified in cell-based chemical screens. Plos Comput. Biol. *11*, e1004153.

Lu, J.J., Pan, W., Hu, Y.J., and Wang, Y.T. (2012). Multi-target drugs: the trend of drug research and development. PLoS One *7*, e40262.

Mak, K.K., and Pichika, M.R. (2019). Artificial intelligence in drug development: present status and future prospects. Drug Discov. Today *24*, 773–780.

Martinez-Jimenez, F., Papadatos, G., Yang, L., Wallace, I.M., Kumar, V., Pieper, U., Sali, A., Brown, J.R., Overington, J.P., and Marti-Renom, M.A. (2013). Target prediction for an open access set of compounds active against. Mycobacterium Tuberculosis. Plos Comput. Biol. *9*, e1003253.

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., et al. (2009). Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. *37*, D619–D622.

McFarland, J.M., Paolella, B.R., Warren, A., Geiger-Schuller, K., Shibue, T., Rothberg, M., Kuksenko, O., Colgan, W.N., Jones, A., Chambers, E., et al. (2020). Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. Nat. Commun. *11*, 4296.

Moffat, J.G., Rudolph, J., and Bailey, D. (2014). Phenotypic screening in cancer drug discovery—past, present and future. Nat. Rev. Drug Discov. *13*, 588–602.

Moffat, J.G., Vincent, F., Lee, J.A., Eder, J., and Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nat. Rev. Drug Discov. *16*, 531–543.

Nidhi, Glick, M., Davies, J.W., and Jenkins, J.L. (2006). Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J. Chem. Inf. Model. *46*, 1124–1133.

Parada, L.F., Dirks, P.B., and Wechsler-Reya, R.J. (2017). Brain tumor stem cells remain in play. J. Clin. Oncol. *35*, 2428–2431.

Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N., Soldatos, T.G., Kossida, S., Aerts, J., Schneider, R., and Bagos, P.G. (2011). Using graph theory to analyze biological networks. BioData Min *4*, 10.

Pawlak, M., Schick, E., Bopp, M.A., Schneider, M.J., Oroszlan, P., and Ehrat, M. (2002). Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. Proteomics *2*, 383–393.

Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M., et al. (2003). Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res. *13*, 2363–2371.

Perlman, Z.E., Slack, M.D., Feng, Y., Mitchison, T.J., Wu, L.F., and Altschuler, S.J. (2004). Multidimensional drug profiling by automated microscopy. Science *306*, 1194–1198.

Plouffe, D., Brinker, A., McNamara, C., Henson, K., Kato, N., Kuhen, K., Nagle, A., Adrian, F., Matzen, J.T., Anderson, P., et al. (2008). In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. Proc. Natl. Acad. Sci. U S A. *105*, 9059–9064.

Poroikov, V., Filimonov, D., Lagunin, A., Gloriozova, T., and Zakharov, A. (2007). PASS: identification of probable targets and mechanisms of toxicity. SAR QSAR Environ. Res. *18*, 101–110.

Qu, X.A., and Rajpal, D.K. (2012). Applications of Connectivity Map in drug discovery and development. Drug Discov. Today *17*, 1289–1298.

Rapkiewicz, A., Espina, V., Zujewski, J.A., Lebowitz, P.F., Filie, A., Wulfkuhle, J., Camphausen, K., Petricoin, E.F., 3rd, Liotta, L.A., and Abati, A. (2007). The needle in the haystack: application of breast fine-needle aspirate samples to quantitative protein microarray technology. Cancer *111*, 173–184.

Ratni, H., Ebeling, M., Baird, J., Bendels, S., Bylund, J., Chen, K.S., Denk, N., Feng, Z., Green, L., Guerard, M., et al. (2018). Discovery of risdiplam, a selective survival of motor neuron-2 (SMN2) gene splicing modifier for the treatment of spinal muscular atrophy (SMA). J. Med. Chem. *61*, 6501–6517.

Rees, M.G., Seashore-Ludlow, B., Cheah, J.H., Adams, D.J., Price, E.V., Gill, S., Javaid, S., Coletti, M.E., Jones, V.L., Bodycombe, N.E., et al. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat. Chem. Biol. *12*, 109–116.

Reisen, F., Sauty de Chalon, A., Pfeifer, M., Zhang, X., Gabriel, D., and Selzer, P. (2015). Linking phenotypes and modes of action through high-content screen fingerprints. Assay Drug Dev. Technol. *13*, 415–427.

Reker, D., Rodrigues, T., Schneider, P., and Schneider, G. (2014). Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc. Natl. Acad. Sci. U S A. *111*, 4067–4072.

Rena, G., Hardie, D.G., and Pearson, E.R. (2017). The mechanisms of action of metformin. Diabetologia *60*, 1577–1585.

Rummun, N., Hughes, R.E., Beesoo, R., Li, W.W., Aldulaimi, O., Macleod, K.G., Bahorun, T., Carragher, N.O., Kagansky, A., and Neergheen-Bhujun, V.S. (2019). Mauritian endemic medicinal plant extracts induce G2/M phase cell cycle arrest and growth inhibition of oesophageal squamous cell carcinoma in vitro. Acta Naturae *11*, 81–90.

Scavone, C., di Mauro, G., Mascolo, A., Berrino, L., Rossi, F., and Capuano, A. (2019). The new paradigms in clinical research: from early access programs to the novel therapeutic approaches for unmet medical needs. Front. Pharmacol. *10*, 111.

Schenone, M., Dancik, V., Wagner, B.K., and Clemons, P.A. (2013). Target identification and mechanism of action in chemical biology and drug discovery. Nat. Chem. Biol. *9*, 232–240.

Schneider, P., and Schneider, G. (2017). De-orphaning the marine natural product (+/–)-marinopyrrole A by computational target prediction and biochemical validation. Chem. Commun. (Camb) *53*, 2272–2274.

Schneidewind, T., Brause, A., Pahl, A., Burhop, A., Mejuch, T., Sievers, S., Waldmann, H., and Ziegler, S. (2020). Morphological profiling identifies a common mode of action for small molecules with different targets. Chembiochem *21*, 3197–3207.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Somody, J.C., MacKinnon, S.S., and Windemuth, A. (2017). Structural coverage of the proteome for pharmaceutical applications. Drug Discov. Today *22*, 1792–1799.

Spitzmuller, A., and Mestres, J. (2013). Prediction of the *P. falciparum* target space relevant to malaria drug discovery. Plos Comput. Biol. *9*, e1003257.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell *171*, 1437–1452 e1417.

Swinney, D.C., and Anthony, J. (2011). How were new medicines discovered? Nat. Rev. Drug Discov. *10*, 507–519.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362–D368.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a cancer dependency map. Cell *170*, 564–576.e16.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

van Oostrum, J., Calonder, C., Rechsteiner, D., Ehrat, M., Mestan, J., Fabbro, D., and Voshol, H. (2009). Tracing pathway activities with kinase inhibitors and reverse phase protein arrays. Proteomics Clin. Appl. *3*, 412–422.

Vincent, F., Loria, P.M., Weston, A.D., Steppan, C.M., Doyonnas, R., Wang, Y.M., Rockwell, K.L., and Peakman, M.C. (2020). Hit triage and validation in phenotypic screening: considerations and strategies. Cell Chem Biol *27*, 1332–1346.

Volochnyuk, D.M., Ryabukhin, S.V., Moroz, Y.S., Savych, O., Chuprina, A., Horvath, D., Zabolotna, Y., Varnek, A., and Judd, D.B. (2019). Evolution of commercially available compounds for HTS. Drug Discov. Today *24*, 390–402.

Wale, N., and Karypis, G. (2009). Target fishing for chemical compounds using target-ligand activity data and ranking based methods. J. Chem. Inf. Model. *49*, 2190–2201.

Wang, I.M., Stone, D.J., Nickle, D., Loboda, A., Puig, O., and Roberts, C. (2013). Systems biology approach for new target and biomarker identification. Curr. Top. Microbiol. Immunol. *363*, 169–199.

Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., and Bryant, S.H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. *37*, W623–W633.

Warchal, S.J., Dawson, J.C., and Carragher, N.O. (2019). Evaluation of machine learning classifiers to predict compound mechanism of action when transferred across distinct cell lines. SLAS Discov. *24*, 224–233.

Warchal, S.J., Unciti-Broceta, A., and Carragher, N.O. (2016). Next-generation phenotypic screening. Future Med. Chem. *8*, 1331–1347.

Willis, C., Nyffeler, J., and Harrill, J. (2020). Phenotypic profiling of reference chemicals across biologically diverse cell types using the cell painting assay. SLAS Discov. *25*, 755–769.

Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. *34*, D668–D672.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.

Ye, C., Ho, D.J., Neri, M., Yang, C., Kulkarni, T., Randhawa, R., Henault, M., Mostacci, N., Farmer, P., Renner, S., et al. (2018). DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. Nat. Commun. *9*, 4307.

Young, D.W., Bender, A., Hoyt, J., McWhinnie, E., Chirn, G.-W., Tao, C.Y., Tallarico, J.A., Labow, M., Jenkins, J., Mitchison, T.J., Feng, Y., et al. (2008). Integrating high-content screening and ligand-target prediction to identify mechanism of action. Nat Chem Biol. *4*, 59–68.

Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., and Cesareni, G. (2002). MINT: a Molecular INTeraction database. FEBS Lett. *513*, 135–140.

Zhang, J.X., Huang, W.J., Zeng, J.H., Huang, W.H., Wang, Y., Zhao, R., Han, B.C., Liu, Q.F., Chen, Y.Z., and Ji, Z.L. (2007). DITOP: drug-induced toxicity related protein database. Bioinformatics *23*, 1710–1712.

Zhao, S., and Iyengar, R. (2012). Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annu. Rev. Pharmacol. Toxicol. *52*, 505–521.