Title

Identification of Genomic Predictors of Response to the CDK4/6 Inhibitor Palbociclib using the UCLATORL Panel of Human Cancer Cell Lines

Permalink

https://escholarship.org/uc/item/8633b778

Author

Conklin, Dylan Francis

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Identification of Genomic Predictors of Response to the

CDK4/6 Inhibitor Palbociclib using the UCLATORL Panel of Human Cancer Cell Lines

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Epidemiology

by

Dylan Francis Conklin

2013

ABSTRACT OF THE DISSERTATION

Identification of Genomic Predictors of Response to the

CDK4/6 Inhibitor Palbociclib using the UCLATORL Panel of Human Cancer Cell Lines

by

Dylan Francis Conklin

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2013

Professor Zuo-Feng Zhang, Co-Chair

Professor Richard Finn, Co-Chair

The landscape of genetic alterations that can lead to cancer is vast and complex. Precancerous cells accumulate mutations that affect the various molecular pathways involved in cancer progression. In the last two decades, hundreds of novel therapeutics designed to inhibit different molecular targets within these oncogenic pathways have recently entered clinical development. The success or failure of these compounds will depend on the ability to correctly identify subpopulations of cancer patients likely to be sensitive or resistant to these therapies. Patient-specific sensitivity to treatment is likely to be determined, at least in part, to the underlying genomic alterations that occurred during the development of the patient's particular cancer.

Palbociclib is a highly specific small molecule inhibitor of cyclin-dependent kinases 4 and 6 currently in clinical development by Pfizer. It is designed to inhibit the cell cycle at the G1/S transition via preventing the phosphorylation of Rb by the Cyclin D/CDK4/6 enzyme complex. Our lab previously identified the estrogen receptor positive (ER+) subpopulation of breast cancer patients as distinctively likely to benefit from treatment with palbociclib. This observation spurred the initiation of a Phase II clinical trial in this patient population where remarkable efficacy was observed. Given this translational success, we wished to investigate the molecular determinants of response to palbociclib in several additional cancer types.

To this end, we assayed the *in vitro* sensitivity to palbociclib across a panel of 416 cancer cell lines derived from 12 distinct cancer types. We observed highly differential response to treatment both within and between cancer types. IC50s (the concentration of palbociclib required to inhibit 50% of population doublings) ranged from the low nanomolar range to above the highest dose tested (1µM). This response distribution was much broader than was observed in similar screens performed by the Broad and Sanger institutes, where the vast majority of cell lines assayed were listed as having IC50s above 1µM. Our ability to identify a higher proportion of palbociclib -sensitive cell lines can likely be credited to methodological innovations aimed at optimizing our screening protocol for the detection of longer-term cytostatic effects, as opposed to shorter-term cytotoxic effects of treatment. The generation of this highly differential response dataset allowed for a unique opportunity to explore the possible genetic mechanisms underlying differential sensitivity to treatment with palbociclib *in vitro*.

We interrogated two large genomics datasets for genotype-response associations. The first of which consisted of whole exome point mutation data downloaded from the Cancer Cell Line Encyclopedia's hybrid capture sequencing database. This data was restricted by various criteria to enrich for functional, somatic point mutations that are known to be causally involved in carcinogenesis. From this restricted dataset we identified three proto-oncogenes (MLL, TSHR, SMO) where presumptive

activating mutations were associated with resistance to treatment with palbociclib across our cell line panel. We further identified five recessive cancer genes where likely loss-of-function point mutations were significantly associated with palbociclib response. Mutations in two of these genes (CDH1, TOPBP1) associated with palbociclib sensitivity, while mutations in the other three (RB1, FANCA, NBN) associated with resistance.

The other dataset interrogated for genotype-response associations was a copy number alteration dataset derived from comparative genomic hybridization arrays. This dataset was organized by gene and also restricted by various criteria to enrich for amplifications or deletions of genes likely to be causally involved in carcinogenesis. From this dataset we identified three chromosomal regions (17q12-21, 11q13, 1q32) where amplification was associated with sensitivity to palbociclib. Two amplified regions (19q13, 8q13) were found to be associated with resistance. Homozygous deletions of the 13q14 chromosomal region were found to be strongly associated with resistance to palbociclib.

Following the identification of these candidate palbociclib response biomarkers from the crude, semi-supervised screens, several post-hoc analyses were performed to strengthen the argument for causation for each of the genotype-response associations. These analyses included: (1) a comprehensive literature search to investigate the possible causal mechanism of each biomarker, (2) Pearson correlation analysis to identify inter-biomarker associations followed by multiple regression to isolate the independent effect of each variable, (3) control of confounding by cell line growth rate and histology, and (4) analysis of misclassification in the genomics datasets.

Following these extensive post-hoc analyses, we pruned our original set of fourteen biomarkers down to the eight most likely to play a causal role in determining sensitivity or resistance to treatment with palbociclib. The final set of candidate sensitivity biomarkers included: loss-of-function point mutations in CDH1, loss-of-function point mutations in TOPBP1, chromosomal amplification of 17q12-21 (ERBB2) and chromosomal amplification of 11q13 (CCND1). The final set of candidate resistance

biomarkers included: activating point mutations in SMO, chromosomal amplification of CCNE1, loss-of-function point mutations in RB1 and chromosomal deletion of 13q14 (RB1).

This final set of eight candidate biomarkers was analyzed by strata representing each of the 12 cancer types in our cell line panel. The four candidate sensitivity biomarkers were sufficiently frequent and associated with sensitivity in 6 of the 12 cancer types in our panel. These were the breast, colon, head/neck, lung, ovarian and upper gastrointestinal strata. The four candidate resistance biomarkers were sufficiently frequent and associated with resistance in 6 of the 12 cancer types in our panel. These were the breast, colon, kidney, lung, ovarian and upper gastrointestinal strata.

The identified palbociclib response biomarkers represent good candidates for clinical translation, as we have shown them to be independently associated with palbociclib response *in vitro* and have a strong biologic rationale for causality. Follow-up experiments can further elucidate the molecular biology behind these associations and further validate these biomarkers before they are applied to the clinical setting. Ultimately, these biomarkers may be clinically implemented across a wide range of cancer types to identify patient subpopulations most likely to benefit from treatment with palbociclib.

The dissertation of Dylan Francis Conklin is approved.

_____

Dennis Slamon

_____

Jian Yu Rao

_____

Frank Sorvillo

_____

Zuo-Feng Zhang, Committee Co-Chair

_____

Richard Finn, Committee Co-Chair

University of California, Los Angeles

2013

Contents

VITA:

EDUCATION:

University of California, Los Angeles
Undergraduate   **2001-2006**
- o   MS: Molecular, Cellular and Developmental Biology
- o   MS: Biochemistry

University of California, Los Angeles
Graduate   **2008-2013**
- o   MPH: Epidemiology
- o   PhD: Epidemiology (In Progress)

RESEARCH:

UCLA Translational Oncology Research Lab, Santa Monica CA
Staff Research Associate   **2006 – 2008**

UCLA Translational Oncology Research Lab, Santa Monica CA
Graduate Student Researcher   **2008 – Present**

PUBLICATIONS AND PAPERS:

*Finn RS, Dering J, **Conklin D**, Kalous O, Cohen DJ, Desai AJ, Ginther C, Atefi M, Chen I, Fowst C, Los G, Slamon DJ.*
PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal
estrogen receptor-positive human breast cancer cell lines in vitro. *Breast cancer research : BCR* **11**,
R77, doi:10.1186/bcr2419 (2009)

*Kalous O, **Conklin D**, Desai AJ, O'Brien NA, Ginther C, Anderson L, Cohen DJ, Britten CD, Taylor I, Christensen JG, Slamon DJ, Finn RS.*
Dacomitnib (PF-00299804), an irreversible Pan-HER inhibitor, inhibits proliferation of HER2-
amplified breast cancer cell lines resistant to trastuzumab and lapatinib. *Molecular cancer
therapeutics* **11**, 1978-1987, doi:10.1158/1535-7163.MCT-11-0730 (2012).

*Logan JE, Mostofizadeh N, Desai AJ, VON Euw E, **Conklin D**, Konkankit V, Hamidi H, Eckardt M, Anderson L, Chen HW, Ginther C, Taschereau E,
Bui PH, Christensen JG, Belldegrun AS, Slamon DJ, Kabbinavar FF.*
PD-0332991, a Potent and Selective Inhibitor of Cyclin-dependent Kinase 4/6, Demonstrates
Inhibition of Proliferation in Renal Cell Carcinoma at Nanomolar Concentrations and Molecular
Markers Predict for Sensitivity. *Anticancer research* **33**, 2997-3004 (2013).

*Wainberg ZA, Anghel A, Rogers AM, Desai AJ, Kalous O, **Conklin D**, Ayala R, O'Brien NA, Quadt C, Akimov M, Slamon DJ, Finn RS.*
Inhibition of HSP90 with AUY922 induces synergy in HER2-amplified trastuzumab-resistant
breast and gastric cancer. *Molecular cancer therapeutics* **12**, 509-519, doi:10.1158/1535-7163.MCT-
12-0507 (2013).

CHAPTER 1: Introduction

1.1. Targeted cancer therapeutics

For many decades, the primary clinical treatments available to cancer patients were cytotoxic chemotherapeutics and radiation therapies[1] designed to preferentially kill cells that are dividing rapidly. Due to a lack of tumor-specificity, most of these interventions cause severe side effects such as nausea, myelosuppression, and neurological problems amongst many others. The efficacy of these traditional treatments is disputed for many cancer types[2], with some going as far to say that the intervention may be doing more harm than good in certain indications[3]. Due to these limitations, there has been a strong push to develop "targeted" cancer drugs that can specifically kill or inhibit the growth of cancerous cells, while leaving normal tissue intact.

Cancer is primarily a genetic disease. Normal cells become cancerous through an accumulation of genetic and epigenetic changes that lead to a dysregulation of the balance between cell growth and cell death[4]. Recent advances in the understanding of these molecular hallmarks of human cancer have allowed for the development of drugs that have been designed to target specific oncogenic alterations[5-7]. Many of these molecularly targeted compounds have shown to be very effective in specific patient populations, while having a much more favorable toxicity profile than traditional non-targeted cytotoxics. These clinical successes include: Herceptin in HER2-Amplified breast cancer[8], Gleevec in chronic myeloid leukemia[9], and Gefitinib in EGFR-mutant non-small cell lung cancer[10] amongst many others.

1.2. Biomarkers of response to treatment

Despite the early successes of several targeted therapies, there exists a high degree of differential response to treatment that cannot be explained simply by characterization of the therapeutic target of the compound. There is strong evidence that tumor-specific, genomic alterations may explain a large portion of this variation[11-20] for each therapy. Specific genetic alterations may increase or decrease the dependence on the molecular target of inhibition by a certain drug, consequently resulting in increased sensitivity or resistance to treatment.

One striking example of this phenomenon is EGFR-targeted therapies in lung cancer. Activating mutations in the oncogene KRAS have been shown to be strongly associated with resistance to EGFR-targeted therapies in non-small cell lung cancers[21]. This is likely due to the fact that KRAS is downstream of EGFR in its signaling cascade, where constitutively activating alterations circumvent the need for activation by the EGFR kinase.

A functionally similar argument has been made for the role of PI3K mutations in determining response to treatment with the anti-HER2 therapy, Trastuzumab[22,23]. PI3K is a necessary downstream effector of HER2 signaling. A large proportion of HER2-amplified breast cancers also harbor downstream activating mutations in PI3K. These cancers presumably have evolved to be independent of the original HER2 activating event. As predicted by this mechanistic rationale, HER2-positive breast cancers that also harbor PI3K activating mutations have been shown to be less responsive to HER2 targeted therapies.

1.3. Preclinical modeling of differential response to treatment using cell line panels

The overwhelming number of new therapies makes the clinical exploration of these potential response biomarkers very difficult. Thus, there is a need for accurate pre-clinical models of the

molecular determinants of response to novel targeted therapeutics. These pre-clinical observations can be employed to guide treatment decisions for existing therapeutics as well as to enrich clinical trials for those patients most likely to respond to treatment.

To this end, large panels of human cancer cell lines have been developed and used to assess the efficacy of new cancer treatments as well as identify subpopulations of tumor types likely to be more responsive to treatment. Cell lines are samples that have been derived from human tumors and conditioned to grow indefinitely in culture. Panels of these cell lines can be interrogated for their response to various cancer treatments *in vitro* via multiple methods including: 2-dimensional proliferation assays, 3-dimensional colony formation assays in gel substrates and mouse xenograft models.

Cell-line based drug screening efforts began with the NCI-60[24], where many of the laboratory techniques for drug screening were developed. However, it quickly became apparent that 60 cell lines was not a large enough sample to adequately capture the huge amount of genetic and epigenetic heterogeneity present in clinical cancer populations[25]. Screening efforts with significantly larger cell line panels (such as the Broad Institute's Cancer Cell Line Encyclopedia (CCLE)[26] and the Wellcome Trust Sanger Institute's Genomics of Drug Sensitivity in Cancer (GDSC)[27] project) have already proven successful in identifying several novel response biomarkers to certain molecular targeted therapies.

The UCLA Translational Oncology Research Laboratory (UCLATORL) runs another large-scale, pre-clinical screening project. UCLATORL aims to add to this body of knowledge by screening compounds that have not yet been looked at by CCLE and GDSC. UCLATORL also implements unique experimental protocols and analytical methods that may provide novel perspective. The UCLATORL database currently holds response data for 100+ compounds tested against a diverse panel of 600+ cell lines.

1.4. The G1/S transition, cyclin D and CDK4/6

1.4.1 Signaling pathways

When cells are not actively proliferating they are generally resting in G_1 phase. This non-mitotic state is maintained by limiting the expression of pro-growth factors like CDK2 and cyclin E. However, in response to mitogenic signals, cells begin to synthesize high levels of cyclin D. Cyclin D forms heterodimeric complexes with CDKs 4 and 6[28]. These cyclin D - CDK4/6 complexes facilitate the G_1/S transition by phosphorylating Rb, which disrupts binding the repressive binding of Rb to the transcription factor E2F. E2F activation results in the transcription of many target genes necessary for cell cycle progression.

The G_1/S transition is further controlled by positive and negative regulators of the cyclin D – CDK4/6 complex. Members of the INK4 family of proteins can associate with CDKs 4 and 6 and cause the release of cyclin D and also dissociation of Cip/Kip proteins from the complex. These Cip/Kip proteins subsequently associate with cyclin E-CDK2 to cause G_1 arrest.

1.4.2. Dysregulation by oncogenic alterations

Oncogenic genetic alterations that disrupt the G_1/S transition are almost omnipresent in human cancers. There is a very high prevalence of p16 (INK4a) loss in human cancers[29], caused by chromosomal deletions, point mutations, transcriptional silencing or some combination of events. Loss of p16 causes an increase in CDK4/6 enzymatic activity, which in turn can lead to an increase in cell proliferation.

CDK4 has been shown to be both genomically amplified or mutated to prevent suppressive INK4 binding in some cases[30,31]. This also has been shown to lead to increased proliferation of the mutant cells[32]. Overexpression of cyclin-D1 by genomic amplification of the CCND1 gene is also common in many cancers[33-38]. This overexpression is thought to stimulate cell proliferation independently of

mitogenic signals. Loss of Rb is another well described genomic abnormality that results in constitutive activation of E2F [39-43].

1.5. CDK inhibitors as cancer therapies

Several small molecule inhibitors of cyclin-dependent kinases have been developed and tested in various stages of clinical development, but to date no CDK inhibitors have been approved for commercial use[44]. Several of the early-generation small molecules were developed as pan-CDK inhibitors, meaning they had enzymatic activity against several kinases in the CDK family.

Flavopiridol (alvociclib) has demonstrated enzymatic activity against CDKs 1, 2, 4, 6, 7 and 9, with enzymatic IC50s ranging from 3-400nM for each of these targets[45]. It has been shown to cause G1 and G2 arrest as well as apoptosis in cell line models[46]. It is currently in Phase II testing.

P276-00 is another pan-CDK inhibitor in development by Nicholas Piramal[47]. It has enzymatic activity against CDKs 1, 4, 6, 7 and 9[47]. Like Flavopiridol, it has shown promising preclinical activity in cell line models[48]. It is currently in Phase I/II testing for patients with multiple myeloma, mantle cell lymphoma, head and neck cancer, and melanoma.

AG-024322 is a potent inhibitor of CDKs 1, 2 and 4 developed by Pfizer that showed strong enzymatic selectivity for these targets over non-CDK kinases[49]. It also showed strong preclinical activity in cell line and animal models[50], but was discontinued after early clinical trials failed to showed increased benefit over other treatment options[51].

Several other pan-CDK inhibitors that have CDK4 and/or CDK6 among their kinase targets have shown similar activity in preclinical models and are at various stages of clinical development. These include: Hoffman-LaRoche's R547[52], Shering-AG's ZK 304709[53], and Astex's AT7519[54]. Most of these pan-CDK inhibitors have shown little to modest benefit in clinical testing to date[51,52,55], and some have had

toxicity profiles that have limited their development[44]. These failures may be due to non-specific, off-targeting effects that contribute to *in vivo* toxicity, unfavorable pharmacodynamic profiles that make it difficult to achieve and sustain functionally relevant doses of the compound, or some combination of the two.

1.6. Palbociclib

1.6.1 Description of compound

Among the small molecule CDKis in clinical development, palbociclib is one of the few inhibitors that is highly selective for CDKs 4 and 6. It inhibits the kinase activity of these two targets with very high potency (enzymatic IC50 ~ 10nmol/L). It has exhibited little to no activity against a panel of 36 additional protein kinases (including other CDKs), indicating that CDK4/6 inhibition is likely the sole mechanism of its anti-proliferative action[56].

It has been extensively mechanistically characterized in cell line models[56], where it has been shown to induce G1 arrest through a concomitant reduction in Rb phosphorylation at Ser^{780}/Ser^{795}. Therapeutic doses *in vitro* have been shown to eliminate phospho-Rb and down-regulate the proliferation marker Ki-67 in Western blot analyses. This was further demonstrated to result in the down-regulation of E2F target genes.

1.6.2. Existing preclinical data

Pre-clinical evidence demonstrates that CDK4/6 inhibition by palbociclib can inhibit cancer cell proliferation both *in vitro* and in mouse xenografts[56,57]. To date, antiproliferative effects have been observed in breast cancer[57], ovarian cancer[58], rhabdoid[59], lymphoma[60], glioblastoma[61] and myeloma[62] cell line models.

Our lab has previously published two papers describing differential sensitivity of palbociclib within ovarian[58] and breast[57] cancer models. These site-specific analyses analyzed candidate genomic biomarkers that may explain some of the differential sensitivity *in vitro*. These candidates included breast cancer subtype, Rb expression, p16 expression and genomic gains of CCND1. Since these papers were published, we have obtained response data on nine more cancer types and have dramatically expanded the available genetic characterization data available for association testing. The purpose of this study is to expand these analyses into other cancer types and screen a broader set of genomic predictors.

1.6.3 Clinical trials

Results from the first Phase I, dose-escalation trial of palbociclib were published in 2011[63]. Palbociclib was administered orally once-daily in Rb-positive, advanced solid tumors. The compound was tolerated very well at doses up to 100mg/day. Dose-limiting toxicities were observed in five patients at the 125mg and 150mg daily dose. Neutropenia was the only dose-limiting effect, and it was non-symptomatic. Ultimately, 125mg per day was chosen as the suggested dose for Phase II testing.

Recently, UCLATORL was able to demonstrate that the ER-positive subtype of breast cancer was uniquely sensitive to palbociclib *in vitro*[57]. This observation served as the basis for the design of a Phase II clinical trial where remarkable efficacy was observed[64]. Post-menopausal women with ER+/HER2-advanced breast cancer were randomized 1:1 to receive letrozole with or without palbociclib. There was a significant improvement in progression free survival in the letrozole + palbociclib arm (HR = 0.35, p=0.006). This clinical efficacy is highly encouraging, and validates the preclinical sensitivity observed in ER+ cell line models.

1.7. Overview of the study

Despite its strong therapeutic potential there is likely to be significant heterogeneity in response to treatment palbociclib, both between and within cancer types. This differential sensitivity may depend on specific genomic alterations that influence a particular cancer's dependence on cyclin D - CDK4/6 function. These alterations may act directly within the target pathway, or exert their influence more indirectly via complex pathway interactions or redundant 'escape' pathways. The goal of this proposed project is to identify genomic alterations that are most strongly associated with response to palbociclib, when other potential confounding factors are fully considered and adjusted for. It is our hope that these observations may help guide the clinical development of palbociclib by identifying those patients most likely to respond to therapy.

We assayed the *in vitro* response to palbociclib across a large panel of 400+ cell lines, representing tumors derived from 12 different primary sites. We screened two large genomics datasets for associations with *in vitro* response to palbociclib. One dataset consists of somatic point mutations identified via whole-exome, hybrid-capture sequencing performed by the Broad Institute. The other dataset consists of whole-genome copy number alterations as identified by internally-performed comparative genomic hybridization arrays.

After a set of candidate biomarkers was identified from these semi-supervised screens, we carefully restricted this list to create a final predictor cassette. These restriction criteria included: a strong mechanistic rationale for the role in palbociclib response, independence of other predictors in the model and robustness to control of confounding factors in multiple regression models.

CHAPTER 2: The UCLATORL cell line panel

2.1. Description of the panel

2.1.1. Distribution of cell lines by primary site

UCLATORL maintains a large panel of 612 human cancer cell lines from 14 different tissue types. A large majority of the cell lines have been obtained from commercial vendors (ATCC, DSMZ, JCRB, ECACC) and academic collaborations. A small subset of the lines has been established internally from patient samples. Many of the cell lines in UCLATORL's panel have growth characteristics that make them difficult or impossible to assay reliably for drug response. UCLATORL was able to generate palbociclib dose-response data on 416 cell lines (detailed in Chapter 3). These 416 cell lines served as the study population in all of the analyses presented in this paper. A quantification of these cell lines by primary site is presented in Figure 1.

2.1.2. Confirmation of cell line identity

Misidentification of cell lines is a very common problem[65-69] and can lead to serious information bias in cell line based screening studies. To minimize this concern, all UCLATORL cell lines have been verified as unique via mitochondrial DNA sequencing. Where possible, STR profiling of cell lines has been performed to match cell line identity to published profiles from vendors as well as public characterization databases such as COSMIC and the CCLE.

2.1.3. Other quality control measures

Cell lines are maintained in T75 culture flasks using the appropriate growth media specified by the vendor. Cell lines are subcultured every 3 to 5 days. Cell lines are grown for at least three passages before proliferation assays are performed to ensure log-phase growth at the time of experiment. Cell

lines are never grown for more than 20 passages to prevent clonal selection effects from altering the characteristics of the line. All cell lines have been verified to be free of *Mycoplasma* contamination at the time of use in proliferation experiments via PCR with *Mycoplasma* -specific primers.

2.2. Genomic characterization datasets

2.2.1. UCLATORL copy number alteration data

Genome-wide characterization of copy number alterations has been performed on 317 of the cell lines in UCLATORL's palbociclib response panel. The distribution of the cell lines by primary site is presented in

Table 1.

Genomic DNA for each cell line was labeled and hybridized to an Agilent 105K CGH array chip according to the manufacturer's protocols. Data was subsequently extracted using Agilent Feature Extraction Software and analyzed using Agilent's CGH analytics software. Log_2(Ratio)s larger than 1 (i.e. a 2 fold copy number gain) were classified as amplified. Log_2(Ratio)s less than -2 were classified as homozygous deletions.

2.2.2. CCLE copy number alteration data

CCLE CNA data was downloaded from the Broad Institute's publicly available database and incorporated into UCLATORL's database for those cell lines where the two panels overlap. This data was available on 262 of the 416 cell lines in UCLATORL's palbociclib response panel. The coverage of this data by primary site to date is also presented in

Table 1.

Again, Log_2(Ratio)s larger than 1 (i.e. a 2 fold copy number gain) were classified as amplified. Log_2(Ratio)s less than -2 were classified as homozygous deletions. UCLATORL treats copy number

alterations as a categorical variable rather than a continuous variable because the Log_2(Ratio)s output from the array software correspond to discrete genomic states.

2.2.2. CCLE hybrid capture sequencing data

The Cancer Cell Line Encyclopedia project has sequenced 4059 protein-coding genes via solution phase hybrid capture and massively parallel sequencing. This point mutation data was downloaded from the Broad Institute's publicly available database[26] and incorporated into UCLATORL's database for those cell lines where the two panels overlap. Data was available on 240 of the 416 cell lines in UCLATORL's palbociclib response panel. The coverage of this data by primary site is also presented in

Table 1.

2.2.4. UCLATORL Agilent microarray data

Transcript microarray analyses have been performed on 396 of the 416 cell lines in UCLATORL's palbociclib response panel. The distribution of these cell lines by primary site is presented in

Table 1.

Cells were grown to log phase in T75 culture flasks. RNA was extracted using Qiagen's RNeasy kit. Purified RNA was quantified using a Nanodrop Spectrophotometer, and quality was determined via capillary electrophoresis using the Agilent 2000 Bioanalyzer. Microarray hybridizations were performed using the Agilent Human 44K array.

Characterization of transcripts will be performed by comparison to a human mixed reference pool on a single chip where the reference cRNA was stained using cyanine-3 and the sample RNA was labeled with cyanine-5. The mixed reference pool will be primary site-specific, to minimize the signature of tissue-specific differential transcription and to allow for the identification of clustered subtypes within each primary site. Each mixed reference pool is constructed using a carefully selected subset of cell lines within each primary site.

Microarray slides were read using an Agilent Scanner. Agilent Feature Extraction software (version 7.5) was used to calculate gene expression values. These feature extracted files were subsequently imported into Rosetta Resolver software (version 7.1) for gene expression data analysis. The intensity ratios between the sample and mixed reference were calculated for each transcript based on the Agilent error model.

CHAPTER 3: Determination of *in vitro* response to palbociclib

3.1. UCLATORL proliferation assay protocol

Cell lines were grown and maintained as an adherent monolayer in plastic 75cm^2 flasks with growth media. During log phase growth, the adherent cells were harvested via trypsinization then homogenized and suspended in growth media. The concentration of suspended cells was determined by running a 100μL aliquot through a Beckman Coulter Z1 particle counter. The suspension was then diluted to a target concentration such that the desired number of cells could be plated out into each well of a 24 well tissue culture plate. The cells were given one day to adhere to the plastic wells and begin logarithmic growth. 24 hours after seeding, six control wells were harvested via trypsinization and counted on a particle counter. These counts were averaged and served as a baseline control count at time of treatment (Day0).

At this point, the remaining wells were treated with palbociclib starting at 1μM and reduced by 6 to 12 two-fold dilutions. Each concentration was treated in duplicate. The plates were allowed to grow under treatment for five days after which all wells were counted on the Z1 particle counter. Six wells were left untreated to serve as Day5 controls. The raw treated counts at each concentration were compared to Day0 and Day5 controls to generate dose response data. The entire protocol has been repeated for each cell line at least twice within four weeks of the initial assay to generate duplicate dose-response data for error estimation. Figure 2 describes the overall protocol.

3.2. Calculation of generational percent inhibition

Palbociclib's inhibitory effect on cell proliferation was calculated at each point on the dose-response curve. The most commonly used metric used on the y-axis of *in vitro* dose-response curves is

13

"percent inhibition". UCLATORL uses a novel method of calculating percent inhibition that differs from other large scale, *in vitro* screening efforts.

Percent inhibition was calculated as a function of the number of cell population doublings from baseline. Average "Day 0" control counts were quantified at time of treatment. This baseline cell count was used to calculate how many generations of logarithmic growth occurred over the 5-day treatment time at each dose using the exponential growth equation below:

$$Doublings\ from\ Baseline = \frac{\log(Avg\ Day5\ Count) - \log(Avg\ Day0\ Count)}{\log 2}$$

Percent Inhibition at each dose was then calculated using the following formula:

$$\%\ Inhibition = \left(1 - \frac{doublings\ under\ treatment}{doublings\ without\ treatment}\right) \times 100\%$$

3.3. Dose-response curve fitting

Curves were fit to the data points from each dose-response assay using the PROC NLIN function in SAS. The basic four-parameter sigmoid model below was implemented:

$$\%\ Inhibition = \left(\frac{\max - min}{1 + 10^{\frac{\log(ec50)-dose}{slope}}}\right) + min$$

Estimates for each of the starting parameters (max, min, slope and ec50) were chosen using a log-stepped grid search. The residuals were weighted inversely by the square root of generational

percent inhibition to reduce the pull of extreme high points on the more informative, lower region of the curve. The resulting parameters were exported and stored for later analysis. Similarly, the r-squared values were exported and stored as a measure of goodness of fit.

3.4. Summary outcome measures

Once curves are fit to the dose-response data sets, there are several response statistics that can be calculated in order to determine the relative sensitivity or resistance to treatment for each cell line. The most widely used of these response measures is the IC50: the concentration of drug that causes 50% inhibition. Since UCLATORL has a novel method of calculating percent inhibition, the IC50 has a unique interpretation for our assays: It is the concentration at which the drug prevents 50% of cell population doublings from occurring. Palbociclib IC50s were interpolated from the fitted dose-response curve. Average IC50s were calculated for each cell line as the geometric mean of individual IC50s from duplicate experiments.

Unlike traditional survival-fraction based percent inhibition, UCLATORL's "generational" percent inhibition is free to vary above the 100% threshold. In our experiments, inhibition exceeding 100% indicates that there was a decrease in the cell population count from baseline. The IC100 can be interpreted as the concentration of drug necessary to completely inhibit any new growth *in vitro* (i.e. 100% inhibition). Any concentration above this threshold value will result in a cross over into the region of the dose response curve where there is clear cell death from baseline. At these concentrations there is an implied cytotoxic effect of the drug rather than a purely inhibitory or cytostatic effect. It is important to note, however, that percent inhibition values below 100% do not necessarily imply a lack of cytotoxic effects in the drug's mechanism of action. IC100s were calculated via interpolation from the

dose-response curve. Average IC100s were calculated as the geometric mean of individual IC100s from duplicate experiments.

Some compounds are characterized by very flat dose-response curves. This can often be the case with irreversible, allosteric inhibitors as well as antibodies where the molecular target can be fully saturated over a large dose range. In these cases, the 'moving' portion of the curve is often at very low concentrations that may not be biologically or clinically relevant. It thus may be more appropriate to use a different summary outcome measure for these compounds. When flat curves can be expected based on drug mechanism, UCLATORL uses $I_{[x]}$ as a summary outcome measure. UCLATORL defines $I_{[x]}$ as the generational percent inhibition at a specified concentration of compound as interpolated from the fitted dose-response curve. Average $I_{[x]}$ is calculated as the linear mean of the individual $I_{[x]}$ values from duplicate experiments.

3.5. Palbociclib response distribution

We were able to generate reliable dose-response data for palbociclib on 416 cancer cell lines representing tumors derived from 12 distinct primary sites. There was highly differential response to treatment across the panel, with IC50 values ranging from 3nM to well above the highest dose tested (1μM). The interquartile range for the entire panel was found to be 143nM to 1μM. Figure 3 shows this distribution of IC50s across the entire panel. Here it can be seen that there are many lines where palbociclib is highly effective at inhibiting cell proliferation, even at very low doses. However, there are also many cell lines whose IC50 is above 1μM, indicating strong resistance to treatment.

Figure 4 shows the cell lines rank ordered by sensitivity to palbociclib using a different outcome measure: percent inhibition at 1μM. Using this measure we also observed wide variation in sensitivity to palbociclib across the panel. Palbociclib had very little to no effect on a subset of lines, even at this maximum palbociclib concentration. Those cell lines that had inhibition < 50% are by definition the same

lines whose IC50s were off scale in Figure 3. There was also a subset of lines whose maximum inhibitory values were above 100%, indicating that there was cell death from baseline at this dose of palbociclib. Both Figure 3 and Figure 4 show that there is a spectrum of response to palbociclib, rather than a simple dichotomy of sensitivity and resistance. This observation will be explored further in section 4.1.

Supplemental Table S1 is a complete description of palbociclib response data, averaged by cell line. Average (geometric mean) values for IC20, IC50, IC80 and IC100 are provided. Also provided is the average (linear mean) number of doublings in the control (untreated) cell population over the five day treatment period for each cell line.

IC50 box-and-whisker plots were generated to examine the differential sensitivity to palbociclib by primary site (Figure 5). The blue box represents the interquartile range of each panel. The top and bottom whiskers stretch to the maximum and minimum IC50 values, respectively. The left most plot represents the IC50 distribution of the entire 416-line panel. The rest of the plots represent the IC50 distribution for a particular cancer type (by primary site of origin). The plots are ranked by median IC50 from lowest (more sensitive) to highest (more resistant). There is clear differential sensitivity both between and within cancer types. While some types were more resistant than others on average, there were at least some cell lines in every primary site panel that were highly sensitive to palbociclib.

3.6. Comparison with Sanger and Broad Institute palbociclib response datasets

UCLATORL is not the only large scale, *in vitro* screening project to have worked with palbociclib. The Sanger Institute's Genomics of Drug Sensitivity in Cancer project[27] as well as the Broad Institute's Cancer Cell Line Encyclopedia (CCLE) project[26] have both assayed response to palbociclib in their own panels of cell lines. While the underlying concepts of the screening protocols are similar between UCLATORL, CCLE and Sanger, there are several methodological distinctions between the labs.

In contrast with our five-day proliferation experiments, Sanger and CCLE both perform 72-hour

assays. While UCLATORL performs direct cell counting on a Z1 particle counter, CCLE and Sanger

measure cell proliferation indirectly with fluorescent stains. Sanger stains with Syto60 (a nucleic acid

stain), while CCLE uses Cell Titer Glo (an indirect measurement of ATP). Both Sanger and CCLE plate

fewer cells in smaller wells than UCLATORL. Sanger plates in 96 or 384 well microplates at 15%

confluency, CCLE plates 250 cells per well in 1536-well plates. In contrast, UCLATORL plates 20,000 cells

per well in 24 well plates. Both Sanger and CCLE calculate percent inhibition using a 'survival fraction'

model where raw counts are compared to controls at the end of treatment only. In contrast, UCLATORL

uses the generational method described in section 3.2 which is based on population doublings from

baseline.

These variations in experimental protocols and analytical methods suggest that it may be useful

to compare the response distributions to palbociclib between labs to quantify how much outcome

misclassification there may be in these datasets. Figure 6 shows box-and-whiskers plots comparing the

IC50 distributions between the three screening projects. The three labs have similarly sized cell line

panels, where some but not all of the cell lines overlap. The left plot compares the full datasets for each

lab, while the right plot only shows IC50 data for overlapping cell lines. It is clear from the figure that

there were more sensitive lines and a broader range of activity in the UCLATORL dataset as compared to

Sanger and CCLE. In the Sanger dataset over half of the panel had IC50s above the highest dose tested

(4μM). In the CCLE dataset over 75% of the lines had IC50s above the highest dose tested (8μM).

Possible explanations for these discrepancies will be explored in the following discussion section.

Next, we performed a correlation analysis directly comparing matched cell lines between the

Sanger and CCLE datasets. There were 159 cell lines where IC50 values were available in both datasets.

Pearson correlation between log(IC50)s was performed using PROC CORR in SAS. For IC50s that were

above the highest dose tested, this maximum dose was used as the IC50 value. The correlation between

the matched log(IC50)s was weak, but significant (Pearson's r = 0.47537, p < 0.0001). A scatter plot comparing log(IC50) values between the Sanger and CCLE datasets is presented as Figure 7.

This correlation analysis may be skewed by the fact that such a high proportion of each panel had IC50s above the maximum dose, and therefore did not have a true quantitative value to use in the analysis. To account for this, a secondary Kappa test was performed to determine the measure of inter-rater agreement on a dichotomous measure of response to palbociclib. Of the 159 overlapping lines, the CCLE only reached IC50s for 25 lines. Thus, we rank ordered the panels on response by IC50 and classified the 25 most sensitive cell lines in each dataset as 'responders' (1) while the rest were classified as 'nonresponders' (0). The agreement between labs on identifying these 25 greatest 'responders' was very low (kappa = 0.16).

3.7. Discussion

We were able to demonstrate a broad differential response to palbociclib *in vitro*, both within and between cancer types. Every cancer type (except liver) had a subset of cell lines that were highly sensitive to palbociclib (IC50 < 100nM). This intra-type variation is encouraging towards the hope of identifying palbociclib -sensitive subpopulations of cancers within several cancer types.

The UCLATORL dataset had a much broader distribution of palbociclib IC50s as compared to similar datasets from Sanger and CCLE, even on the exact same cell lines that were assayed independently by each group. On the whole, UCLATORL was able to identify many more lines that were "sensitive" to palbociclib in clinically relevant dose ranges. This discrepancy in outcome classification between parallel screening efforts is disconcerting, and warrants a detailed discussion on the methodological and analytical differences that may have given rise to the inter-lab differences observed.

It is our belief that the major factor underlying this discrepancy is the difference in treatment time between labs. The 72-hour, fluorescence-based assays implemented by CCLE and Sanger were

originally optimized to observe the effect of *cytotoxic* agents on cell populations. Palbociclib, however, is

assumed to act predominantly through a *cytostatic* mechanism. In order to see this predominantly

cytostatic effect of palbociclib, cell populations must be treated for a long enough period of time such

that the cell population has time to grow significantly from baseline. Three days of treatment may not

allow for enough growth in the control wells to see a difference in the treated wells, especially for slow

growing cell lines.

The mathematical formulae used to calculate percent inhibition (the response variable in dose-

response data) may compound this problem in the CCLE and Sanger datasets. Both labs use a "survival

fraction" method to calculate percent inhibition, as opposed to the "generational" method employed by

UCLATORL. It is our belief that the "survival fraction" method likely underestimates the effect of

treatment on slow growing lines and overestimates the effect of treatment on fast growing lines. To

highlight this problem consider the following hypothetical examples.

Figure 8 compares the two methods using hypothetical response data at one dose of drug for a

slow-growing cell line. Figure 9 compares the two methods for a hypothetical fast-growing cell line. In

the slow-growing example, the dose of drug was enough to completely inhibit and new growth from

baseline. In the fast-growing example, the dose of drug had very little effect and allowed several rounds

of population doublings from baseline. The drug very clearly had a stronger cytostatic effect on the

slow-growing cell line; however the "survival fraction" method would calculate a 50% inhibitory effect

for both examples. In contrast, the generational method calculates an effect of 100% in the slow-

growing example and 25% in the fast-growing example. We believe these "generational" inhibition

values are much more indicative of the true cytostatic effect of the drug in each example.

Ultimately, we believe that the combination of short treatment windows and 'survival-fraction'-

based analytical methods has led to a severe underestimation of the cytostatic effect of palbociclib for a

large portion of the cell lines in the Sanger and CCLE response datasets. Almost none of the IC50s

observed *in vitro* by either lab below the clinically achievable plasma levels of palbociclib (~100nM).

Nearly a quarter of UCLATORL's cell lines were below this cutoff. We believe the UCLATORL response

dataset is therefore uniquely suited for the identification of palbociclib sensitivity biomarkers, an

assertion already partially borne out through the successful phase II clinical trial results based on our *in*

vitro breast cancer observations.

There was very little correlation between the Sanger and CCLE palbociclib IC50 datasets, either

as a continuous measure or by agreement on the most sensitive lines. This may be a consequence of the

aforementioned biases induced by their inhibition calculations, or may be due to random experimental

error. UCLATORL has incorporated many features into our proliferation protocols towards the end of

reducing experimental error in our response data. Our experiments are set up in 24 well plates as

opposed to the 96, 384 or 1536 well microtiter plates used in other screening efforts. The larger surface

area allows UCLATORL to seed more cells at baseline, lessening the chance for the accidental

enrichment of a specific subpopulation of cells. CCLE, in contrast, seeds only 250 cells/well in their

experiments. It is our belief that this low seeding number could lead to clonal selection effects that may

influence outcome.

UCLATORL also implements direct particle counting to assess cell number, rather than

measuring cell proliferation indirectly with fluorescent stains. The need to interpolate cell counts from a

fluorescent standard can cause error in these estimates. The Cell Titer-Glo assay employed by CCLE uses

a luciferase reaction to indirectly measure ATP levels in cell cultures, and then generalizes ATP levels to

cell count. However it is well known that ATP levels per cell can vary considerably from cell line to cell

line[70,71]. This imprecise relationship between cell number and fluorescence can cause considerable noise

in inter-cell line comparisons.

CHAPTER 4: Association testing methodology

4.1. Dichotomous classification of *in vitro* response to palbociclib:

Following the fitting of dose-response curves and interpolation of summary outcome measures, response to treatment with palbociclib was then dichotomized using cutoffs in the continuous summary outcome measures (1 - "sensitive", 0 - "resistant"). This decision to treat outcome as dichotomous, rather than continuous, is another distinction between UCLATORL's analytical approach as compared to other large screening efforts. We believe that dichotomizing response is mathematically appropriate and has interpretational benefits.

In a clinical setting, palbociclib is prescribed at one specified dose. Any differential effects above or below this dose cannot be clinically observed. In our *in vitro* work, IC50s at either very high or low extremes are often observed and could exert a strong pull on a continuous analysis in one direction, even though these concentrations are likely to be clinically irrelevant. It is important to note, however, that clinical concentrations of drug may not be directly generalizable to the *in vitro* setting. This problem will be discussed further in section 4.4.

Another major benefit of a categorical analysis is that it makes it unnecessary to extrapolate summary outcome measures (such as IC50 or IC100 values) when they lie outside the tested concentration range. This is often the case for resistant cell lines. Over 25% of our cell lines had palbociclib IC50s above the tested concentration range. CCLE and Sanger had even higher proportions of these highly resistant lines. We believe that inputting an extrapolated IC50 value for these lines into a regression model is statistically inappropriate. These very high, extrapolated IC50s may artificially increase the significance of observed associations, even though these concentrations may be well above a clinically meaningful dose.

The relative risks (response ratios) that result from the dichotomized response analysis are also more easily interpreted into clinical analogy than the coefficients from a continuous response analysis. These response ratios relate whether a subpopulation of cancers with a particular genetic alteration will have a higher proportion of responders at a set dose. A linear coefficient relating the difference in the average IC50s of the two populations may be less useful, both intuitively and practically.

4.2. Modified Poisson regression analysis

The *in vitro* proliferation experiments conducted by UCLATORL can be modeled like a prospective cohort study where the outcome (response to treatment) is common and person-time is the same for all subjects. To quantify the association between individual biomarkers and a dichotomized response variable, UCLATORL uses a modified Poisson regression model with robust error variance. This method was first introduced in a 2004 paper by Zou[72], and recommended by Greenland[73] later that year. Advantages of this approach include better estimation of confidence intervals and lower likelihood of convergence problems as compared to other available models.

This analysis was implemented using the PROC GENMOD feature in SAS using a Poisson distribution, a log link function and with the 'repeated' statement added. The model can accommodate categorical predictors such as mutation or copy number status, as well as continuous variables such as cell line growth rate or gene expression level. This model was used in all association studies whose results are presented in later sections.

The primary output from this analysis is a relative risk estimate for each of the independent variables in the model. In the context of UCLATORL's genotype-response association studies, these relative risk estimates will be referred to as "response ratios" and can be interpreted as the average fold increase in likelihood to be classified as sensitive to treatment for cell lines that have a particular genetic alteration compared to those that do not.

Many of the genomic alterations used as independent predictors in these regression models are relatively rare in UCLATORL's cell line panel. Over the course of the study, there were several occasions where all cell lines with a particular alteration were classified as resistant to palbociclib, corresponding to a crude response ratio of zero. Under these conditions, the Poisson regression model run by PROC GENMOD failed to converge. When this occurred, a dummy entry was added to the association testing dataset coded as mutant (1) and sensitive (1). This entry was down-weighted relative to the true data in proportion to the frequency of the alteration in the cell line population.

4.3. Controlling the false discovery rate

The large number of tests being performed increases the likelihood of false positive predictors falling out of the analysis. To address this concern UCLATORL will control the false discovery rate using the two-stage linear step up model introduced by Benjamini, Krieger, and Yekutieli in 2006[74]. These adjustments will be performed separately for each genomic dataset that is interrogated for associations with palbociclib response using the PROC MULTTEST function in SAS.

4.4. Sensitivity analysis

Section 4.1 outlined the justification for classifying response to palbociclib as a dichotomous, rather than continuous, variable. However, there are many non-generalizable experimental conditions that differ between *in vitro* proliferation assays and *in vivo* pharmacodynamics. There will necessarily be some level of arbitrariness in the choice of the palbociclib dose cutoff used to categorize response. To account for this, all crude association tests will be performed using three separate cutoffs for palbociclib sensitivity.

The cutoffs used in association testing was informed by the range of dose-response summary outcome measures observed, as well as known molecular characteristics of the compound. Palbociclib has been shown to have on-target inhibitory effects at concentrations above 15nM[56]. No known off-

target effects exist at concentrations < 2μM. A cell proliferation dose-response curve based on target-specific effects can be expected to exist within this range. The slope, ec50 and maximum of this curve are expected to be variable between cell lines and depend on cytoplasmic/nuclear drug availability, enzyme kinetics and relative dependence on the target.

A phase I, dose-escalation study of PD 0332991 has shown that the compound is tolerable at moderate to high doses in patients, and has a relatively long half-life in serum. The recommended phase II dose of PD 0332991 was concluded to be 125mg orally once a day. Pharmacodynamic analysis demonstrated that this resulted in sustainable serum levels >50ng/mL (>112nM) with maximum concentrations reaching significantly higher levels.

The first cutoff used in the analysis was the most conservative, and was primarily based on these clinical pharmacodynamic observations. Cell lines were classified as sensitive to palbociclib if they had an average *in vitro* IC50 above 100nM. Those cell lines with average IC50s above 100nM were classified as resistant. The second and third cutoffs were informed by the IC50 distribution across the entire panel. Cutoff 2 was set at an IC50 above or below the lower quartile (144nM). Cutoff 3 was set at an IC50 above or below the median (365nM).

CHAPTER 5: Somatic point mutations versus *in vitro* palbociclib sensitivity

Once response to treatment with palbociclib was classified and a regression model was established, we began to interrogate our genomic characterization datasets with the intention of identifying candidate genetic alterations that predict for palbociclib sensitivity *in vitro*. The first genotype-response association screen was performed using the CCLE's whole-exome, hybrid-capture sequencing data. The intention of this analysis was to identify somatic, acquired point mutations in genes that code for proteins that may play a functional role in palbociclib sensitivity.

5.1. Restriction of point mutation dataset

Genomic instability is a hallmark of most cancers. Cancer genomes are usually highly complex with many acquired alterations existing on top of the normal genetic background of the patient. The genetic alterations that accumulate during oncogenesis may be broadly separated into two groups: "driver" alterations that play some causal role in cancer, versus neutral "passenger" alterations that have neither a strong oncogenic effect nor are actively selected against. When attempting to identify acquired alterations that may play a functional role in response to treatment, it is important to take steps to enrich the predictor dataset for "driver" mutations where possible. Otherwise, there is the risk of identifying "passenger" mutations as false positives or missing true "driver" effects due to overly stringent multiple testing corrections.

Recognizing this issue, the Cancer Cell Line Encyclopedia has pre-restricted its whole-exome sequencing dataset for several inclusion criteria with the intention of enriching for "driver" alterations[26]. First, they list the following criteria for genes to be included in the dataset: "(1) genes identified as somatically altered in cancer based on (A) occurrence in at least 4 instances, collectively, from recently

published literature, (B) occurrence in 2 – 3 instances from the aforementioned studies and also present in a significantly amplified or deleted focal peak in primary tumors or cell lines (this study), (C) membership in the Cancer Gene Census, or (D) significant mutation frequency across 441 tumors; (2) genes identified in either the literature, or meeting abstracts and presentations, as putative oncogenes, tumor suppressors, members of cancer related pathways, or having a cancer-related function(s); or (3) protein kinases."

The CCLE also had several post-sequencing exclusion criteria for the types of alterations that are included in their final dataset. The following variants were excluded by CCLE: (1) variants with allelic fractions lower than 0.25, (2) common germline variants (>0.1%) listed in dbSNP134, (3) putative neutral variants, and (4) variants observed in a panel of 278 whole 'normal' exomes. CCLE claims that these steps help to eliminate "…common false positives that originate predominantly from alignment artifacts" and enrich the dataset for putatively oncogenic functional mutations. At the time of integration into UCLATORL's internal database, this list from the CCLE contained sequencing information on 4059 genes. Of these, alterations in 2380 genes were found to be present in at least one cell line in UCLATORL's panel.

This number of genes was still far too large to be used for association testing considering the relatively small number of cell lines available for the genotype-response screen. Therefore, UCLATORL had its own internal further restriction criteria for inclusion in our final dataset to be used for association testing. We wished to limit the dataset to only those genes where a strong causal role in cancer has been established. To this end, genes were included in our restriction dataset only for those genes listed in the Sanger Institute's COSMIC gene census[75]. The COSMIC gene census is a continuously updated list of cancer genes where the literature has suggested a causal implication in oncogenesis. The gene census lists the following inclusion criteria for point mutations: "Genes have been included in the census if there exist at least two independent reports showing mutations in primary patient material. In

considering somatic mutations for inclusion in the census, we included only genes for which there was evidence of the somatic origin of at least a subset of mutations, based on analysis of normal tissue from the same individuals." To date, the gene census includes somatic mutations in 452 genes. Alterations in many of these 452 genes were either absent in our cell line panel, or were present at such a low frequency that we did not have sufficient power to test for associations (< 6 out of 240 cell lines). This further limited our list to just 84 genes.

The COSMIC gene census also distinguishes between dominant, likely activating point mutations in oncogenes versus recessive, likely loss-of-function mutations in tumor suppressor genes. We separated these two functionally distinct groups into separate datasets where the false discovery rate was controlled independently for each gene set. One dataset consisted of 27 oncogenes where likely activating mutations existed in our cell line panel. The other consisted of 57 recessive cancer genes where likely loss-of-function point mutations existed. These were the final datasets used for association testing. The full gene sets are presented in Supplemental Table S2.

5.2. Results

5.2.1. Dominant, likely-activating point mutations

The 27-gene, dominant, likely-activating point mutation dataset was interrogated for associations with palbociclib response using the modified Poisson regression model outlined in section 4.2. Alterations were grouped by gene and classified as a dichotomous variable: 1 = mutant, 0 = wild type. Each gene was run independently in the regression model to obtain a crude response ratio. Three separate association screens were performed using the three dichotomous palbociclib response variables generated using the IC50 cutoffs described in section 4.4.

The outcome statistics for each of the three screens performed were visualized using volcano plots. Figure 10 shows the resulting volcano plot for the genotype-response screen using a cutoff for

sensitivity where palbociclib IC50 = 100nM. Each circle on the volcano plot represents a separate regression analysis for a particular gene's association with palbociclib response. The x-axis plots the effect size (response ratio) of the association. The further to the left of the plot a circle lies, the more the gene predicted for sensitivity to palbociclib whereas the right-most circle represent genes that predicted for resistance. The y-axis plots the p-value on a reverse log scale, such that the more significant the association the higher the circle falls on the y-axis. The relative size of the circles represents the frequency of the alteration in our panel.

After the crude response ratios, confidence-intervals and p-values were calculated, the list of significant associations was restricted via controlling the false discovery rate using the method described in section 4.3. The false discovery rate was controlled separately in each of the three screens representing the three palbociclib sensitivity cutoffs. In order to be classified as a significant predictor, the gene must have been identified as significant (post multiple test correction) in at least one of the three screens and must have predicted in a consistent direction using all three cutoffs. This limited our results to just three significant associations, all of which predicted in the direction of resistance. Table 2 relates the response ratios, confidence intervals and p-values for the significant associations at each of the three cutoffs used for palbociclib response classification.

The first of these genes, MLL, encodes for a histone-methyltransferase protein called "Mixed-Lineage Leukemia" that plays a critical role in early embryonic development. This gene was found to be mutated in 14 of the 240 lines in UCLATORL's panel where point mutation data was available. It was highly correlated with resistance to palbociclib at the 100nM IC50 cutoff. It also predicted in the resistant direction using the other two response cutoffs, but did not reach statistical significance. The causal role MLL may play in determining palbociclib response will be discussed in detail in Chapter 8.

Alterations in the TSHR gene were also found to be associated with *in vitro* resistance to palbociclib. The TSHR gene encodes for thyroid stimulating hormone receptor, a membrane protein that

controls thyroid cell metabolism. This gene was found to be mutated in 10 of the 240 lines in

UCLATORL's panel where point mutation data was available. Again, alterations in this gene were found

to be highly associated with resistance to palbociclib at the 100nM IC50 cutoff, but less significantly at

the other two cutoffs. The causal role TSHR may play in determining palbociclib response will be

discussed in detail in Chapter 8.

The SMO gene also fell out of our screen as a candidate resistance biomarker. The SMO gene

codes for the Smoothened protein; a G-protein coupled receptor that interacts with patched (Ptc)

protein as part of the Hedgehog (Hh) signaling pathway. This gene was found to be mutated in 9 of the

240 lines in UCLATORL's panel where point mutation data was available. Alterations in SMO were

strongly associated with resistance to palbociclib at both Cutoff 1 (IC50 = 100nM) and Cutoff 2 (IC50 =

Q3), but were less significantly associated at Cutoff 3 (IC50 = Median). The causal role SMO may play in

determining palbociclib response will be discussed in detail in Chapter 8.

5.2.2. Recessive, likely-loss-of-function point mutations

Next we tested the 57-gene, loss-of-function, point mutation dataset for associations with

palbociclib response *in vitro*. The same methods were used as for the activating mutations. One extra

complicating factor in the loss-of-function dataset is that zygosity is more of a concern. For activating

mutations, one altered allele is enough to grant the oncogenic phenotype. In contrast, loss-of-function

mutations typically require all alleles to be lost, as one functional allele is often enough to maintain a

normal phenotype.

The point mutation data provided by CCLE does not contain definitive calls on zygosity, but

rather provides "read counts" for normal versus alternative alleles. In a typical Mendelian setting with a

homogenous cell population, the alternative read counts would be 0% for wild-type, 50% for

heterozygous mutations and 100% for homozygous mutations. However, the CCLE read count data

varies continuously between these discrete, Mendelian values. This deviation from the Mendelian

expectation can reflect several scenarios.

First, many cancer cell lines are polyploid, meaning they have chromosomal duplications of large

portions of their genome. Having more than two alleles at certain loci can lead to non-Mendelian

mutant fractions. Second, the cell population that was genotyped may have not been completely

homogenous for a given genotype. While the phenomenon is rare, cultured cell lines may contain

subpopulations of cells that differ in their mutation status for a given gene. Third, there are common

epigenetic means of silencing the wild-type allele in a cell line that harbors a heterozygous mutation,

leading to a homozygous phenotype. Finally, there is a certain amount of "noise" in any chip-based

genotyping platform due to mechanical error in the assay. Considering these complicating factors we

chose to classify any cell line where CCLE discovered a likely functional alteration as mutated,

independent of read count.

After association testing was performed and the false discovery rate was properly controlled, we

found that loss-of-function alterations in two genes (CDH1, TOPBP1) were significantly associated with

sensitivity to palbociclib whereas loss-of-function alterations in three genes (FANCA, RB1, NBN) were

significantly associated with resistance. The response ratios, confidence intervals and p-values for these

associations at each of the three dichotomous response cutoffs are presented in Table 2. A

representative volcano plot for all association tests performed at dichotomous response cutoff 1 is

presented as Figure 11.

CDH1 encodes a calcium-dependent cell adhesion protein, E-Cadherin. Presumptive loss-of-

function alterations in CDH1 were present in 13 cell lines out of the 240 that were genotyped. CDH1

mutant cell lines were significantly more likely to be classified as sensitive to palbociclib at all three

dichotomous response cutoffs, with the strongest association occurring at the IC50=100nM cutoff. The

mechanistic role that alterations in CDH1 may play in conferring palbociclib sensitivity will be discussed in Chapter 8.

The TOPBP1 gene codes for Topoisomerase II Binding Protein 1. This protein interacts with the C-terminal region of topoisomerase II beta and is necessary for DNA replication. Presumptive loss-of-function alterations in TOPBP1 were found in 8 cell lines out of the 240 that were genotyped. Alterations in TOPBP1 predicted in the direction of sensitivity for all three dichotomous response cutoffs, but only reached statistical significance at cutoff 3 (IC50 < Median). The mechanistic role that alterations in TOPBP1 may play in conferring palbociclib sensitivity will be discussed in Chapter 8.

FANCA codes for a DNA repair protein called fanconi anemia, complementation group A. Presumptive loss-of-function point mutations in FANCA were present in 12 cell lines of the 240 that were genotyped. Alterations were significantly associated with resistance to palbociclib at two of the three dichotomous response cutoffs used in the analysis. It also predicted in the resistant direction for the third cutoff (IC50 < Median), but did not reach statistical significance. The mechanistic role that alterations in FANCA may play in conferring palbociclib sensitivity will be discussed in Chapter 8.

NBN encodes a protein called nibrin that is involved in several cellular processes including: DNA damage signaling, double-strand break repair, telomere maintenance and cell-cycle checkpoint control. Presumptive loss-of-function mutations in NBN were present in 9 cell lines of the 240 that were genotyped. Alterations were significantly associated with resistance to palbociclib at two of the three dichotomous response cutoffs used in the analysis. It also predicted in the resistant direction for the third cutoff (IC50 < Median), but did not reach statistical significance. The mechanistic role that alterations in NBN may play in conferring palbociclib sensitivity will be discussed in Chapter 8.

RB1 encodes retinoblastoma 1, a protein that is directly downstream of the target of palbociclib inhibition, CDK4/6. It is a tumor suppressor that negatively regulates the G1/S transition. Presumptive loss-of-function alterations in RB1 were present in 12 cell lines of the 240 that were genotyped.

Alterations predicted in the direction of resistance for all three dichotomous response cutoffs, but only

reached statistical significance at cutoff 1 (IC50 = 100nM). The mechanistic role that alterations in RB1

may play in conferring palbociclib sensitivity will be explored further in Chapter 8.

CHAPTER 6: Copy number alterations versus *in vitro* palbociclib sensitivity

Duplications and/or losses of large parts of chromosomal DNA are a very common type of genomic alteration during oncogenic transformation. UCLATORL has assayed our cell line panel for copy number alterations via high-throughput comparative genomic hybridization. Copy number alterations resulting in the amplification of oncogenes or the deletion of tumor suppressor genes were interrogated for their association with *in vitro* response to palbociclib.

6.1. Restriction of copy number alteration dataset

The comparative genomic hybridization dataset generated by UCLATORL provides genome-wide copy number data. As was the case for point mutations, many of the alterations discovered may be "passenger" mutations that are unlikely to play a causal role in determining response to treatment with palbociclib. Leaving these likely-neutral alterations in our dataset could either increase the risk for false positives, or increase the chance of missing a true effect due to overly stringent multiple testing corrections. This restriction process is philosophically similar to what has been discussed for the point mutation dataset; however there are many logistical considerations that are unique to the copy number alteration dataset.

Genomic amplification events often create amplicons that often encompass the coding region of many genes. Many of these amplicons may be neutral, passenger alterations that do not contain an oncogenic target gene. Others may include multiple candidate oncogenes, making it difficult to identify a single driver. In 2010, Santarius et al[76] published a review in Nature that outlined a set of criteria that may be used to identify genes where amplification is likely to be causally linked to cancer. UCLATORL

used the same "weight-of-evidence" selection criteria to create the gene set used for association testing.

First, amplification was defined as having a $\log_2$(ratio) > 1 for at least two adjacent probes on the CGH chip platform. This corresponds to a gain of at least twice as many copies of the amplified DNA as compared to centromeric DNA on the same chromosome. If a chromosomal region was amplified in several distinct cell lines, the amplicons were mapped and cross-compared to establish a minimal region of amplification (MRA). Genes that were within this MRA were considered more likely to be oncogenic drivers and up-weighted for inclusion into the final dataset. Genes were excluded if they were amplified in less than six of UCLATORL's cell lines.

Another criterion that was used to justify inclusion was whether the amplification event is known to lead to overexpression of the mRNA transcripts from genes within the amplicon. This was determined either from historical observations in the literature, of from UCLATORL's own mRNA expression microarray data.

Several other criteria from the Santarius[76] paper were also utilized in the restriction process including: correlation of amplification with clinical outcome data, biological investigations of function as well as drug response data targeting an amplified gene product. Ultimately, 107 genes from 26 distinct amplicons were included in the final dataset used for association testing.

Chromosomal deletion events are often smaller in size than amplifications, making it easier to determine the deleted gene most likely to be causally implicated in oncogenesis. However, there are many deletions that occur in known fragile sites, where the deletion event is likely non-oncogenic. We used criteria similar to those outlined in another 2010 Nature article by Bignell et al[77] to restrict our chromosomal deletion dataset.

First, the gene must be completely (homozygous) deleted rather than partially (hemizygous) deleted in each cell line in order to be classified as mutant in our association screen. We used a cutoff of

log$_2$(ratio) < -2 from at least 2 adjacent probes to identify these homozygous deletions. Next, deleted regions were investigated for overlap with known recessive cancer genes. Using these criteria, we identified 17 genes from 14 distinct deleted regions for inclusion in our association screen.

6.2. Results

6.2.1. Chromosomal amplification of oncogenes

Our 107-gene, chromosomal amplification dataset was interrogated for associations with palbociclib response using the modified Poisson regression model described in section 4.2. Amplifications were grouped by gene classified as a dichotomous variable: amplified (1) versus non-amplified (0). Each gene was run independently in the regression model to obtain a crude response ratio. Again, three separate association screens were performed using the three dichotomous palbociclib response variables generated using the IC50 cutoffs described in section 4.4.

The volcano plot in Figure 12 shows the outcome measures for all association tests performed at dichotomous response cutoff 1. The large number of significant associations reflects the fact that significantly predictive amplicons often contain several co-amplified candidate oncogenes whose amplification status are strongly correlated with one another. The strong correlation of co-amplified genes has several consequences on our subsequent analyses.

The false discovery control methods used for the point mutation dataset assumed independence of the predictor variables for which crude effect estimates were calculated. This is clearly not the case for the copy number amplification dataset. Using the same method would likely result in too stringent corrections for multiple testing. Therefore we chose to base our multiple test corrections on the number of distinct amplicons tested (N = 26) rather than the number of candidate genes (N = 107).

After controlling the false discovery rate by this method, we found genes on three amplicons (17q12-q21, 11q13, 1q32) to be significantly associated with sensitivity to palbociclib and genes on three amplicons (19q13, 8q13, 5q33) to be associated with resistance. Table 3 relates the response ratios, confidence intervals and p-values for the significant associations at each of the three palbociclib response cutoffs used. In most cases, several genes on each amplicon were found to be significantly associated with palbociclib response. For each amplicon, we attempted to identify a candidate gene that was most likely to casually underlay the association with palbociclib response. This rationale is discussed at length in Chapter 8.

6.2.2. Homozygous deletion of tumor suppressor genes

Our 17-gene, chromosomal deletion dataset was interrogated for associations with palbociclib response using the modified Poisson regression model described in section 4.2. Homozygous deletions were grouped by gene classified as a dichotomous variable: amplified (1) versus non-amplified (0). Each gene was run independently in the regression model to obtain a crude response ratio. Again, three separate association screens were performed using the three dichotomous palbociclib response cutoffs generated using the IC50 cutoffs described in section 4.4.

Figure 13 shows a volcano plot representing the outcome statistics for each of the association tests performed using dichotomous response cutoff 3 (IC50 < Median). Most genes in the dataset were the only candidate recessive cancer gene discovered in the deleted region. Therefore, each gene was treated as an independent predictor and the false discovery rate was controlled using the same method as for the point mutations dataset.

After controlling for the false discovery rate, only one deleted region was associated with response to palbociclib *in vitro*. Small, focused deletions at 13q14 strongly predicted for sensitivity to palbociclib at all three dichotomous response cutoffs. One candidate recessive cancer gene was

identified on this amplicon: RB1. The mechanism describing the role of RB1 loss in determining

palbociclib response will be discussed further in Chapter 8.

CHAPTER 7: Investigation of candidate genomic predictors proposed in existing literature

Several clinical and preclinical analyses of palbociclib and other functionally similar compounds have been performed. Many more are currently ongoing. Many of these studies have either postulated or identified several candidate biomarkers that may have predictive utility in determining response to CDK4/6 inhibition. We wished to test these candidate biomarkers from existing literature for their performance in predicting response to palbociclib in our *in vitro* response dataset.

7.1. Selection of previously implicated candidate genomic predictors

We performed an extensive literature search to identify candidate genomic biomarkers that have been proposed as potential mediators of response to CDK4/6 inhibition. We focused our search to include genomic alterations that have been mechanistically implicated as biomarkers in preclinical models and/or that have been used as selection criteria for clinical trials. Table 4 relates a list of the biomarkers we chose to include in our association screen based on our literature search. All of the alterations on this list were included as part of the semi-supervised screens performed in sections 5 and 6, however many did not fall out of the analysis as independent predictors of palbociclib response. We felt it was appropriate to re-run a regression on these candidate biomarkers with more lenient multiple testing corrections, as each has a strong mechanistic rationale for a causal role in determining palbociclib sensitivity.

7.1.1. CDK4 amplification

The CDK4 gene codes for a Serine/Threonine kinase that phosphorylates and inhibits RB1, allowing for the dissociation of E2F from the RB/E2F complex. This ultimately allows for the transcription of E2F target genes that are needed for cell-cycle progression through the G1/S transition.

CDK4 amplification is known to occur as a somatic mutation in several cancer types, including gliomas[78], certain types of sarcomas[79], breast[80], esophageal[81] and pancreatic[82] cancers. It has been shown to be a prognostic biomarker of increases cell proliferation and poor outcome in several cancer types[80,82].

CDK4 is part of the protein complex that serves as the molecular target of palbociclib. Cell lines whose proliferation is at least partly driven by CDK4 amplification may be dependent on this oncogenic effect, and therefore may be more sensitive to inhibition with a CDK4 inhibitor such as palbociclib. However, to our knowledge there have been no empirical observations in preclinical testing that have validated this hypothesis.

Despite the relative lack of preclinical evidence, several clinical trials are currently recruiting in populations that are enriched for CDK4 amplified patients. The Memorial Sloan-Kettering Cancer Center is running Phase II trial in advanced or metastatic liposarcomas[83] where CDK4 alterations are common. The Abramsom Cancer Center is running a phase II trial where CDK4 amplified melanoma lines are listed in the inclusion criteria[84].

7.1.2. CDK6 amplification

The CDK6 gene also codes for a Serine/Threonine kinase that is part of the same complex as CDK4. It also interacts with cyclin D and Rb to stimulate cell-cycle progression. CDK6 amplification has been shown to occur somatically in T-cell lymphomas[85], gliomas[86], esophageal adenocarcinomas[81] as well as myxofibrosarcomas[87]. It has been shown to be a negative prognostic biomarker in several cancer types[81,87].

CDK6 amplification may play a similar causal role as CDK4 in determining response to palbociclib. It is part of the target complex of palbociclib inhibition. As with CDK4, amplification may

increase the cellular dependence on cyclin D - CDK4/6 function and therefore sensitize cells to palbociclib inhibition.

However, to our knowledge no preclinical studies have demonstrated an empirical association between CDK6 amplification and sensitivity to CDK4/6 inhibitors. Nor have any clinical trials have yet used CDK6 amplification status as a selection criterion for inclusion. This may be due to the rarity of the alteration in human cancers. However, we believe that because it is part of the target complex there is a strong enough causal link to justify including in the candidate analysis.

7.1.3. CCND1 amplification

The CCND1 gene codes for cyclin D1, which is also part of the palbociclib target complex. CCND1 amplifications are quite common in several cancer types including breast[88], ovarian[89], colorectal[37], head and neck[90], esophageal[91] and gastric[91] cancers. CCND1 has been shown to be a prognostic biomarker[36] as well as a predictor of response to chemotherapy[92] in certain cancers.

Similar to the rationale for CDK4/6, CCND1 amplification may drive proliferation and therefore increase cellular dependence on cyclin D – CDK4/6 function. CCND1 amplification was one of the biomarkers identified in the semi-supervised screen performed in section 6. To our knowledge, until this study no preclinical analysis had demonstrated this association empirically. The Abramsom Cancer Center is running a phase II trial where an inclusion criterion is CCND1 amplified patients from several cancer types including: esophageal cancers, head and neck cancers, breast cancers and liposarcomas[84].

7.1.4. CDKN2A homozygous deletions

The CDKN2A gene codes for the p16 (INK4A) protein. It acts as a tumor suppressor protein that induces cell cycle arrest at G0/G1 by inhibiting the CDK4 kinase. It is one of the most commonly deleted

genes in human cancers[77,93,94] and has been shown to be a negative prognostic biomarker in many cancer types[95,96].

Deletion of CDKN2A has been hypothesized to play a causal role in determining response to CDK4/6 inhibitors. CDKN2A deletion relieves the repressing effect on CDK4 kinase, effectively activating CDK4/6-driven cell proliferation. Loss (or low expression) of p16 has been shown to be associated with sensitivity to palbociclib *in vitro* within panels of ovarian[58] and glioblastoma[61] cell lines. This year, a palbociclib clinical trial was opened by the Jonsson Comprehensive Cancer Center for ovarian cancer patients, using low p16 expression as a selection criterion[97].

7.1.5. RB1 homozygous deletions and loss-of-function point mutations

RB1 codes for the retinoblastoma (Rb) protein. It acts as a negative regulator of the cell cycle via repression of E2F1. Unbound E2F acts as a transcriptional activator on many genes necessary for cell cycle progression. Loss of RB1 expression via homozygous deletion occurs in several cancer types[77]. Loss via deleterious point mutations has also been observed[98].

Loss of Rb protein via deletion or point mutation has been hypothesized to play a crucial role in determining response to CDK4/6 inhibition. Palbociclib was designed to prevent phosphorylation of Rb by CDK4/6, and therefore enhance its role as a tumor suppressor. If Rb is missing from cancer cells, it may induce resistance to CDK4/6 treatment as the necessary downstream effector is missing.

Both homozygous deletions in RB1 as well as loss-of-function point mutations were identified as independently associated with resistance to palbociclib in our semi-supervised screens. This year, a palbociclib clinical trial was opened by the Jonsson Comprehensive Cancer Center for ovarian cancer patients, using Rb proficiency as a selection criterion[97].

7.2. Results

All six candidate alterations discussed in this section were listed in the COSMIC gene census, and therefore included in our genomic alteration datasets. Of the six, alterations of RB1 and CCND1 had already fallen out of our semi-supervised screen as significant resistance and sensitivity biomarkers, respectively. The other three candidate biomarkers were also tested for associations as part of the previous screens, but did not fall out as significant. The response rations, confidence intervals and p-values for each of these candidate biomarkers are given in Table 4.

Amplification of CDK4 and CDK6 did not significantly predict for response to palbociclib in either direction at any of the three response cutoffs, even when multiple test corrections were not applied. It should be noted that these alterations were very rare in our panel. A larger number of mutant lines may be necessary to observe an association if it exists.

Loss of p16 via homozygous deletion of the CDKN2A gene did not significantly predict for sensitivity or resistance at the lower two cutoffs. It did, however, have a small but significant association with palbociclib sensitivity at response cutoff 3 (IC50 = Median). We ultimately chose not to include CDKN2A deletions in our final predictor cassette. This rationale will be discussed at length in Chapter 9.

CHAPTER 8: Creation of final predictor set and confounder control

8.1. Restriction criteria for inclusion in final predictor set

The genotype-response association screens discussed in sections 5 and 6 yielded fourteen candidate biomarkers that were associated with either sensitivity or resistance to *in vitro* treatment with palbociclib. Several measures were undertaken to minimize the occurrence of false-positive associations, including careful restriction of the genomic characterization datasets as well as post-hoc control of the false discovery rate. However, there is still a reasonable concern that several of these associations may have arisen by chance or alternatively by a strong association with one of the other independent biomarkers. We wished to restrict our list of candidate biomarkers to a "final predictor set" where each biomarker had been verified to have a strong mechanistic rationale for the association with palbociclib response, and that this association was independent of the other biomarkers in the set.

A literature search was performed on each prospective biomarker to build a mechanistic rationale for its inclusion in the final predictor set. Several criteria were used to strengthen the case for causality and potential for clinical translatability. First, the distribution of each candidate biomarker by primary site in our cell line was compared to the distribution in patient samples from the literature. Certain genetic alterations can occur de novo during the establishment of a cell line, and did not exist in the primary tumor from which it was derived. However, for any particular alteration it is impossible to know this for certain without original tumor material for comparison. Thus, the only validation that could be reasonably performed at this step is to confirm the existence of the alteration in patient samples for each cancer type where the alteration was found in our cell line panel. We primarily used the comprehensive patient sample genotyping information made public by the Cancer Genome Atlas project[99].

Next, we carefully investigated each candidate biomarker for correlations with the other biomarkers that were identified as crudely associated with palbociclib response in our screen. Though oncogenic somatic alterations are "random" events, clonal selection pressures in early development of a tumor can increase or decrease the likelihood of co-occurrence of certain alterations. These inter-correlations between genetic alterations in nature may have led to the identification of a non-causal biomarker through a strong correlation with the "true" driver of response to treatment. To this end, a Pearson correlation table was created for all fourteen candidate biomarkers that were identified in our screen. These results are presented in Table 6 and will be discussed independently for each biomarker in Chapter 8. When a potentially confounding correlation was present, a multiple regression model was constructed to isolate the independent effect of the biomarker or interest on palbociclib response. The same modified-Poisson model was implemented in these analyses, using a dichotomous response variable at the cutoff where the strongest crude effect was observed.

Finally, existing literature was consulted to establish a mechanistic rationale for the role each alteration may play in determining response to palbociclib. Generally, a case was made for each sensitivity biomarker to describe how the alteration may increase the cancer cell's dependence on cyclin D - CDK4/6 activity. Conversely, each resistance biomarker was investigated for a role in subverting or lessening cellular dependence on cyclin D - CDK4/6 activity.

8.2. Palbociclib response biomarkers included in "final predictor set"

8.2.1. 11q13 (CCND1): chromosomal amplifications associate with palbociclib sensitivity

Chromosomal amplification of the region mapping to 11q13 is a common feature of many cancer types including breast[88], ovarian[89], colorectal[37], head and neck[90], esophageal[91] and gastric[91]

cancers. Several candidate oncogenes on 11q13 have been identified including CCND1, ORAVO1, EMS1, FGF3/4/19, RPS6KB2, and PAK1 amongst others[100-103]. Of these candidates, CCND1 has accumulated the most compelling evidence for being the oncogenic 'driver' on the amplicon. CCND1 amplification has been shown to be a prognostic biomarker[36] as well as a predictor of response to chemotherapy[92] in certain cancers.

CCND1 amplification was the most common sensitivity biomarker to fall out of our semi-supervised genotype-response association screen. Amplification was discovered in 38 out of 317 genotyped lines in our panel. Amplification events were present in all but three (colon, liver, pancreas) of the ten cancer types that have been genotyped. Table 5 shows the distribution of these alterations by primary site, compared with the same distribution in the clinical samples genotyped by TCGA. The distribution of CCND1 amplifications was very similar between cell lines and patient samples.

CCND1 amplification was found to be significantly associated with sensitivity to palbociclib at two of the three response cutoffs used in our screen (Table 3). There was a small correlation between ERBB2 amplification and CCND1 amplification in our panel. 5 out of 38 CCND1-amplified lines were also ERBB2 amplified. However, CCND1 was still a significant predictor of response when added to a multiple regression model with a variable for ERBB2 amplification included (ARR= 1.74, 95% CI: 1.06-2.86).

The hypothesized role of CCND1 amplification in determining response to palbociclib has already been discussed in section 7.1.3, but will be reiterated here. The CCND1 gene codes for the cyclin D1 protein. Cyclin D1 is part of the palbociclib target complex, where it serves to regulate the activity of CDKs 4 and 6. Cyclin D - CDK4/6 phosphorylates Rb, causing in the release of the transcription factor E2F and the upregulation of E2F target genes. These transcriptional changes allow for cell cycle progression from G1 into S-phase.

Other cyclins such as cyclin E can play a redundant role to cyclin D in promoting the G1/S transition, and it has been shown that certain cancers can be preferentially dependent on the activity of one cyclin versus the other[104,105]. Amplification of CCND1 often leads to overexpression of cyclin D protein[88], which could be a marker for increased dependence on cyclin D rather than cyclin E towards Rb-dependent cell cycle progression. Cyclin D-preferring cancers would thus be expected to be more sensitive to the inhibitory effects of palbociclib.

8.2.2. 19q12 (CCNE1): chromosomal amplifications associate with palbociclib resistance

Amplification of the 19q12 chromosomal region is common oncogenic alteration in many cancer types including: breast[106], ovarian[107], endometrial[108], lung[109], bladder[110], esophageal[111] and gastric[112] cancers. The CCNE1 gene is thought to be the most likely 'driver' of amplification on this amplicon, as it is the only well-established dominantly acting oncogene in the minimally amplified region. Strengthening this claim, transgenic mouse models with induced overexpression of CCNE1 have been shown to form hyperplasias and carcinomas at higher rates than non-amplified controls[113]. CCNE1 amplification has also been shown to be a negative prognostic biomarker in many cancer types[107,109,114].

Amplification of CCNE1 was discovered in 14 out of 317 genotyped cell lines in UCLATORL's panel. These mutants were enriched for breast, ovarian and lung cancer cell lines; but at least one mutant was also found in the melanoma, colon, upper GI and head and neck panels. Table 5 shows the distribution of these alterations by primary site, compared with the same distribution in the clinical samples genotyped by TCGA. CCNE1 amplifications were similarly distributed in both datasets.

Amplification of CCNE1 was found to be significantly associated with resistance to palbociclib in our semi-supervised genotype-response association screen. 12 of the 14 amplified lines were resistant at all three dichotomous response cutoffs. The response ratios, confidence intervals and p-values are given in Table 3. Alterations in CCNE1 overlapped slightly with alterations in SGK3 and FGFR4, however

both of these genes were dropped from the final model for reasons to be discussed in section 8.2.9.

CCNE1 amplification was not significantly correlated with any of the other remaining response

biomarkers (Table 6).

Both cyclin D1 - CDK4/6 and cyclin E - CDK2 drive cellular proliferation via the phosphorylation of

Rb[115]. Hyperphosphorylated Rb subsequently releases E2F, which in turn results in the transcriptional

activation of many factors involved in cell cycle progression. Normally, cyclin D - CDK4/6 acts prior to

cyclin E/CDK2. However, it has been shown that cancer cells may preferentially use one cyclin over the

other to phosphorylate Rb, indicating that there is redundancy in their functions[104,105]. Cyclin E has also

been shown to promote entry into S phase via mechanisms that bypass Rb phosphorylation. These

mechanisms are not fully understood, but are hypothesized to involve activation of non-E2F

transcriptional regulators of cell cycle proteins such as B-MYB[116] and NPAT[117].

We propose the following mechanism for the role CCNE1 amplification plays in determining

response to palbociclib: Amplification of CCNE1 as part of the 19q12 amplicon results in overexpression

of the cyclin E protein. In amplified cancers, cyclin E overexpression is a marker of preferential utilization

of cyclin E over cyclin D in promoting S-phase entry. CCNE1-amplified cells may therefore progress

through the cell cycle independently of cyclin D - CDK4/6 activity. These cells are consequently resistant

to the inhibition of cyclin D - CDK4/6 activity by palbociclib.

8.2.3. 17q12-q21 (ERBB2): chromosomal amplifications associate with palbociclib sensitivity

Amplification of 17q12-q21 is a common genetic event in several cancer types including:

breast[118], ovarian[119], endometrial[120], gastric[121], esophageal[122] and non-small cell lung[123] cancers. The

amplicon contains several proto-oncogenes including GRB7, T2A, RARA and THRA1, but it is largely

accepted that ERBB2 is the primary oncogenic actor on the amplicon[124]. ERBB2 amplification has been

shown to be correlated with overexpression and to serve as an independent prognostic biomarker in several cancer types[118,125-128], although these associations are by far the strongest in breast cancer.

ERBB2 amplifications were identified in 25/317 genotyped lines in our panel. Table 5 shows the distribution of these alterations by primary site, compared with the same distribution in the clinical samples genotyped by TCGA. The distribution of ERBB2 amplification by primary site in our panel of cell lines largely reflected the distribution in patient samples. At least one ERBB2 amplified cell line was identified in each of the breast, ovarian, lung, and upper GI panels.

ERBB2 amplification was found to be highly associated with sensitivity to palbociclib at all three response cutoffs used in our semi-supervised genotype-response association screen (Table 3). There was some overlap observed with CCND1 amplification as well as CDH1 mutations. After controlling for these co-alterations in a multiple regression model, ERBB2 amplification was still found to be independently associated with palbociclib (ARR= 1.95, 95% CI: 1.01 – 3.74)

The ERBB2 gene encodes the human epidermal growth factor receptor 2 (HER2) protein. HER2 has no known ligand that binds its extracellular domain, and thus forms obligate heterodimers with EGFR or HER3/4 for ligand-dependent signaling activity[129]. HER2 signaling has been implicated in many distinct cell processes including: proliferation, differentiation, migration and apoptosis. Mouse models of HER2 overexpression have been used to conclusively demonstrate that ERBB2 amplification or overexpression can play a causal role in oncogenesis[130].

The cyclin D - CDK4/6 complex has been shown to be a necessary mediator of the proliferative, oncogenic effect of HER2 overexpression. Several studies have shown that intact cyclin D1 function is necessary for oncogenic transformation by HER2 overexpression in transgenic mouse and cell line models[131-133]. Cell lines whose growth is predominantly driven by HER2 signaling may therefore be

distinctly sensitive to the downstream inhibition of cyclin D - CDK4/6 by palbociclib, as it would also abrogate the oncogenic effect of HER2 signaling.

8.2.4. RB1: loss-of-function point mutations and chromosomal deletions associate with palbociclib resistance

The role of Rb loss in determining sensitivity to CDK inhibitors has already been well established and was previously discussed in section 7.1.5., but will be reiterated here. RB1 codes for the retinoblastoma protein. It acts as a negative regulator of the cell cycle via repression of E2F. Unbound E2F acts as a transcriptional activator on many genes necessary for cell cycle progression. Chromosomal deletion events resulting in the loss of the RB1 gene are known to occur in several cancer types[77]. Rb loss via deleterious point mutations has also been observed[98].

Palbociclib was originally designed to exert its anti-proliferative effect through the prevention of Rb phosphorylation by the cyclin D - CDK4/6 complex. When Rb is missing, this action by the cyclin D - CDK4/6 complex is no longer necessary to activate E2F and drive the cell cycle. It is expected that these cancers would be resistant to CDK4/6 targeted therapies as they are missing the necessary downstream effector.

RB1 loss by chromosomal deletion was relatively rare in our panel. 10 out of 317 genotyped cell lines were found to have homozygous deletions of the Rb gene. The deleted region in most of these lines was very small such that RB1 could be easily identified as the driver recessive cancer gene within the deleted region. Presumptive loss-of-function point mutations in RB1 were identified in 12 out of 240 genotyped cell lines. Loss of RB1 was most common in the breast and lung panels, but alterations were also identified in each of the ovarian, liver, kidney, lymphoma, melanoma and upper GI panels. Table 5 shows the distribution of these alterations by primary site, compared with the same distribution in the clinical samples genotyped by TCGA. RB1 alterations were similarly distributed in both mutational

datasets. RB1 alterations were not significantly correlated with any of the other response biomarkers (Table 6).

Rb loss by either genetic mechanism was strongly associated with resistance in our semi-supervised genotype-response screen. In fact, every cell line with homozygous deletion of the RB1 gene was classified as resistant at every dichotomous response cutoff. Table 2 and Table 3 give the relative risks, confidence intervals and p-values for RB1 point mutations and RB1 homozygous deletions, respectively.

8.2.5. CDH1: loss-of-function point mutations associate with palbociclib sensitivity

The CDH1 gene encodes the E-Cadherin protein. It is involved in various mechanisms regulated proliferation, mobility, and adhesion of epithelial cells. It is a substrate of the anaphase-promoting complex (APC). CDH1 is listed in the COSMIC gene census as a known recessive cancer gene that acts a tumor suppressor. Germline mutations in CDH1 have shown to increase lifetime risk of colorectal[134] and gastric[135] cancers. Somatic mutations have also been shown to occur in lobular invasive breast cancers[136,137] as well as gastric cancers[138,139].

CDH1 point mutations were found in 13/240 of UCLATORL's cell lines that were genotyped by the Broad institute. Table 5 shows the distribution of these alterations by primary site, compared with the same distribution in the clinical samples genotyped by TCGA. Our *in vitro* prevalence mirrored the clinical data in that alterations were primarily observed in breast, colorectal and gastric (upper GI) cancers.

Alterations in CDH1 were strongly associated with sensitivity to palbociclib in our semi-supervised genotype-response screen. The crude response ratio at the IC50=100nM cutoff was found to be 3.25, meaning that cell lines harboring CDH1 alterations were 3.25 times as likely to be classified as

sensitive to palbociclib. The response ratios were slightly lower for the other two cutoffs, but the association was statistically significant at all three (Table 2).

There was some overlap between CDH1 alterations and ERBB2 amplification in our panel. 5 out of 13 CDH1 mutants were also ERBB2 amplified. The Pearson correlation statistics and associated p-values are presented in Table 6. We were concerned that the observed association may have been driven by the ERBB2 effect so we created a multiple regression model with both predictors included. The independent CDH1 effect size at Cutoff 1 was slightly lower (ARR = 1.70, 95% CI: 0.95– 3.05), but still strongly in the direction of sensitivity. All of the ERBB2/CDH1 co-mutated lines were sensitive at all response cutoffs.

CDH1 plays an essential role in the regulation of the G1/S transition[140]. APC[CDH1] works to maintain cells in G1 phase by keeping levels of cyclin D low, and consequently preventing phosphorylation of Rb. This effect on cyclin D levels is mediated through the Ets2 protein. Cdh1 ultimately acts a tumor suppressor that keeps a check on cell proliferation by preventing entry into S-phase. If Cdh1 function is abrogated via genetic alteration or loss of expression, this check on the cell cycle is lost and cell proliferation is stimulated.

We propose the following mechanism for the role of CDH1 alterations in determining palbociclib sensitivity: Loss-of-function mutations in CDH1 result in the overexpression of Ets2 which in turn stimulates the expression of cyclin D. This has the effect of driving cell cycle progression through the G1/S transition via increased cyclin D - CDK4/6 activity. Cell lines harboring these alterations are consequently more dependent on CDK4/6 function, as it is likely one of the major genetic drivers of cell proliferation in these lines. Thus, these lines are uniquely sensitive to the inhibition of CDK4/6 kinase activity by palbociclib. ERBB2/CDH1 co-mutated cancers may be even more sensitive to palbociclib

treatment, as they have two upstream alterations that may serve as drivers of cyclin D - CDK4/6-mediated cell proliferation.

8.2.6. TOPBP1: loss-of-function point mutations associate with palbociclib sensitivity

The TOPBP1 gene codes for the topoisomerase II-beta-binding protein. It is required for DNA replication and plays a role in cell cycle checkpoint control in response to DNA damage via direct interaction with E2F1[141]. Germline alterations in TOPBP1 have been shown to associate with hereditary susceptibility to breast and ovarian cancer[142]. Studies examining models of TopBP1-deficiency in mice have indicated that it may behave as a tumor suppressor[143]. Aberrant TopBP1 expression has been implicated in breast cancer tumorigenesis[144].

TOPBP1 point mutations were relatively rare in our panel. Alterations were discovered in 8/240 genotyped cell lines. Table 5 shows the distribution of these alterations by primary site, compared with the same distribution in the clinical samples genotyped by TCGA. Though somewhat rare, these alterations were broadly distributed throughout different cancer types, with at least one mutant cell line present in each of the breast, liver, colon, melanoma, upper GI, lymphoma and sarcoma panels. The distribution in TCGA patient samples did not entirely mimic the distribution *in vitro*, but a similar broadly-distributed pattern was observed.

TOPBP1 mutant cell lines were found to be significantly associated with sensitivity to palbociclib treatment in our genotype-response association screen, especially at the highest sensitivity cutoff (Table 2). 7/8 of these mutants had palbociclib IC50s less than the median value. There were no significant correlations between alterations in TOPBP1 and any of the other sensitivity-associated alterations in our biomarker set.

Normal TopBP1 protein has been shown to play a necessary, dual role in the G1/S transition[145]. First, TopBP1 activates the cyclin E/CDK2 complex via the down-regulation of cell cycle inhibitor proteins p21 and p27. Second, TopBP1 recruits replication components to chromatin during initiation of DNA synthesis. Models of TopBP1 deficiency have shown down-regulation of cyclin E/CDK2 activity, G0/G1 arrest and induction of cellular senescence[143].

We propose the following mechanism for the role of TOPBP1 alterations in determining palbociclib sensitivity: Loss-of-function alterations in the TOPBP1 gene result in abrogated function of the TopBP1 protein. This leads to the up-regulation of the cell cycle inhibitors p21 and p27. High levels of p21 and p27 cause significant down-regulation of cyclin E/CDK2 complex. Cancer cells with this alteration are consequently more dependent on cyclin D - CDK4/6 activity, as it plays a semi-redundant role to cyclin E - CDK2. When TOPBP1-mutant cells are treated with palbociclib, cyclin E/CDK2 is unable to compensate for the loss CDK4/6 activity, leading to cell cycle arrest.

8.2.7. SMO: activating point mutations associate with palbociclib resistance

The SMO gene codes for the smoothened protein; a G-protein coupled receptor that is normally bound and repressed by patched, the receptor for the hedgehog ligand. Upon hedgehog-patched binding, smoothened is released and activated via phosphorylation[146]. This triggers a signaling cascade that ultimately results in the activation of GLI transcription factors. Several GLI transcriptional targets are involved in cell proliferation and cell cycle regulation[147] including: cyclin D1, cyclin E, and IGF2 amongst others. Activating mutations in SMO have been discovered in basal cell carcinomas[148], medulloblastomas[149], and gastric[150] cancers. Transgenic mouse models with constitutively active SMO produce legions similar to basal cell carcinomas[151].

In UCLATORL's cell line panel, 9 out of 240 genotyped cell lines were found to have presumptively activating point mutations in SMO. These mutants were found in diverse cancer types

with at least one mutant in each of the colon, ovarian, kidney, liver and lung cell line panels. Table 5 shows the distribution of these alterations by primary site, compared with the same distribution in the clinical samples genotyped by TCGA. Point mutations in SMO were more common in our cell lines than in the TCGA patient samples, especially in the colon, kidney and liver cell line panels. The consequence of these discrepancies will be discussed in detail in section 9.2.

Point mutations of SMO were found to be significantly associated with resistance to palbociclib in our semi-supervised genotype-response screen (Table 2). In fact, all nine mutated cell lines were classified as resistant to palbociclib at the lower two response cutoffs. Mutations in SMO were not mutually exclusive with other resistance biomarkers that fell out of our screen. Five out of nine SMO-mutated lines also had mutations in MLL, TSHR, FANCA or NBN. However, all four of these alterations were excluded from our final predictor set for reasons outlined in section 8.2.9.

We propose the following mechanism for the role that activating mutations in SMO play in determining response to palbociclib: point mutations in SMO lead to constitutive activation of the hedgehog signaling pathway. This signaling cascade bypasses the cyclin D-Rb-E2F axis for S phase entry and directly results in the transcription of factors involved in cell cycle progression (notably cyclin E). This relieves the dependence on cyclin D - CDK4/6 activation for S-phase entry. When cyclin D - CDK4/6 is inhibited by palbociclib in SMO-mutant lines, the GLI-mediated upregulation of cyclin E and other proliferative factors compensates to drive the cell cycle through the G1/S transition. Consequently, SMO-mutant lines are more resistant to the inhibitory effects of palbociclib.

8.3. Candidate predictors excluded from "final predictor set"

Of the fourteen original candidate biomarkers identified in the preceding genotype-response association screen, six were rejected for failing to meet the criteria for predictive/translational capacity presented in section 8.1.

Amplifications in the chromosomal region mapping to 1q32 were found to be associated with sensitivity to palbociclib. However, this alteration was strongly inter-correlated with other sensitivity biomarkers including: CCND1 amplification, ERBB2 amplification and CDH1 mutation (Table 6). In fact only three of the nine 1q32 amplified lines lacked co-mutation of other sensitivity genes. The independent effect of 1q32 alterations was found to be non-significant when added to a multiple regression model with the other biomarkers (ARR=0.70 , 95% CI: 0.27–1.88). Furthermore, no clear proto-oncogene could be identified within the amplified region that had a plausible mechanistic relationship to CDK4/6 inhibition.

Amplifications in the chromosomal region mapping to 8q13 were found to be associated with resistance to palbociclib. This association was significant at only one of the three response cutoffs analyzed (Table 2). 8q13 amplification was strongly inter-correlated with 19q13 (CCNE1) amplification (Table 6) and overlapped slightly with RB1 loss as well. The independent effect of 8q13 amplification was no longer found to be significant (and in fact switched direction) when added to a multiple regression model with the other biomarkers (ARR= 1.33, 95% CI: 0.71-2.49). Furthermore, no clear proto-oncogene could be identified within the amplified region that had a plausible mechanistic relationship to CDK4/6 inhibition.

Presumptive activating point mutations in the MLL gene were found to be crudely associated with resistance to treatment with palbociclib in our original screen. This association was significant at only one of the three response cutoffs analyzed (Table 2). Alterations in MLL were found to be significantly associated with TSHR mutations, and overlapped with several other of the resistance biomarkers including: SMO mutations, FANCA mutations, RB1 loss and CCNE1 amplification. When added to a multiple regression model with these extra variables, the independent effect of MLL1 mutation was no longer significant (ARR= 0.97, 95% CI: 0.55-1.73). No clear mechanism for the role of

MLL activation in determining response to CDK4/6 inhibition could be elucidated from the literature. MLL point mutations have not been identified in patient tumor samples outside of hematological malignancies, where rearrangements seem to be more common than point mutations[152].

Presumptive activating mutations in the TSHR gene were found to be crudely associated with resistance to treatment with palbociclib in our original screen. This association was significant at only one of the three response cutoffs analyzed (Table 2). Mutations of TSHR were significantly correlated with other resistance biomarkers including: MLL mutations, SMO mutations, FANCA mutations, NBN mutations and RB1 mutations. Only three of the ten TSHR mutants were not co-mutated with another resistance gene. When added to a multiple regression model with these extra variables, the independent effect of TSHR mutation was no longer significant (ARR= 1.13, 95% CI: 0.57-2.24). Additionally, no clear relationship between TSHR activation and CDK4/6 biology could be elucidated from the literature.

Presumptive loss-of-function alterations in the NBN gene were found to be crudely associated with resistance to treatment with palbociclib in our original screen. Mutations of NBN were significantly correlated with other resistance biomarkers: including TSHR and SMO. When added to a multiple regression model with these extra variables, the independent effect of NBN mutation was no longer significant (ARR= 1.31, 95% CI: 0.69-2.49). No clear relationship between NBN loss and CDK4/6 biology could be elucidated from the literature.

Presumptive loss-of-function alterations in the FANCA gene were found to be crudely associated with resistance to treatment with palbociclib in our original screen. Mutations of FANCA were significantly correlated with TSHR and overlapped slightly with other resistance biomarkers: including SMO mutations, NBN mutations, RB1 loss, and CCNE1 amplification. When added to a multiple regression model with these extra variables, the independent effect of FANCA mutation was no longer

57

significant (ARR=0.94, 95% CI: 0.52-1.68). No clear relationship between FANCA loss and CDK4/6 biology could be elucidated from the literature.

8.4. Creation of collapsed biomarker variables and performance by cancer type

Sections 8.1-3 described the restriction of the list of candidate biomarkers to just those that had a plausible biologic mechanism and whose association with response was independent from the other biomarkers in the set. This limited the list to four genes (ERBB2, CCND1, CDH1, TOPBP1) where genetic alterations were associated with sensitivity and three genes (SMO, CCNE1, RB1) where genetic alterations were associated with resistance. These associations were made in a pooled analysis where all cancer types were included and given equal weight. However, clinical development of novel cancer therapeutics is often conducted in a cancer-type specific setting. Clinical trials are generally restricted to just one or few cancer types.

Given this consideration, we wished to describe how well our cassette of candidate biomarkers performed in a cancer type-specific analysis. To accomplish this, we first created indicator variables for "any sensitivity biomarker" or "any resistance biomarker" for the 387 cell lines where genomic data was available. One logistical concern that arose during this step was the handling of missing data. We have copy number data on 317 cell lines and point mutation data on 240 cell lines. 362 cell lines had either dataset available and 196 had both datasets. Since each individual genetic alteration was rare in our panel, we chose to keep those cell lines that were missing one of the two datasets and impute 'wild-type' for their mutational status. These collapsed indicator variables were then tested for association

with palbociclib response using the same modified Poisson regression model that was used for the semi-supervised biomarker screens.

Overall, 71 out of 362 lines that had genetic data available harbored at least one of the four sensitivity biomarkers included in our final set. Table 8 gives the response ratios and p-values for this collapsed biomarker variable for each cancer type and at each of the three dichotomous response cutoffs. This collapsed set of biomarkers associated in the direction of sensitivity in almost every cancer type where enough data was available to run the regression. However, sparse data made difficult to achieve statistical significance in many of the cancer type sub-strata. The overall, pooled, crude association was highly significant at all three cutoffs, with the highest effect size at the 100nM IC50 cutoff for response (RR = 2.66, 95% CI: 1.79-3.97).

44 out of 362 cell lines where genotyping data was available harbored at least one of the four resistance biomarkers included in our final set. Table 9 provides the response ratios and p-values for the indicator variable representing "any resistance biomarker" for each cancer type and at each of the three dichotomous response cutoffs. This collapsed variable predicted strongly in the direction of resistance in nearly every cancer type where there was enough data to run the regression, the only exceptions being cancer types with very small numbers of biomarker positive cell lines. The pooled effect across all primary sites was highly significant at all three response cutoffs. The effect size was highest at the 100nM IC50 cutoff (RR = 0.22, 95% CI: 0.08 – 0.58).

8.5. Control of confounding variables

To this point all of the response ratios presented for each biomarker association in our results tables have been crude, unadjusted risk estimates. The only covariate control performed so far was for

the inter-correlation between candidate biomarkers (sections 8.1-3). This analysis was performed as part of a model building schema, and not as a post-hoc adjustment to our effect estimates. The highly inter-correlated biomarkers were removed from our model, the remaining biomarkers are largely independent of one another. However, there are two additional covariates that have a potential biasing effect on our biomarker-response associations.

The first of these potential confounders is the differential growth rate between cell lines. We have already described unique methods (section 3.2) to minimize the mathematical artifact of these differential growth rates in our outcome measure calculations. However, there remains the possibility that growth rate itself may be causally involved in response determination, especially to a cell cycle inhibitor like palbociclib. Indeed, growth rate (measured as untreated population doublings per five days) was found to be significantly associated with response to palbociclib within many cancer types (Supplemental Table S3). Slow growing cell lines tended to be more sensitive to treatment with palbociclib on average than faster growing lines. To the extent that growth rate is also associated with any of the biomarkers in our model; this could directionally bias our results.

The second potentially biasing covariate is the primary site of origin of the tumor from which each cell line was derived. Different cancer types arise from epigenetically distinct progenitor cells. There can be considerable differences in the phenotypic characteristics of cancer cells due to the developmental epigenetic processes that occur during cellular differentiation. These broad morphological and histologic characteristics of certain cell types may themselves have a causal relationship with palbociclib sensitivity. To the extent to which each biomarker in our model is differentially distributed between cancer types; this could directionally bias our effect estimates.

There is also the possibility that either of these covariates could be an intermediate, rather than a confounder of the genotype-response association. For example, slower growth rate may be a

phenotype determined by an underlying genotype. If genetics determine growth rate, which in part determines response to treatment with palbociclib, than this intermediary effect should not be adjusted out of our effect size estimates. The same argument can apply to primary site. If a certain cancer type with high prevalence of sensitivity biomarkers is more sensitive to treatment with palbociclib, it could be that the sensitivity biomarkers are solely responsible for the association, and not any histological or epigenetic characteristics of that cancer type. We chose to present the adjusted effect estimates with the caveat that they may reflect an over-adjustment away from the crude estimate if either covariate behaves partially as an intermediate rather than a confounder.

Table 10 shows the crude and adjusted response ratios and confidence intervals for the two indicator variables representing "any sensitivity biomarker" and "any resistance biomarker" after controlling for growth rate and cancer type. Here it can be seen that the adjusted effect estimates for the "any sensitivity biomarker" variable moved significantly towards the null, while the adjustments did not appreciably change the effect estimates for "any resistance biomarker."

8.6. Quantification of misclassification in genomics datasets

The genotyping technologies used to assay the genetic alterations harbored by our cell lines are known to have some level of inaccuracy. Imperfect probes, noise in the readout, threshold choices and other limitations can result in genotype misclassification. Therefore there is a reasonable concern that some of the cell lines in our analysis have been misclassified on their biomarker status. We wished to quantify this potential for misclassification in our characterization datasets.

The Broad Institute's Cancer Cell Line Encyclopedia project has performed redundant genotyping of copy number alterations on a large fraction of cell lines in our panel. This affords the opportunity of multiple source comparison to quantify the likely level of misclassification in each of our

biomarkers. There exist 192 overlapping cell lines where both the CCLE and UCLATORL have performed

whole-genome copy number alteration genotyping. CCLE implemented a different array technology

(SNP6.0 arrays) than we did for our internal genotyping (comparative genomic hybridization arrays).

However the readouts of the two assays are comparable.

First both datasets were reclassified from continuous $\log_2$(ratio) data to dichotomized calls on

amplification or deletion status. The thresholds used for this classification were described in sections

2.2.1. and 2.2.2. The inter-rater agreement for calls in each dataset was quantified using Cohen's kappa

coefficient. Table 11 provides the frequency of each copy number alteration in both datasets, the

frequency of overlap as well as kappa coefficients. There was generally very good agreement between

the CCLE and UCLA genotyping calls. The kappa coefficients for CCND1 amplification, ERBB2

amplification and RB1 deletion were all around 0.9.

CCNE1 amplification was the only copy number alteration in our biomarker set to have

significant disagreement between labs. Are closer analysis of the raw data revealed that this was mostly

due to threshold effects. There were six cell lines that UCLATORL classified as CCNE1 amplified that the

CCLE did not. These six lines all had $\log_2$(ratio)s just under to the cutoff for classification as amplified in

the CCLE SNP6.0 dataset, while they were just above the threshold in our CGH dataset. The fact that the

$\log_2$(ratio)s were directionally biased for only this gene likely indicates that the probe-binding stringency

in this particular region varied between the two different chips used in the respective assays. A follow-

up assay such as FISH probing would definitively answer which labs genotype calls more accurately

reflect the discrete amplification state of these cell lines.

The point mutation dataset could not be investigated for misclassification by multiple source

comparison, as only one whole-exome sequencing dataset was available on the cell line panel. The

misread rate inherent to the hybrid capture technology used for exome sequencing has been shown to

be quite low[153]. It is likely that the majority of misclassification present in this dataset is not due to

sequencing errors, but rather the difficulty in functionally validating the point mutations as truly

activating or loss-of-function alterations. Section 5.1 described many of the restriction criteria used to

enrich our dataset for likely functional alterations, but there remains a strong possibility that many of

the point mutations in our dataset do not have a strong functional effect on the proto-oncogene in

which they reside. The CCLE plans to add functional validation data to their publically available hybrid

capture dataset. In the future, this data can be incorporated to further reduce misclassification rates in

this context.

CHAPTER 9: Discussion

9.1. Palbociclib response *in vitro* and genomic predictors of differential sensitivity

Palbociclib is a potent and highly selective inhibitor of the cyclin D - CDK4/6 complex. It belongs to an exciting class of cell-cycle targeted compounds that may have significant clinical utility in slowing the growth of human cancers. The success of these compounds hinges on the ability to accurately identify subpopulations of patients whose cancers may be comparatively sensitive or resistant to treatment. Our lab previously published the observation that ER-positive breast cancers may be uniquely sensitive to CDK4/6 inhibition[57]. This discovery led to the initiation of a Phase II clinical trial where remarkable efficacy was observed[64]. However, to our knowledge no clear biomarker (or set of biomarkers) exists in other cancer types that may be similarly used to enrich future clinical trials and open up new indications for CDK4/6 inhibitors. We wished to explore the sensitivity profile of palbociclib across several cancer types in an *in vitro*, preclinical setting and use the observed response distribution to inform the development of candidate biomarkers of palbociclib sensitivity and resistance.

To this end, we modeled differential sensitivity to palbociclib across a panel of 416 human cancer cell lines derived from 12 distinct cancer types. We observed remarkable anti-proliferative activity in a subset of sensitive cell lines within nearly every cancer type tested, suggesting that there is potential for very diverse clinical application of this compound. This response profile was unique to the UCLATORL dataset; parallel screening efforts run by the Broad[26] and Sanger[27] institutes failed to identify this large palbociclib-sensitive cohort. We employed unique proliferation assay protocols and analytics that have been optimized for the identification of long-term cytostatic effects. We believe our methodology allowed for superior *in vitro* response characterization in this context. The primary factors

differentiating UCLATORL's protocols (as discussed in detail in section 3) included: longer treatment

windows, seeding of larger cell populations at baseline, incorporation of day 1 counts and

implementation of logarithmic (generational) formulae for calculation of percent inhibition.

This response data was incorporated into a custom-designed database for the high-throughput

identification of genotype-response associations. This database also holds highly-detailed genetic

characterization data on each subject. This genomic data has been curated by UCLATORL to enrich for

somatic alterations likely to be causally involved in the cellular mechanisms of cancer development. This

genomics dataset was interrogated for biomarkers of sensitivity and resistance in a large, semi-

supervised screen. The statistical models employed for the identification of genotype-response

associations was generalized from epidemiological tools used in the analysis of prospective cohort

studies. Namely, a modified Poisson regression model with robust error handling was used. After

controlling the false discovery rate, alterations in thirteen genes were found to be significantly

associated with either sensitivity or resistance to palbociclib *in vitro*.

These crude associations were further investigated via extensive post-hoc analyses. This

included: analysis of biomarker inter-correlation, comparison of biomarker distribution in cell lines

versus clinical samples, and a literature search to develop hypothetical mechanisms of response

determination. Overall, six of the fourteen crude associations were rejected for use as candidate

biomarkers due to strong inter-correlations with other alterations that had a more defined mechanistic

role in CDK4/6 biology.

Several biomarkers identified in our screen validated some of the existing hypotheses based on

a priori knowledge of palbociclib's mechanism of action. Loss of retinoblastoma protein function by

point mutation or chromosomal deletion had been long-presumed to confer resistance to CDK4/6-

targeted therapies, as it is the immediate downstream effector of CDK4/6. Indeed, we found that nearly

every cell line with a genetic alteration in the RB1 gene was highly resistant to palbociclib treatment. Conversely, amplification of the CCND1 gene has been hypothesized to confer increased dependence on cyclin D - CDK4/6 activity and therefore sensitivity to CDK4/6 inhibition. We verified that CCND1 amplification was strongly associated with, but not completely sufficient for *in vitro* sensitivity to palbociclib.

We were not able to validate several other a priori assumptions from the literature in our *in vitro* models. CDK4 amplification, CDK6 amplification and CDKN2A/B deletion have been previously hypothesized to confer sensitivity to CDK4/6 inhibitors, and in some cases are already being used as inclusion criteria for clinical trials. The crude association with palbociclib sensitivity was modest or non-existent for these biomarkers in our genotype-response screen. The lack of an association with CDK4 and CDK6 amplification may simply be that we were underpowered to observe an association due to the relative rarity of these alterations in our cell lines. However, the lack of an association with CDKN2A deletion may be a more complicated story.

Loss of CDKN2A by genetic or epigenetic mechanisms is one of the most common somatic alterations in human cancers[77,93,94]. Even so, mutation frequencies in cell lines have been discovered to be significantly higher than patient samples from the same cancer type[99]. Indeed, frequency of homozygous deletion of CDKN2A UCLATORL cell line panel is significantly higher than the clinical rates listed by the TCGA in most cancer types (Supplemental Table S4). The very high prevalence of CDKN2A alteration in cell lines makes its utility as a clinical biomarker difficult to analyze, especially when sensitivity to treatment is comparatively rare. Therefore, the lack of an association in our dataset may not imply that a true association does not exist; rather it was difficult and inappropriate to model using our cell line panel.

In addition to confirming some of the a priori assumptions regarding palbociclib response biomarkers, we found several novel genotype-response associations that no previous literature had hypothesized. In retinoblastoma-competent cell lines, these biomarkers appear to exert their influence on palbociclib response through affecting the molecular preference of cyclin D versus cyclin E. The mechanistic result of alterations associated with palbociclib sensitivity (CCND1 amplification, ERBB2 amplification, CDH1 loss-of-function and TOPBP1 loss-of-function) is the activation and resulting dependence on cyclin D or the deactivation and resulting independence of cyclin E. Conversely, the shared mechanism between alterations associated with palbociclib resistance (CCNE1 amplification, SMO activating mutations) is the activation and resulting dependence on cyclin E over cyclin D.

The final set of candidate biomarkers was unevenly distributed amongst the cancer types represented in the UCLATORL cell line panel. The strength of the genotype-response associations also varied between cancer types. However, there are several cancer types that have high enough biomarker frequencies in both cell lines and clinical samples that a justifiable clinical development strategy could be implemented. In the context of breast cancer, our data suggest that there may be cause for expanding use of palbociclib to HER2-positive (ERBB2-amplified) patient populations, as cyclin D is a necessary intermediate of the oncogenic effect of HER2 pathway activation.

Our biomarker set may be used to develop similar treatment rationales in other cancer types. The upper gastrointestinal (stomach, esophageal) cancer cell lines were the second most sensitive cancer type on average in our screen. The upper GI cell line panel also had high frequencies of sensitivity biomarkers (CDH1 mutation, CCND1/ERBB2 amplification) that were relatively consistent to clinical frequencies. The head and neck cancer cell line panel was also comparatively sensitive on average compared to other cancer types. Head and neck cancers have a high frequency of CCND1 amplification, whereas resistance biomarkers are relatively rare both in our cell line panel and in clinical samples.

The lung and ovarian panels were not as sensitive on the whole as compared to the breast, head and neck and upper GI panels. However, the sensitivity and resistance biomarkers in our final cassette performed very well in these panels. Ovarian and lung are among the most common cancer types worldwide. Even if just a small fraction of palbociclib-sensitive lung or ovarian cancers could be identified, these would represent sufficiently large patient cohorts to justify the clinical development of palbociclib for use in these subpopulations.

9.2. Limitations of the study

There are several limitations that should be considered when interpreting the results of our *in vitro* genotype-response association screen. First, there is the potential for misclassification in both the response and genomic characterization datasets used in our analysis. Second, there is the potential for residual confounding by unmeasured genetic or epigenetic effects. Third, the *in vitro* models used in the analysis have many limitations on generalizability to the more complex biological systems that govern response to treatment *in vivo*.

Misclassification of palbociclib sensitivity and resistance is likely the largest single source of error and/or bias in our study. The proliferation assays performed by UCLATORL to quantify the cytostatic effect of palbociclib have an inherent 'noisiness' in their output. Tissue culture work is notoriously imprecise, though UCLATORL has made many modifications to our protocol to minimize this inherent error. We have also implemented several unique analytic tools that distinguish UCLATORL from other large, cell line-based screening efforts. These protocol and analytic distinctions have been discussed at length in Chapter 3.

Despite our effort, there likely remain many cell lines where IC50s and other summary outcome measures have been estimated with non-negligible error, and therefore do not exactly reflect the true biology of response. Furthermore, the dichotomous response cutoffs used to classify sensitivity and resistance are necessarily somewhat arbitrary. To account for these limitations, we built a sensitivity analysis into our genotype-response screen where three separate dichotomous cutoffs were used. This methodology was discussed in detail in Section 4.

There is also likely to be some misclassification of the genomic data used in our screen. The chip technology used for both the copy number and point mutation genotyping are known to have very low read error rates, but it is not impossible for some of our mutant/wild-type calls to be misclassified. An issue of higher concern is threshold effects in the copy number datasets. Copy number calls for amplification and deletion represent discrete genomic states within a particular cancer, but the calls are made from somewhat imprecise continuous data. In addition, probe quality can bias these continuous measurements in one direction or the other, leading to a systematic under- or overestimation of rates of amplification or deletion for that gene. Finally, the copy number data has yet to be functionally verified (gene amplification leads to overexpression, deletion leads to loss of expression) for every biomarker-positive cell line identified in our screen.

The point mutation dataset obtained from the Cancer Cell Line Encyclopedia also may have similar misclassification issues. In this context, the lack of functional verification of each mutation is a potentially larger problem. While some of the specific mutation loci included in our dataset are functionally verified "hotspots", the majority of the putative activating or loss-of-function mutations in the dataset have an unverified functional consequence on the resulting protein. Chapter 5 discusses the criteria used for enriching this mutation dataset for likely functional alterations, though it is possible that some non-functional mutations were included.

The majority of the misclassification issues discussed so far would likely have the effect of biasing our results towards the null. As a result, we may have missed some genotype-response associations with rare alterations or small effect size. As we further refine our genomic and response datasets to reduce this noise, it is possible that we will have the ability to identify more biomarkers of palbociclib response.

One major source of directional bias in our study could be confounding by unmeasured epigenetic variation between cell lines. Cancer cells can share epigenetic similarities based on their cell type of origin, owing to the molecular processes underlying cellular development and differentiation. Some of these epigenetic similarities and differences may affect response to palbociclib. To the extent that they also correlate with the genetic alterations in our analysis, they may confound the genotype-response associations we identified. We attempted to control for this potential source of confounding by including cancer type (or primary site) in a multiple regression model with our candidate biomarkers. However, cancer type is an imperfect surrogate for the complex epigenetic variation between cancers. As the field of cancer epigenetics progresses, we may be able to better account for this source of confounding in future analyses using data sources such as mRNA expression, DNA methylation, histone acetylation, etc.

Another potential source of confounding is non-random genetic variation between individual cancers. It is well known that clonal selection effects can lead to co-enrichment of certain genetic alterations. We performed a correlation analysis to investigate this possibility in our set of candidate biomarkers, and adjusted for these effects when they were observed. However, there remains the possibility that other unmeasured genetic alterations could be confounding our analyses. If one of these unmeasured alterations is strongly correlated with one of our biomarkers as well as response to palbociclib, it could confound our genotype-response associations.

It should be noted that these directional biases may be of slightly less concern in this context than they would be in traditional exposure-disease epidemiologic analyses. For biomarkers of response to treatment, true causality in the genotype-response association is a secondary concern to its performance towards successfully identifying sensitive and/or resistant patients. As long as the factors that caused the directional bias in our preclinical work are similarly distributed in the clinical setting, non-causal biomarkers can still be clinically useful. However, we realize this is a fairly strong assumption and therefore we attempted to limit our list of candidate biomarkers to those most likely to play a causal role in response determination.

Finally, there are many experimental conditions in our cell-line based modeling that may limit the generalizability of our results to the clinical setting. *In vitro* experiments performed on cell line panels can only model molecular and cellular level effects of drug response. Higher-level biologic effects such as tumor micro-environment, vascularization, immune involvement, drug metabolism, etc. cannot be modeled in our limited, *in vitro* platform. It is likely that some or all of these factors may effect palbociclib sensitivity in the clinic. The cellular and molecular associations described in our analysis must ultimately be combined with whole organism-level observations made in the clinical setting.

Another limit on the generalizability of our genotype-response associations may be that the distribution/frequency of genetic alterations in our cell line panel may not reflect the true distribution in patients. Certain alterations may confer a growth advantage *in vitro* and therefore be enriched in cell lines. Furthermore, certain alterations may have even spontaneously occurred de novo in the development of the cell line. We attempted to quantify these *in vitro* / *in vivo* differences by comparing the frequency of alterations in our sample to those in the clinical samples genotyped as part of the Cancer Genome Atlas.

In general, we observed higher rates of alteration than were seen in clinical samples. However, these differences may be partially attributed to the sampling and genotyping methods employed be the TCGA. Contamination by surrounding non-cancerous tissues may also be a concern in genotyping patient samples. This contamination may reduce the ability to identify a point mutation, and may also affect the sample-averaged $\log_2$(ratio)s from copy number data. Overall, this would have the effect of lowering the mutation call rate in patient samples below what the true rate may be, and may partially explain the higher rates observed in cell lines. Despite slight discordances, all of the biomarkers in our final set were sufficiently frequent in patient samples to be considered for use as inclusion/exclusion criteria for future clinical trials.

9.3. Future directions

There are still many unanswered questions that should be investigated before any of the genotype-response associations identified in this study are translated to the clinical setting. First, the specific genomic alterations present in our cell lines must be functionally verified. For copy number alterations, this would entail verification that DNA-level changes correlate with RNA expression level changes, and that this in turn causes downstream activation/deactivation of the proto-oncogene's signaling pathway. Point mutations must also be functionally confirmed to be truly activating or loss-of-function via molecular and biochemical assays.

Next, causation in the genotype-response relationship can be explored for each candidate biomarker via carefully designed engineered cell line models. Alterations can be engineered into a wild-type background in order to see if the addition of the alteration induces a change in palbociclib sensitivity. Similarly, siRNA knockdowns of activated oncogenes can be performed to see if palbociclib sensitivity is affected. *In vivo* models can also be developed to verify that the genotype-response

associations observed in our 2-dimensional tissue culture assays hold up in the context of 3-dimensional mouse xenografts.

Finally, the ultimate confirmation of the utility of these biomarkers will have to come from human clinical data. This may be difficult to achieve in practice, as many clinical trials are already being enriched for candidate genomic biomarkers and therefore lack sufficient controls for the investigation of the efficacy of these biomarkers. However, there are some 'all-comer' trials with palbociclib currently underway where some of these questions may be investigated. However, the ability to explore our candidate biomarker associations in these clinical datasets may be hindered by the sample collection and genotyping practices employed by each clinical study.

Figure 1: Distribution of cell lines in the UCLATORL panel organized by primary site of tumor from which cell line was established. UCLATORL: UCLA Translational Oncology

Research Laboratory. GI: Gastrointestinal.

UCLATORL Cell Line Panel by Primary Site
(N = 416)
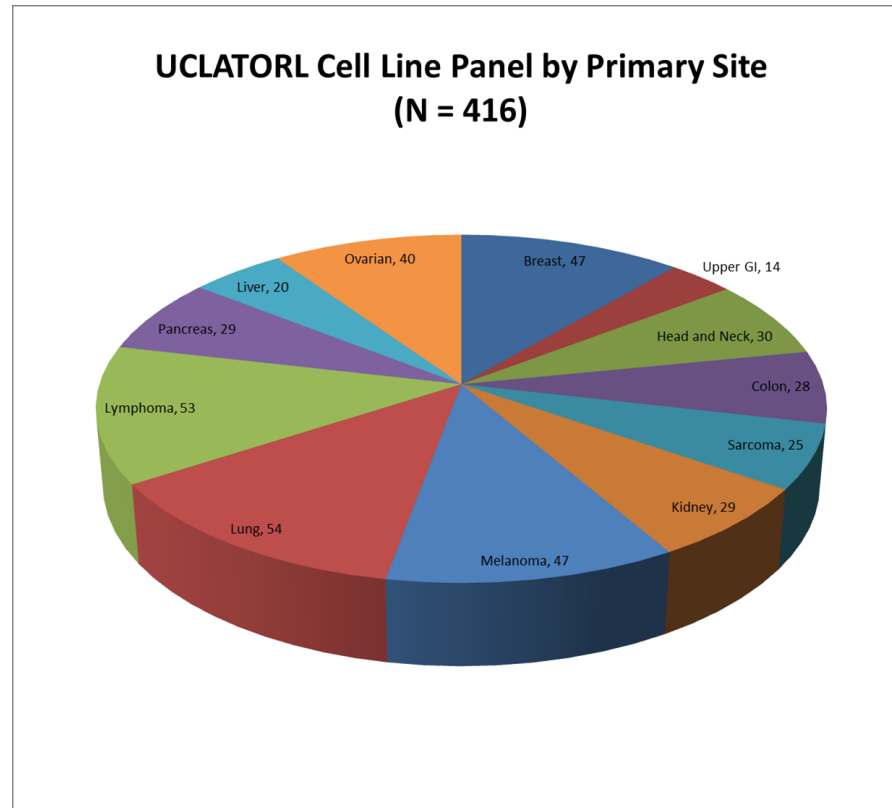
Figure 2: Flowchart for the UCLATORL 5-day proliferation assay protocol used to determine *in vitro* response to palbociclib.
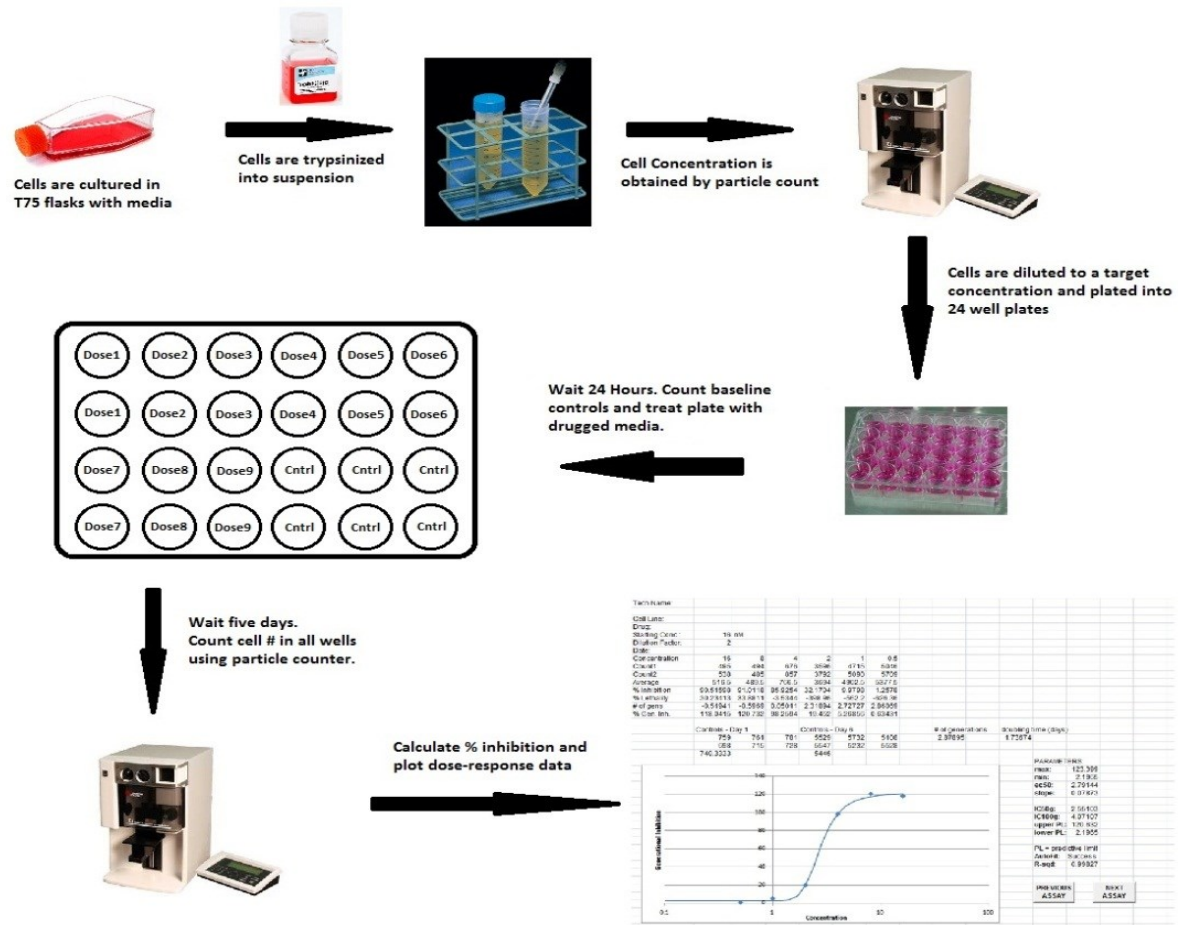
Figure 3: Palbociclib response distribution by IC50. Each bar in the chart represents the average (geometric mean) IC50 value for one cell line in the UCLATORL panel. Cell

lines are rank ordered from most sensitive (left) to most resistant (right).

Rank Order of Palbociclib Sensitivity by IC50
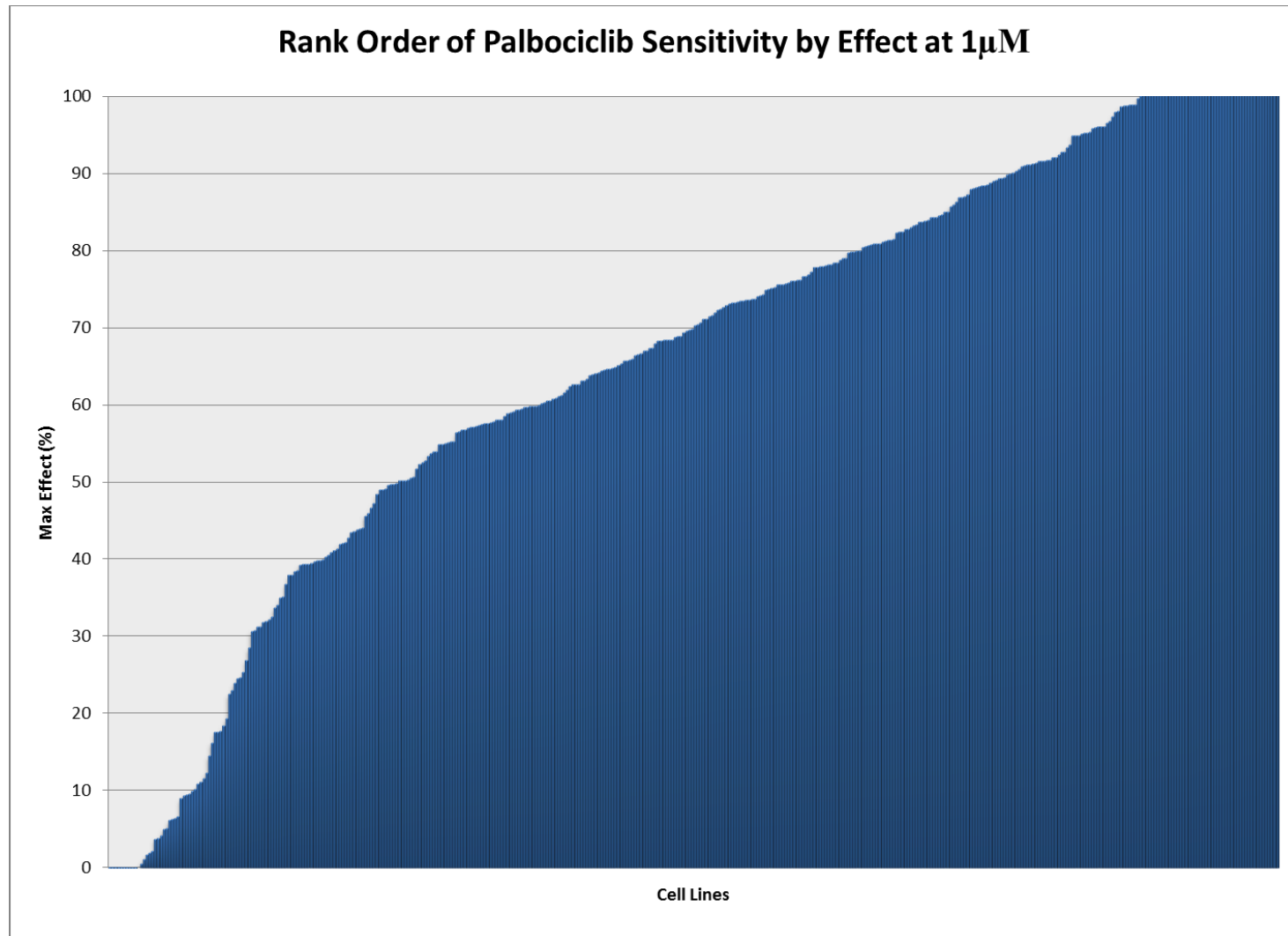
Figure 5: Box plot of average palbociclib IC50s by Cancer Type.
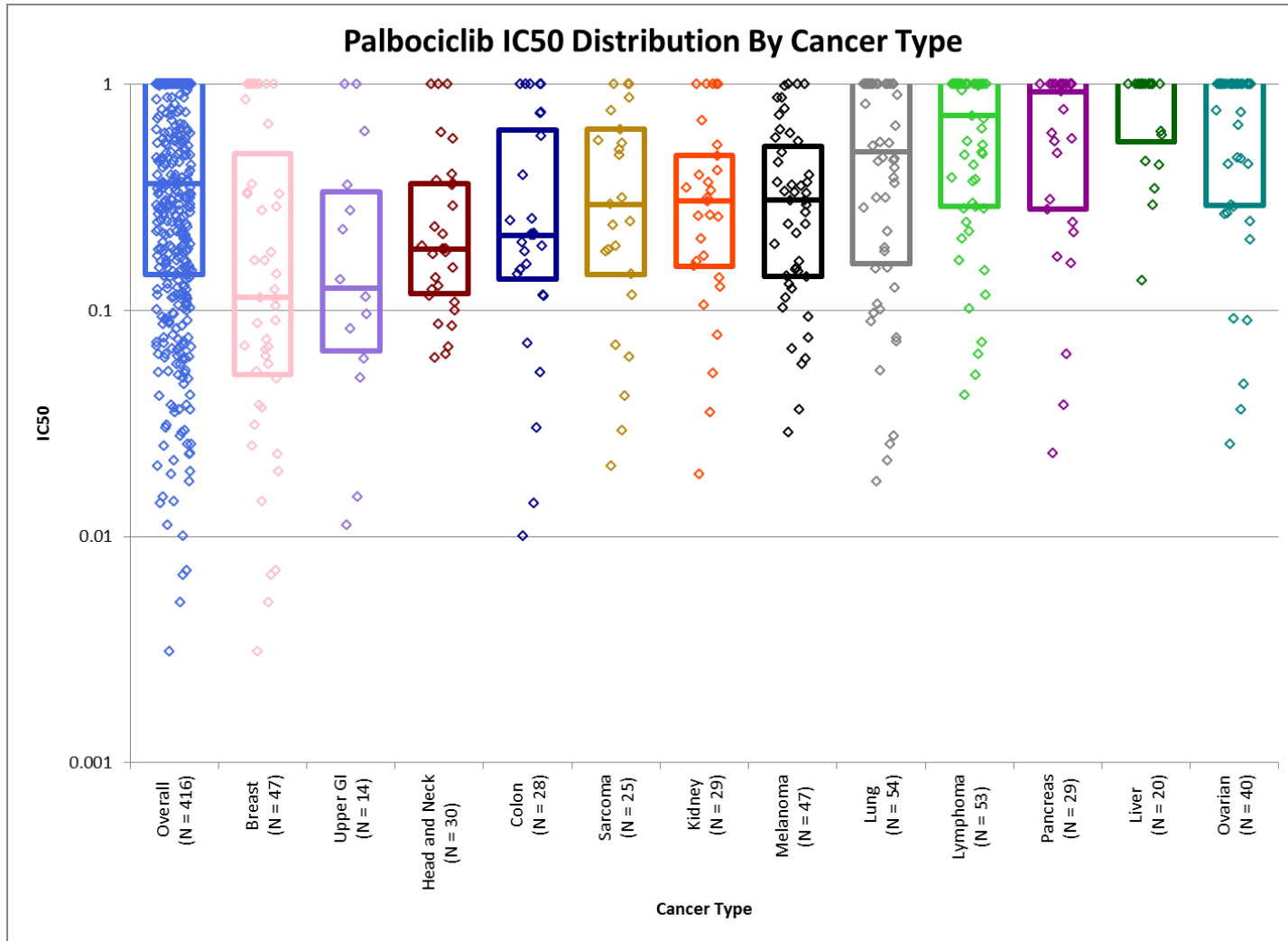
Figure 6: Box plots comparing palbociclib IC50 distributions between UCLATORL, the Sanger Institute, and the Broad Institute Cancer Cell Line Encyclopedia (CCLE). Left: IC50 distribution for each screening project's entire cell line panel. Right: IC50 distribution for each screening project comparing only overlapping cell lines.
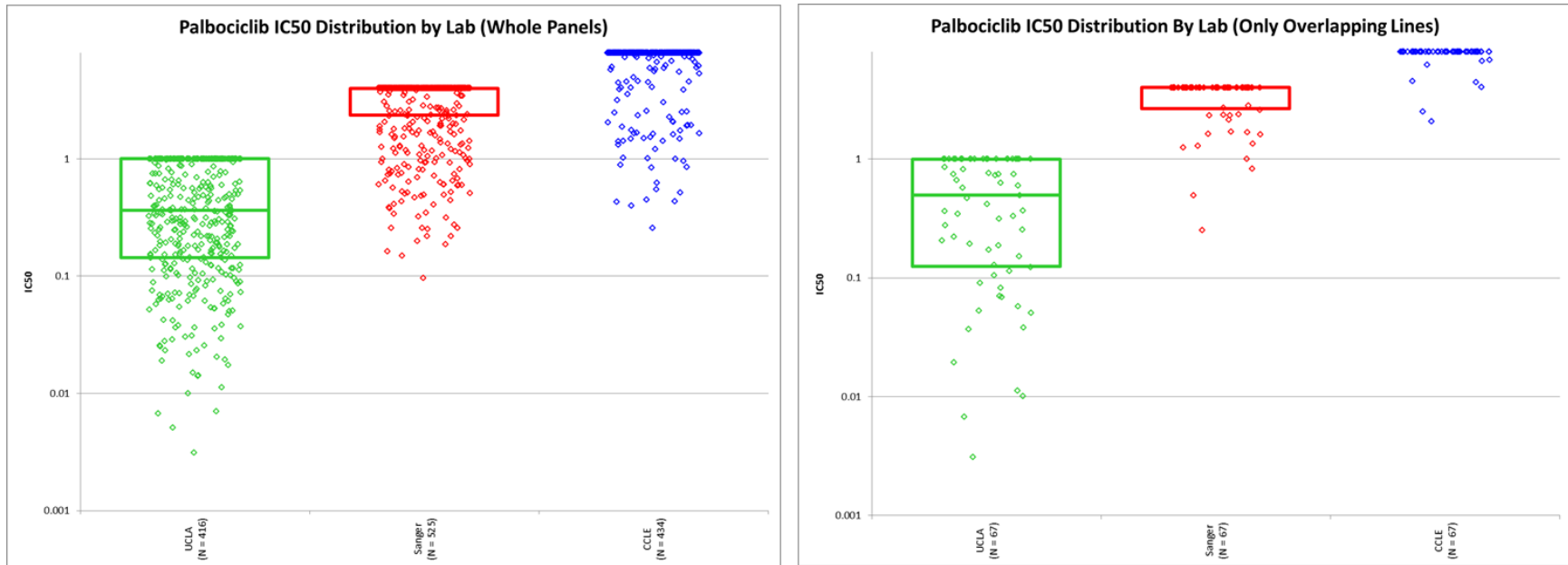
Figure 7: Scatterplot comparing IC50 distribution between the Sanger Institute and the Broad Institute's Cancer Cell Line Encyclopedia. Each point represents a cell line where IC50 data was generated independently by each lab. X-coordinate: IC50 from Sanger. Y-coordinate: IC50 from CCLE.

Example: Slow Growing Cell Line
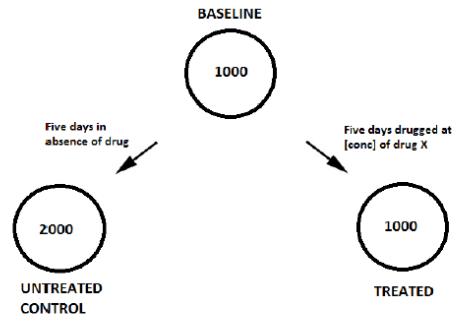
BASELINE

1000

Five days in absence of drug

Five days drugged at [conc] of drug X

2000

1000

UNTREATED CONTROL

TREATED

"Survival Fraction" Method

$$\% \, Inhibition = \left(1 - \frac{Average \, Treatment \, Count}{Average \, Untreated \, Control \, Count}\right) \times 100\%$$

$$\% \, Inhibition = \left(1 - \frac{1000}{2000}\right) \times 100\%$$

$$\% \, Inhibition = 50\%$$

"Generational" Method

$$\% \, Inhibition = \left(1 - \frac{doublings \, from \, basline \, under \, treatment}{doublings \, from \, basline \, without \, treatment}\right) \times 100\%$$

$$\% \, Inhibition = \left(1 - \frac{0}{1}\right) \times 100\%$$

$$\% \, Inhibition = 100\%$$

Figure 9: Comparison between "survival fraction" and "generational" methods for calculating percent inhibition for a hypothetical, fast growing cell line

Example: Fast Growing Cell Line
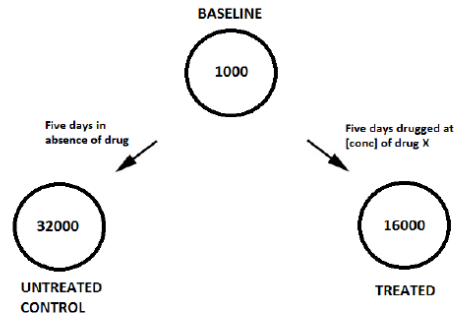
BASELINE

1000

Five days in
absence of drug

Five days drugged at
[conc] of drug X

32000

16000

UNTREATED
CONTROL

TREATED

"Survival Fraction" Method

$$\% \, Inhibition = \left(1 - \frac{Average \, Treatment \, Count}{Average \, Untreated \, Control \, Count}\right) \times 100\%$$

$$\% \, Inhibition = \left(1 - \frac{16000}{32000}\right) \times 100\%$$

$$\% \, Inhibition = 50\%$$

"Generational" Method

$$\% \, Inhibition = \left(1 - \frac{doublings \, from \, basline \, under \, treatment}{doublings \, from \, basline \, without \, treatment}\right) \times 100\%$$

$$\% \, Inhibition = \left(1 - \frac{4}{5}\right) \times 100\%$$

$$\% \, Inhibition = 25\%$$

Figure 10: Volcano plot of observed associations between presumptive activating point mutations and palbociclib response. Cutoff for sensitivity at IC50 = 100nM. Each circle represents one genotype-response association statistic. Size of circles represents frequency of alteration. DomActs: Dominant Activating Point Mutations.

85

Figure 13: Volcano plot of observed associations between chromosomally deleted (homozygous) genes and palbociclib response. Each circle represents one genotype-response association statistic. Size of circles represents frequency of alteration. Cutoff for sensitivity at IC50 = 365nM
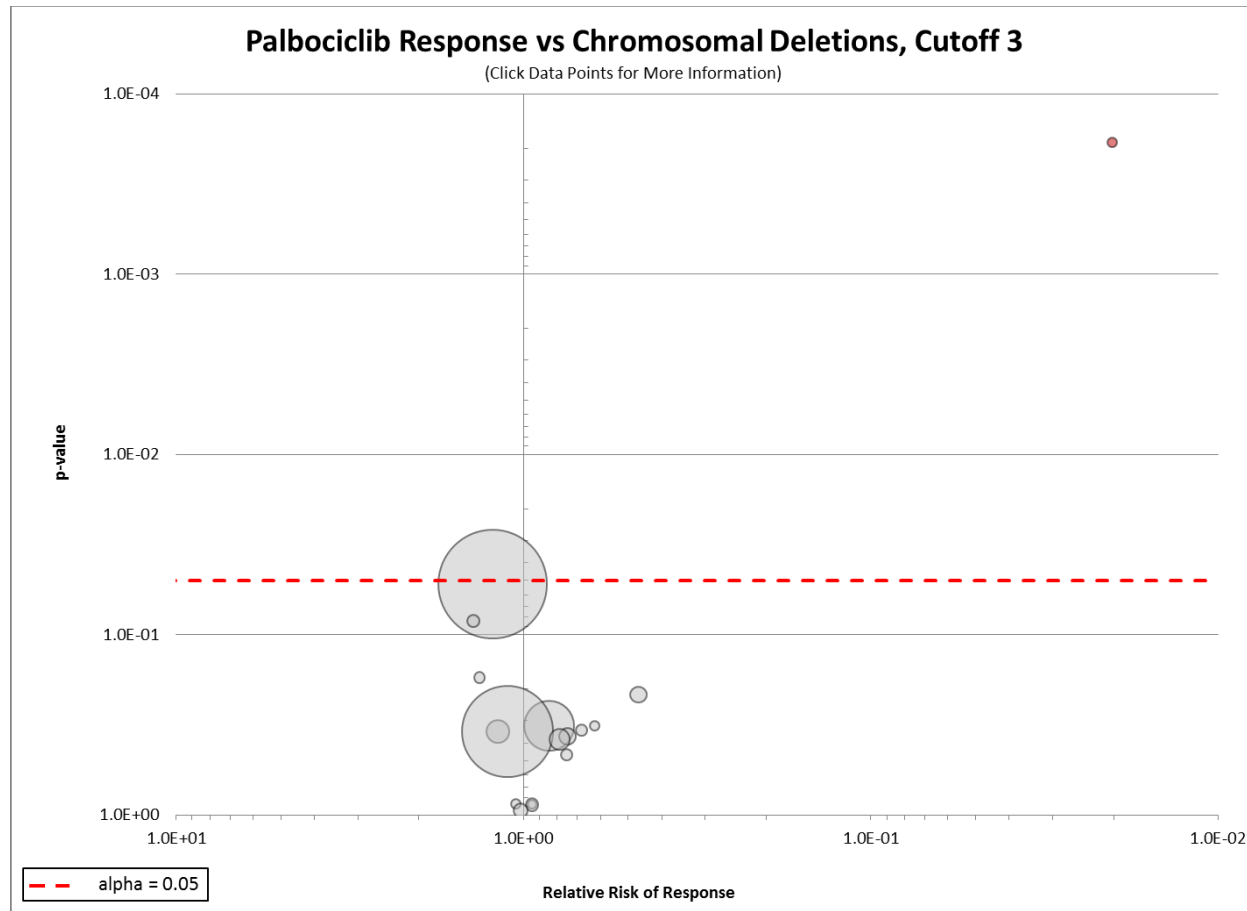
Table 1: Cell line characterization data coverage by cancer type. CNAs: Copy Number Alterations. CGH: Comparative Genomic Hybridization. Affy: Affymetrix. SNP: Single Nucleotide Polymorphism. UCLATORL: UCLA Translational Oncology Research Laboratory. CCLE: Cancer Cell Line Encyclopedia.

Data Type:	Palbociclib Response	CNAs	CNAs	mRNA Expression	Point Mutations
Assay Method:	5-day proliferation	CGH Chip	Affy SNP6.0 Array	Agilent 44K	Hybrid Capture
Data Source:	UCLATORL	UCLATORL	CCLE	UCLATORL	CCLE
Breast	47	47	39	47	36
Colon	28	23	23	27	21
Head and Neck	30	30	5	30	5
Kidney	29	25	7	27	7
Lung	54	48	50	53	48
Liver	20	20	20	20	16
Melanoma	47	44	15	42	15
Ovarian	40	39	22	40	22
Pancreas	29	29	27	29	21
Sarcoma	25	0	7	14	5
Upper GI	14	12	14	14	13
Lymphoma	53	0	33	53	31
Total:	**416**	**317**	**262**	**396**	**240**

Table 2: Significant associations from both point mutation datasets. LOF: Loss-Of-Function. ACT: Activating.

Gene	Frequency	Type of Alteration	Direction of Association	CUTOFF 1 IC50 < 100nM			CUTOFF 2 IC50 < Q3			CUTOFF 3 IC50 < Median		
				RR	95% CI	p-value	RR	95% CI	p-value	RR	95% CI	p-value
CDH1	13/240	LOF	Sensitivity	3.25	1.96 - 5.39	< 0.0001	2.71	1.77 - 4.14	< 0.0001	1.81	1.38 - 2.37	< 0.0001
TOPBP1	8/240	LOF	Sensitivity	1.18	0.35 - 4.03	0.79	1.36	0.54 - 3.41	0.51	1.85	1.37 - 2.48	< 0.0001
MLL	12/240	ACT	Resistance	0.04	0.005 - 0.28	0.002	0.29	0.04 - 1.9	0.20	0.85	0.43 - 1.68	0.63
TSHR	10/240	ACT	Resistance	0.05	0.006 - 0.35	0.003	0.35	0.05 - 2.26	0.27	0.81	0.38 - 1.76	0.60
SMO	9/240	ACT	Resistance	0.05	0.006 - 0.39	0.004	0.04	0.005 - 0.30	0.002	0.91	0.43 - 1.91	0.80
FANCA	12/240	LOF	Resistance	0.04	0.005 - 0.28	0.002	0.03	0.004 - 0.21	0.0006	0.50	0.19 - 1.34	0.17
NBN	9/240	LOF	Resistance	0.05	0.006 - 0.39	0.004	0.04	0.005 - 0.30	0.002	1.15	0.63 - 2.1	0.66
RB1	12/240	LOF	Resistance	0.04	0.005 - 0.28	0.002	0.58	0.16 - 2.11	0.41	0.50	0.19 - 1.34	0.17

Table 3: Significant associations from both copy number alteration datasets. AMP: Amplification. HD: Homozygous Deletion.

Chromosomal Region	Candidate Gene	Type of Alteration	Frequency	Direction of Association	CUTOFF 1 IC50 < 100nM			CUTOFF 2 IC50 < Q3			CUTOFF 3 IC50 < Median		
					RR	95% CI	p-value	RR	95% CI	p-value	RR	95% CI	p-value
17q12-21	ERBB2	AMP	25/317	Sensitivity	3.23	2.04 - 5.11	< 0.0001	2.50	1.71 - 3.66	< 0.0001	1.69	1.38 - 2.08	< 0.0001
11q13	CCND1	AMP	38/317	Sensitivity	2.45	1.52 - 3.94	0.0002	1.70	1.11 - 2.61	0.01	1.24	0.95 - 1.62	0.12
1q32	???	AMP	9/317	Sensitivity	3.11	1.65 - 5.85	0.0004	2.60	1.58 - 4.28	0.0002	1.51	1.05 - 2.17	0.03
19q13	CCNE1	AMP	14/317	Resistance	0.37	0.05 - 2.46	0.30	0.52	0.14 - 1.91	0.32	0.26	0.07 - 0.96	0.04
8q13	???	AMP	9/317	Resistance	0.06	0.007 - 0.44	0.006	0.82	0.24 - 2.84	0.76	0.63	0.24 - 1.60	0.33
13q14	RB1	HD	10/217	Resistance	0.05	0.007 - 0.39	0.004	0.04	0.005 - 0.28	0.001	0.02	0.002 - 0.14	0.0001

Table 4: Candidate predictive biomarkers implicated in previous literature. AMP: Amplification, HD: Homozygous Deletion, LOF: Loss-of-function, NS: Not Significant.

Chromosomal Region	Candidate Gene	Type of Alteration	Frequency	Direction of Association	CUTOFF 1 IC50 < 100nM			CUTOFF 2 IC50 < Q3			CUTOFF 3 IC50 < Median		
					RR	95% CI	p-value	RR	95% CI	p-value	RR	95% CI	p-value
12q13	CDK4	AMP	7/317	NS	0.75	0.12 - 4.67	0.76	1.07	0.33 - 3.49	0.91	1.09	0.57 - 2.10	0.79
7q21	CDK6	AMP	15/317	NS	0.34	0.05 - 2.30	0.27	0.49	0.13 - 1.79	0.28	1.15	0.75 - 1.77	0.51
9p21	CDKN2A	HD	99/317	NS	0.80	0.48 - 1.35	0.40	0.92	0.61 - 1.37	0.67	1.24	1.01 - 1.54	0.04
11q13	CCND1	AMP	38/317	Sensitivity	**2.45**	**1.52 - 3.94**	**0.0002**	**1.70**	**1.11 - 2.61**	**0.01**	1.24	0.95 - 1.62	0.12
13q14	RB1	HD	9/317	Resistance	**0.05**	**0.007 - 0.39**	**0.004**	**0.04**	**0.005 - 0.28**	**0.001**	**0.02**	**0.002 - 0.14**	**0.0001**
	RB1	LOF Point	12/240	Resistance	**0.04**	**0.005 - 0.28**	**0.002**	0.58	0.16 - 2.11	0.41	0.50	0.19 - 1.34	0.17

Table 5: Distribution of candidate predictive genomic alterations by primary site in the UCLATORL cell line panel and the Cancer Genome Atlas patient sample database.

TCGA: The Cancer Genome Atlas. AMP: Chromosomal Amplification. HD: Homozygous Deletion. DAPM: Dominant, Activating Point Mutation. RLOFPM: Recessive Loss-Of-Function Point Mutation.

| Gene: | CDH1 | | TOPBP1 | | ERBB2 | | CCND1 | | SMO | | CCNE1 | | RB1 | | RB1 | |
| Alteration: | RLOFPM | | RLOFPM | | AMP | | AMP | | DAPM | | AMP | | RLOFPM | | HD | |
	UCLA	TCGA	UCLA	TCGA	UCLA	TCGA	UCLA	TCGA	UCLA	TCGA	UCLA	TCGA	UCLA	TCGA	UCLA	TCGA
Breast	0.14	0.09	0.03	0.02	0.34	0.13	0.26	0.16	0.00	0.01	0.09	0.04	0.03	0.02	0.09	0.03
Colon	0.10	0.03	0.04	0.01	0.00	0.03	0.00	0.01	0.14	0.01	0.04	0.01	0.00	0.02	0.00	0.00
Head and Neck	0.00	0.03	0.00	0.02	0.00	0.03	0.40	0.27	0.00	0.00	0.03	0.02	0.00	0.03	0.00	0.01
Kidney	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.14	0.01	0.00	0.00	0.14	0.00	0.00	0.00
Liver	0.00	ND	0.06	ND	0.00	ND	0.20	0.09	0.13	0.01	0.00	0.01	0.00	ND	0.10	0.05
Lung	0.00	0.04	0.00	0.03	0.06	0.02	0.08	0.12	0.00	0.02	0.06	0.06	0.10	0.06	0.04	0.01
Lymphoma	0.00	ND	0.03	ND	ND	ND	ND	ND	0.00	ND	ND	ND	0.06	ND	ND	ND
Melanoma	0.00	0.03	0.13	0.01	0.00	0.00	0.05	0.06	0.07	0.04	0.02	0.01	0.07	0.04	0.00	0.01
Ovarian	0.00	0.06	0.00	0.01	0.08	0.03	0.05	0.07	0.05	0.00	0.08	0.21	0.05	0.03	0.05	0.09
Pancreas	0.00	ND	0.00	ND	0.03	0.00	0.00	0.00	0.00	ND	0.00	0.02	0.00	ND	0.00	0.00
Sarcoma	0.00	0.01	0.25	0.00	ND	0.00	ND	0.08	0.00	0.00	ND	0.00	0.00	0.01	ND	0.09
Upper GI	0.46	0.09	0.08	0.04	0.17	0.13	0.17	0.07	0.07	0.04	0.08	0.10	0.08	0.03	0.00	0.03

Table 6: Pearson correlation statistics for inter-correlations between candidate palbociclib response biomarkers. r: Pearson correlation coefficient, p: p-value, N: number of cell lines used to determine correlation.

		Sensitivity Biomarkers					Resistance Biomarkers								
		CDH1 RLOFPM	TOPBP1 RLOFPM	17q12-21 AMP	11q13 AMP	1q32 AMP	MLL DAPM	TSHR DAPM	SMO DAPM	FANCA RLOFPM	NBN RLOFPM	RB1 RLOFPM	19q13 AMP	8q13 AMP	13q14 HD
CDH1 RLOFPM	r	1	-0.04444	0.23875	0.08761	0.21678	0.02956	0.04222	0.04966	-0.0549	-0.04724	-0.0549	-0.06844	-0.04762	-0.05157
	p		0.4932	0.0008	0.2233	0.0023	0.6486	0.5151	0.4438	0.3971	0.4664	0.3971	0.3418	0.5086	0.474
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
TOPBP1 RLOFPM	r	-0.04444	1	0.04831	-0.06077	-0.02632	-0.0426	-0.03872	-0.03665	-0.0426	-0.03665	-0.0426	-0.04154	-0.0289	-0.0313
	p	0.4932		0.5025	0.3987	0.715	0.5113	0.5505	0.572	0.5113	0.572	0.5113	0.5642	0.6883	0.664
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
17q12-21 AMP	r	0.23875	0.04831	1	0.10821	0.09092	-0.01324	-0.07642	-0.07642	0.02189	0.00243	-0.07642	-0.00593	0.09092	-0.05281
	p	0.0008	0.5025		0.0543	0.1061	0.8543	0.2883	0.2883	0.7613	0.9731	0.2883	0.9162	0.1061	0.3487
	N	195	195	317	317	317	195	195	195	195	195	195	317	317	317
11q13 AMP	r	0.08761	-0.06077	0.10821	1	0.22929	-0.0916	-0.00801	-0.08241	0.01162	0.06638	-0.08241	-0.07933	0.05386	-0.06661
	p	0.2233	0.3987	0.0543		<.0001	0.2028	0.9115	0.2521	0.8719	0.3565	0.2521	0.1588	0.3391	0.237
	N	195	195	317	317	317	195	195	195	195	195	195	317	317	317
1q32 AMP	r	0.21678	-0.02632	0.09092	0.22929	1	-0.03966	-0.03568	-0.03568	-0.0313	-0.03568	-0.03568	0.0557	0.08514	-0.03085
	p	0.0023	0.715	0.1061	<.0001		0.582	0.6204	0.6204	0.664	0.6204	0.6204	0.3229	0.1304	0.5842
	N	195	195	317	317	317	195	195	195	195	195	195	317	317	317
MLL DAPM	r	0.02956	-0.0426	-0.01324	-0.0916	-0.03966	1	0.23918	0.05535	0.12281	-0.04528	0.03509	0.02988	0.08515	0.0723
	p	0.6486	0.5113	0.8543	0.2028	0.582		0.0002	0.3933	0.0575	0.485	0.5886	0.6784	0.2366	0.3152
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
TSHR DAPM	r	0.04222	-0.03872	-0.07642	-0.00801	-0.03568	0.23918	1	0.17835	0.14351	0.17835	0.14351	-0.05633	-0.03919	-0.04245
	p	0.5151	0.5505	0.2883	0.9115	0.6204	0.0002		0.0056	0.0262	0.0056	0.0262	0.4341	0.5864	0.5557
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
SMO DAPM	r	0.04966	-0.03665	-0.07642	-0.08241	-0.03568	0.05535	0.17835	1	0.05535	0.19192	-0.04528	-0.05633	-0.03919	-0.04245
	p	0.4438	0.572	0.2883	0.2521	0.6204	0.3933	0.0056		0.3933	0.0028	0.485	0.4341	0.5864	0.5557
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
FANCA RLOFPM	r	-0.0549	-0.0426	0.02189	0.01162	-0.0313	0.12281	0.14351	0.05535	1	0.05535	0.03509	0.06529	-0.03438	0.11094
	p	0.3971	0.5113	0.7613	0.8719	0.664	0.0575	0.0262	0.3933		0.5886	0.5886	0.3645	0.6333	0.1226
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
NBN RLOFPM	r	-0.04724	-0.03665	0.00243	0.06638	-0.03568	-0.04528	0.17835	0.19192	0.05535	1	-0.04528	-0.05633	-0.03919	-0.04245
	p	0.4664	0.572	0.9731	0.3565	0.6204	0.485	0.0056	0.0028	0.3933		0.485	0.4341	0.5864	0.5557
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
RB1 RLOFPM	r	-0.0549	-0.0426	-0.07642	-0.08241	-0.03568	0.03509	0.14351	-0.04528	0.03509	-0.04528	1	-0.05633	-0.03919	-0.04245
	p	0.3971	0.5113	0.2883	0.2521	0.6204	0.5886	0.0262	0.485	0.5886	0.485		0.4341	0.5864	0.5557
	N	240	240	195	195	195	240	240	240	240	240	240	195	195	195
19q13 AMP	r	-0.06844	-0.04154	-0.00593	-0.07933	0.0557	0.02988	-0.05633	-0.05633	0.06529	-0.05633	-0.05633	1	0.14814	0.04905
	p	0.3418	0.5642	0.9162	0.1588	0.3229	0.6784	0.4341	0.4341	0.3645	0.4341	0.4341		0.0082	0.3841
	N	195	195	317	317	317	195	195	195	195	195	195	317	317	317
8q13 AMP	r	-0.04762	-0.0289	0.09092	0.05386	0.08514	0.08515	-0.03919	-0.03919	-0.03438	-0.03919	-0.03919	0.14814	1	0.07781
	p	0.5086	0.6883	0.1061	0.3391	0.1304	0.2366	0.5864	0.5864	0.6333	0.5864	0.5864	0.0082		0.167
	N	195	195	317	317	317	195	195	195	195	195	195	317	317	317
RB1 HD	r	-0.05157	-0.0313	-0.05281	-0.06661	-0.03085	0.0723	-0.04245	-0.04245	0.11094	-0.04245	-0.04245	0.04905	0.07781	1
	p	0.474	0.664	0.3487	0.237	0.5842	0.3152	0.5557	0.5557	0.1226	0.5557	0.5557	0.3841	0.167	
	N	195	195	317	317	317	195	195	195	195	195	195	317	317	317

Table 7: Final set of candidate biomarkers of palbociclib sensitivity and resistance. AMP: Chromosomal Amplification. HD: Homozygous Deletion. DAPM: Dominant, Activating Point Mutation. RLOFPM: Recessive Loss-Of-Function Point Mutation.

SENSITIVITY				RESISTANCE			
DAPMS	RLOFPMS	AMPS	HDS	DAPMS	LOFPOINTS	AMPS	HDS
	CDH1	ERBB2		SMO	RB1	CCNE1	RB1
	TOPBP1	CCND1					

Table 8: Genotype-response associations for "any sensitivity biomarker" by cancer type. RR: Response Ratio. DNC: Regression Model Did Not Converge.

Cancer Type	Frequency	CUTOFF 1 IC50 < 100nM			CUTOFF 2 IC50 < Q3			CUTOFF 3 IC50 < Median		
		RR	95% CI	p-value	RR	95% CI	p-value	RR	95% CI	p-value
Breast	23/47	3.07	1.34 - 6.98	0.01	2.46	1.27 - 4.75	0.01	1.62	1.11 - 2.36	0.01
Colon	3/25	1.83	0.34 - 10.44	0.52	1.47	0.29 - 7.89	0.67	1.05	0.45 - 2.36	0.92
Head and Neck	12/30	1.50	0.37 - 6.20	0.57	0.86	0.33 - 2.29	0.76	1.25	0.83 - 1.87	0.29
Kidney	0/25	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Liver	5/20	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Lung	7/52	1.84	0.51 - 6.94	0.38	1.29	0.38 - 4.56	0.70	1.07	0.44 - 2.66	0.88
Lymphoma	1/31	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Melanoma	4/45	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Ovarian	5/39	1.70	0.28 - 11.39	0.60	1.70	0.28 - 11.40	0.60	1.70	0.52 - 5.64	0.40
Pancreas	1/29	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Sarcoma	1/5	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Upper GI	8/14	1.50	0.40 - 5.61	0.55	1.25	0.48 - 3.24	0.65	1.31	0.70 - 2.42	2.45
Overall	**71/362**	**2.63**	**1.75 - 3.96**	**< 0.0001**	**2.02**	**1.44 - 2.82**	**< 0.0001**	**1.56**	**1.29 - 1.89**	**< 0.0001**

Table 9: Genotype-response associations for "any resistance biomarker" by cancer type. RR: Response Ratio. DNC: Regression Model Did Not Converge.

Cancer Type	Frequency	CUTOFF 1 IC50 < 100nM			CUTOFF 2 IC50 < Q3			CUTOFF 3 IC50 < Median		
		RR	95% CI	p-value	RR	95% CI	p-value	RR	95% CI	p-value
Breast	8/47	0.02	0.003 - 0.18	< 0.01	0.22	0.04 - 1.18	0.08	0.15	0.03 - 0.82	0.03
Colon	4/25	0.10	0.01 - 0.96	0.05	0.08	0.009 - 0.78	0.03	0.75	0.28 - 2.00	0.57
Head and Neck	1/30	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Kidney	2/25	0.26	0.02 - 2.96	0.28	0.15	0.01 - 1.55	0.11	0.67	0.18 - 2.46	0.55
Liver	4/20	DNC	DNC	DNC	DNC	DNC	DNC	0.13	0.01 - 1.35	0.01
Lung	10/52	0.05	0.006 - 0.38	< 0.01	0.03	0.004 - 0.28	< 0.01	0.02	0.002 - 0.15	< 0.01
Lymphoma	2/31	0.66	0.05 - 9.09	0.76	0.33	0.03 - 3.76	0.37	1.61	0.41 - 6.39	0.40
Melanoma	3/45	2.41	0.05 - 12.38	0.30	1.20	0.26 - 5.60	0.81	0.53	0.12 - 2.35	0.41
Ovarian	7/32	0.09	0.01 - 0.81	0.03	0.09	0.01 - 0.81	0.01	0.54	0.09 - 3.12	0.09
Pancreas	0/29	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Sarcoma	0/5	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
Upper GI	3/11	0.06	0.006 - 0.54	0.01	0.54	0.12 - 2.49	0.43	0.38	0.09 - 1.65	0.20
Overall	**44/362**	**0.11**	**0.02 - 0.75**	**0.02**	**0.32**	**0.12 - 0.80**	**0.02**	**0.36**	**0.20-0.66**	**0.001**

Table 10: Crude and adjusted genotype response associations for the collapsed biomarker variables. Adjusted by growth rate and cancer type.

	CUTOFF 1 IC50 < 100nM		CUTOFF 2 IC50 < Q3		CUTOFF 3 IC50 < Median	
	Crude (95% CI)	Adj (95% CI)	Crude (95% CI)	Adj (95% CI)	Crude (95% CI)	Adj (95% CI)
Any Sensitivity Biomarker	2.63 (1.75-3.96)	1.38 (0.86-2.24)	2.02 (1.44-2.82)	1.15 (0.79-1.68)	1.56 (1.29-1.89)	1.21 (0.96-1.51)
Any Resistance Biomarker	0.11 (0.02-0.75)	0.24 (0.08-0.68)	0.32 (0.12-0.80)	0.42 (0.20-0.88)	0.36 (0.20-0.66)	0.42 (0.25-0.72)

Table 11: Kappa test of inter-rater agreement for genotyping calls in UCLA and CCLE copy number alteration datasets. CAN: Copy Number Alteration. Amp: Amplification. HD: Homozygous Deletion. CCLE: Cancer Cell Line Encyclopedia.

CNA	# Cell Lines Altered in UCLA Dataset	# Cell Lines Altered in CCLE Dataset	Overlapping	Kappa
CCND1 Amp	25	21	20	0.87
CCNE1 Amp	12	6	6	0.65
ERBB2 Amp	21	20	19	0.92
RB1 HD	7	6	6	0.92

Supplemental tables:

Supplemental Table S1 (attached as excel file): Palbociclib response data for all UCLATORL cell lines. IC values are the geometric mean of duplicate runs.

Supplemental Table S2 (attached as excel file): List of candidate oncogenes included in all four genomics datasets used in genotype-response association testing. AMP: Chromosomal Amplification, HD: Chromosomal Homozygous Deletion, DAPM: Dominant Activating Point Mutation, RLOFPM: Recessive Loss-Of-Function Point Mutation.

Supplemental Table S3: Association between growth rate and palbociclib sensitivity by cancer type.

PrimarySite	CUTOFF 1 IC50 < 100nM		CUTOFF 2 IC50 < Q3		CUTOFF 3 IC50 < Median	
	RR	95% CI	RR	95% CI	RR	95% CI
Breast	0.55	0.39 -0 .78	0.64	0.49 - 0.83	0.88	0.75 - 1.04
Colon	0.43	0.24 - 0.76	0.59	0.33 - 1.06	0.94	0.75 - 1.19
Head and Neck	0.23	0.10 - 0.50	0.52	0.29 - 0.93	0.82	0.58 - 1.16
Kidney	0.46	0.23 - 0.92	0.42	0.27 - 0.65	0.72	0.53 - 0.97
Liver	DNC	DNC	3.27	1.67 - 6.39	0.89	0.40 - 1.99
Lung	0.38	0.20 - 0.73	0.51	0.23 - 0.88	0.63	0.47 - 0.85
Lymphoma	0.29	0.14 - 0.57	0.43	0.22 - 0.82	0.65	0.41 - 1.02
Melanoma	0.54	0.34 - 0.85	0.67	0.47 - 0.96	0.78	0.66 - 0.91
Ovarian	0.61	0.32 - 1.17	0.61	0.32 - 1.17	0.99	0.61 - 1.59
Pancreas	0.35	0.16 - 0.77	0.35	0.16 - 0.77	0.47	0.30 - 0.74
Sarcoma	DNC	DNC	0.25	0.12 - 0.52	0.25	0.12 - 0.52
Upper GI	0.51	0.22 - 1.18	0.7	0.38 - 1.29	0.81	0.56 - 1.17
Total:	**0.48**	**0.41 - 0.56**	**0.59**	**0.52 - 0.68**	**0.78**	**0.72 - 0.84**

Supplemental Table S4: Frequency of CDKN2A deletion in UCLATORL and TCGA datasets

	CDKN2A	
Gene:		
Alteration:	HD	
	UCLA	TCGA
Breast	0.20	0.04
Colon	0.00	0.02
Head and Neck	0.44	0.28
Kidney	0.59	0.03
Liver	0.15	0.03
Lung	0.36	0.20
Lymphoma	ND	ND
Melanoma	0.45	0.31
Ovarian	0.23	0.06
Pancreas	0.48	0.22
Sarcoma	ND	0.15
Upper GI	0.46	0.10

References

1   Kaplan, H. S. Basic principles in radiation oncology. *Cancer* **39**, 689-693 (1977).

2   Morgan, G., Ward, R. & Barton, M. The contribution of cytotoxic chemotherapy to 5-year survival in adult malignancies. *Clinical oncology* **16**, 549-560 (2004).

3   Sun, Y. *et al.* Treatment-induced damage to the tumor microenvironment promotes prostate cancer therapy resistance through WNT16B. *Nature medicine* **18**, 1359-1368, doi:10.1038/nm.2890 (2012).

4   Croce, C. M. Oncogenes and cancer. *The New England journal of medicine* **358**, 502-511, doi:10.1056/NEJMra072367 (2008).

5   Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

6   Hainaut, P. & Plymoth, A. Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. *Current opinion in oncology* **25**, 50-51, doi:10.1097/CCO.0b013e32835b651e (2013).

7   Barros Costa, R. L. Targeted therapy: comprehensive review. *The American journal of hospice & palliative care* **26**, 137-146, doi:10.1177/1049909108330026 (2009).

8   Cobleigh, M. A. *et al.* Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *Journal of Clinical Oncology* **17**, 2639-2648 (1999).

9   Druker, B. J. *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *The New England journal of medicine* **344**, 1031-1037, doi:10.1056/NEJM200104053441401 (2001).

10   Sequist, L. V. First-generation epidermal growth factor receptor tyrosine kinase inhibitors in EGFR mutation: positive non-small cell lung cancer patients. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* **3**, S143-145, doi:10.1097/JTO.0b013e318174e981 (2008).

11   West, M., Ginsburg, G. S., Huang, A. T. & Nevins, J. R. Embracing the complexity of genomic data for personalized medicine. *Genome research* **16**, 559-566, doi:10.1101/gr.3851306 (2006).

12   Vollan, H. K. & Caldas, C. The breast cancer genome--a key for better oncology. *BMC cancer* **11**, 501, doi:10.1186/1471-2407-11-501 (2011).

13   Vineis, P. & Perera, F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **16**, 1954-1965, doi:10.1158/1055-9965.EPI-07-0457 (2007).

14   van't Veer, L. J. & Bernards, R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* **452**, 564-570, doi:10.1038/nature06915 (2008).

15   Toft, D. J. & Cryns, V. L. Minireview: Basal-like breast cancer: from molecular profiles to targeted therapies. *Molecular endocrinology* **25**, 199-211, doi:10.1210/me.2010-0164 (2011).

16   Sriram, K. B., Larsen, J. E., Yang, I. A., Bowman, R. V. & Fong, K. M. Genomic medicine in non-small cell lung cancer: paving the path to personalized care. *Respirology* **16**, 257-263, doi:10.1111/j.1440-1843.2010.01892.x (2011).

17   Sharma, S. V. & Settleman, J. Oncogene addiction: setting the stage for molecularly targeted cancer therapy. *Genes & development* **21**, 3214-3231, doi:10.1101/gad.1609907 (2007).

18   Kim, E. S. *et al.* The BATTLE trial: personalizing therapy for lung cancer. *Cancer discovery* **1**, 44-53, doi:10.1158/2159-8274.CD-10-0010 (2011).

19   Janne, P. A., Gray, N. & Settleman, J. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nature reviews. Drug discovery* **8**, 709-723, doi:10.1038/nrd2871 (2009).

20   Gonzalez-Angulo, A. M., Hennessy, B. T. & Mills, G. B. Future of personalized medicine in oncology: a systems biology approach. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 2777-2783, doi:10.1200/JCO.2009.27.0777 (2010).

21   Linardou, H. *et al.* Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. *Lancet Oncology* **9**, 962-972, doi:Doi 10.1016/S1470-2045(08)70206-7 (2008).

22   Cizkova, M. *et al.* Outcome impact of PIK3CA mutations in HER2-positive breast cancer patients treated with trastuzumab. *British journal of cancer* **108**, 1807-1809, doi:10.1038/bjc.2013.164 (2013).

23   Berns, K. *et al.* A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer cell* **12**, 395-402, doi:10.1016/j.ccr.2007.08.030 (2007).

24   Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nature reviews. Cancer* **6**, 813-823, doi:10.1038/nrc1951 (2006).

25   Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nature communications* **4**, 2126, doi:10.1038/ncomms3126 (2013).

26   Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).

27   Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research* **41**, D955-961, doi:10.1093/nar/gks1111 (2013).

28   Massague, J. G1 cell-cycle control and cancer. *Nature* **432**, 298-306, doi:10.1038/nature03094 (2004).

29   Shapiro, G. I. *et al.* Multiple mechanisms of p16INK4A inactivation in non-small cell lung cancer cell lines. *Cancer research* **55**, 6200-6209 (1995).

30   Wolfel, T. *et al.* A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science* **269**, 1281-1284 (1995).

31   Schmidt, E. E., Ichimura, K., Reifenberger, G. & Collins, V. P. CDKN2 (p16/MTS1) gene deletion or CDK4 amplification occurs in the majority of glioblastomas. *Cancer research* **54**, 6321-6324 (1994).

32   Sotillo, R. *et al.* Wide spectrum of tumors in knock-in mice carrying a Cdk4 protein insensitive to INK4 inhibitors. *Embo J* **20**, 6637-6647, doi:10.1093/emboj/20.23.6637 (2001).

33   Wang, M. B. *et al.* Detection of chromosome 11q13 amplification in head and neck cancer using fluorescence in situ hybridization. *Anticancer research* **19**, 925-931 (1999).

34   Moreno-Bueno, G. *et al.* Cyclin D1 gene (CCND1) mutations in endometrial cancer. *Oncogene* **22**, 6115-6118, doi:10.1038/sj.onc.1206868 (2003).

35   Zaharieva, B. M. *et al.* High-throughput tissue microarray analysis of 11q13 gene amplification (CCND1, FGF3, FGF4, EMS1) in urinary bladder cancer. *The Journal of pathology* **201**, 603-608, doi:10.1002/path.1481 (2003).

36   Roy, P. G. *et al.* High CCND1 amplification identifies a group of poor prognosis women with estrogen receptor positive breast cancer. *International journal of cancer. Journal international du cancer* **127**, 355-360, doi:10.1002/ijc.25034 (2010).

37   Al-Kuraya, K. *et al.* HER2, TOP2A, CCND1, EGFR and C-MYC oncogene amplification in colorectal cancer. *Journal of clinical pathology* **60**, 768-772, doi:10.1136/jcp.2006.038281 (2007).

38   Adelaide, J., Monges, G., Derderian, C., Seitz, J. F. & Birnbaum, D. Oesophageal cancer and amplification of the human cyclin D gene CCND1/PRAD1. *British journal of cancer* **71**, 64-68 (1995).

39   Xu, H. J., Cairns, P., Hu, S. X., Knowles, M. A. & Benedict, W. F. Loss of RB protein expression in primary bladder cancer correlates with loss of heterozygosity at the RB locus and tumor progression. *International journal of cancer. Journal international du cancer* **53**, 781-784 (1993).

40   Takano, H., Okamoto, A., Terashima, Y. & Yokota, J. High-incidence of allelic loss at the rb gene locus in advanced human ovarian-cancer. *Int J Oncol* **6**, 129-135 (1995).

41   Venesio, T. *et al.* Loss of retinoblastoma gene (RB1) is associated with deletions at the 17p13.3 chromosome and S-phase index in human breast cancer. *Annals of the New York Academy of Sciences* **784**, 462-466 (1996).

42   Latil, A. *et al.* Loss of heterozygosity at chromosome arm 13q and RB1 status in human prostate cancer. *Human pathology* **30**, 809-815 (1999).

43   Bieche, I. & Lidereau, R. Loss of heterozygosity at 13q14 correlates with RB1 gene underexpression in human breast cancer. *Molecular carcinogenesis* **29**, 151-158 (2000).

44   Lapenna, S. & Giordano, A. Cell cycle kinases as therapeutic targets for cancer. *Nature reviews. Drug discovery* **8**, 547-566, doi:10.1038/nrd2907 (2009).

45   Sedlacek, H. H. Mechanisms of action of flavopiridol. *Critical reviews in oncology/hematology* **38**, 139-170 (2001).

46   Christian, B. A., Grever, M. R., Byrd, J. C. & Lin, T. S. Flavopiridol in the treatment of chronic lymphocytic leukemia. *Current opinion in oncology* **19**, 573-578, doi:10.1097/CCO.0b013e3282efb9da (2007).

47   Joshi, K. S. *et al.* In vitro antitumor properties of a novel cyclin-dependent kinase inhibitor, P276-00. *Molecular cancer therapeutics* **6**, 918-925, doi:10.1158/1535-7163.MCT-06-0613 (2007).

48   Joshi, K. S. *et al.* P276-00, a novel cyclin-dependent inhibitor induces G1-G2 arrest, shows antitumor activity on cisplatin-resistant cells and significant in vivo efficacy in tumor models. *Molecular cancer therapeutics* **6**, 926-934, doi:10.1158/1535-7163.MCT-06-0614 (2007).

49   Zhang, C. C. e. a. in *Proc Amer. Assoc. Cancer Res.*

50   Brown, A. P. *et al.* Toxicity and toxicokinetics of the cyclin-dependent kinase inhibitor AG-024322 in cynomolgus monkeys following intravenous infusion. *Cancer chemotherapy and pharmacology* **62**, 1091-1101, doi:10.1007/s00280-008-0771-1 (2008).

51   Pfizer. *A Phase 1 Study Of An Intravenously Administered Cyclin-Dependent Kinase Inhibitor In Patients With Advanced Cancer*, <http://clinicaltrials.gov/show/NCT00147485> (2008).

52   DePinto, W. *et al.* In vitro and in vivo activity of R547: a potent and selective cyclin-dependent kinase inhibitor currently in phase I clinical trials. *Molecular cancer therapeutics* **5**, 2644-2658, doi:10.1158/1535-7163.MCT-06-0355 (2006).

53   Siemeister, G. *et al.* Molecular and pharmacodynamic characteristics of the novel multi-target tumor growth inhibitor ZK 304709. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **60**, 269-272, doi:10.1016/j.biopha.2006.06.003 (2006).

54   Squires, M. S. *et al.* Biological characterization of AT7519, a small-molecule inhibitor of cyclin-dependent kinases, in human tumor cell lines. *Molecular cancer therapeutics* **8**, 324-332, doi:10.1158/1535-7163.MCT-08-0890 (2009).

55   Scott, E. N. *et al.* A phase I dose escalation study of the pharmacokinetics and tolerability of ZK 304709, an oral multi-targeted growth inhibitor (MTGI), in patients with advanced solid tumours. *Cancer chemotherapy and pharmacology* **64**, 425-429, doi:10.1007/s00280-009-0968-y (2009).

56   Fry, D. W. *et al.* Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts. *Molecular cancer therapeutics* **3**, 1427-1438 (2004).

57   Finn, R. S. *et al.* PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast cancer research : BCR* **11**, R77, doi:10.1186/bcr2419 (2009).

58   Konecny, G. E. *et al.* Expression of p16 and retinoblastoma determines response to CDK4/6 inhibition in ovarian cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 1591-1602, doi:10.1158/1078-0432.CCR-10-2307 (2011).

59   Katsumi, Y. *et al.* Sensitivity of malignant rhabdoid tumor cell lines to PD 0332991 is inversely correlated with p16 expression. *Biochemical and biophysical research communications* **413**, 62-68, doi:10.1016/j.bbrc.2011.08.047 (2011).

60   Leonard, J. P. *et al.* Selective CDK4/6 inhibition with tumor responses by PD0332991 in patients with mantle cell lymphoma. *Blood* **119**, 4597-4607, doi:10.1182/blood-2011-10-388298 (2012).

61   Cen, L. *et al.* p16-Cdk4-Rb axis controls sensitivity to a cyclin-dependent kinase inhibitor PD0332991 in glioblastoma xenograft cells. *Neuro-oncology* **14**, 870-881, doi:10.1093/neuonc/nos114 (2012).

62   Menu, E. *et al.* A novel therapeutic combination using PD 0332991 and bortezomib: study in the 5T33MM myeloma model. *Cancer research* **68**, 5519-5523, doi:10.1158/0008-5472.CAN-07-6404 (2008).

63   Flaherty, K. T. *et al.* Phase I, dose-escalation trial of the oral cyclin-dependent kinase 4/6 inhibitor PD 0332991, administered using a 21-day schedule in patients with advanced cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 568-576, doi:10.1158/1078-0432.CCR-11-0509 (2012).

64   RS Finn, J. C., I Lang, K Boer, IM Bondarenko, SO Kulyk, J Ettl, R Patel, T Pinter, M Schmidt, Y Shparyk, AR Thummala, NL Voytko, A Breazna, ST Kim, S Randolph, and DJ Slamon. Results of a randomized phase 2 study of PD 0332991, a cyclin-dependent kinase (CDK) 4/6 inhibitor, in combination with letrozole vs letrozole alone for first-line treatment of ER+/HER2– advanced breast cancer *Cancer research* **72**, S1-6, doi:10.1158/0008-5472.SABCS12-S1-6 (2012).

65   Korch, C. *et al.* DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. *Gynecologic oncology* **127**, 241-248, doi:10.1016/j.ygyno.2012.06.017 (2012).

66   Schweppe, R. E. Thyroid cancer cell line misidentification: an update. *The Journal of clinical endocrinology and metabolism* **98**, 956-957, doi:10.1210/jc.2012-4182 (2013).

67   Dirks, W. G. & Drexler, H. G. Online verification of human cell line identity by STR DNA typing. *Methods in molecular biology* **731**, 45-55, doi:10.1007/978-1-61779-080-5_5 (2011).

68   Boonstra, J. J. *et al.* Mistaken identity of widely used esophageal adenocarcinoma cell line TE-7. *Cancer research* **67**, 7996-8001, doi:10.1158/0008-5472.CAN-07-2064 (2007).

69   American Type Culture Collection Standards Development Organization Workgroup, A. S. N. Cell line misidentification: the beginning of the end. *Nature reviews. Cancer* **10**, 441-448, doi:10.1038/nrc2852 (2010).

70   Lundin, A., Hasenson, M., Persson, J. & Pousette, A. Estimation of biomass in growing cell lines by adenosine triphosphate assay. *Methods in enzymology* **133**, 27-42 (1986).

71   Kangas, L., Gronroos, M. & Nieminen, A. L. Bioluminescence of cellular ATP: a new method for evaluating cytotoxic agents in vitro. *Medical biology* **62**, 338-343 (1984).

72   Zou, G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol* **159**, 702-706, doi:10.1093/aje/kwh090 (2004).

73   Greenland, S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol* **160**, 301-305, doi:10.1093/aje/kwh221 (2004).

74   Benjamini, Y., Krieger, A. M. & Yekutieli, D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491-507, doi:DOI 10.1093/biomet/93.3.491 (2006).

75   Futreal, P. A. *et al.* A census of human cancer genes. *Nature reviews. Cancer* **4**, 177-183, doi:10.1038/nrc1299 (2004).

76   Santarius, T., Shipley, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nature reviews. Cancer* **10**, 59-64, doi:10.1038/nrc2771 (2010).

77   Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893-898, doi:10.1038/nature08768 (2010).

78   Collins, V. P. Gene amplification in human gliomas. *Glia* **15**, 289-296, doi:10.1002/glia.440150309 (1995).

79   Pires de Camargo, V. *et al.* Other targetable sarcomas. *Seminars in oncology* **36**, 358-371, doi:10.1053/j.seminoncol.2009.06.008 (2009).

80   An, H. X., Beckmann, M. W., Reifenberger, G., Bender, H. G. & Niederacher, D. Gene amplification and overexpression of CDK4 in sporadic breast carcinomas is associated with high tumor cell proliferation. *The American journal of pathology* **154**, 113-118, doi:10.1016/S0002-9440(10)65257-1 (1999).

81   Ismail, A. *et al.* Early G(1) cyclin-dependent kinases as prognostic markers and potential therapeutic targets in esophageal adenocarcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 4513-4522, doi:10.1158/1078-0432.CCR-11-0244 (2011).

82   Tang, L. H. *et al.* Attenuation of the retinoblastoma pathway in pancreatic neuroendocrine tumors due to increased cdk4/cdk6. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 4612-4620, doi:10.1158/1078-0432.CCR-11-3264 (2012).

83   Dickson, M. A. *et al.* Phase II Trial of the CDK4 Inhibitor PD0332991 in Patients With Advanced CDK4-Amplified Well-Differentiated or Dedifferentiated Liposarcoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**, 2024-2028, doi:10.1200/JCO.2012.46.5476 (2013).

84   Pennsylvania, A. C. C. o. t. U. o. *PD 0332991 in Treating Patients With Refractory Solid Tumors*, <http://clinicaltrials.gov/show/NCT01037790> (2013).

85   Nagel, S. *et al.* Amplification at 7q22 targets cyclin-dependent kinase 6 in T-cell lymphoma. *Leukemia* **22**, 387-392, doi:10.1038/sj.leu.2405028 (2008).

86   Costello, J. F. *et al.* Cyclin-dependent kinase 6 (CDK6) amplification in human gliomas identified using two-dimensional separation of genomic DNA. *Cancer research* **57**, 1250-1254 (1997).

87   Tsai, J. W. *et al.* Recurrent amplification at 7q21.2 Targets CDK6 gene in primary myxofibrosarcomas and identifies CDK6 overexpression as an independent adverse prognosticator. *Annals of surgical oncology* **19**, 2716-2725, doi:10.1245/s10434-012-2317-3 (2012).

88   Elsheikh, S. *et al.* CCND1 amplification and cyclin D1 expression in breast cancer and their relation with proteomic subgroups and patient outcome. *Breast cancer research and treatment* **109**, 325-335, doi:10.1007/s10549-007-9659-8 (2008).

89   Dimova, I., Raicheva, S., Dimitrov, R., Doganov, N. & Toncheva, D. Coexistence of copy number increases of c-Myc, ZNF217, CCND1, ErbB1 and ErbB2 in ovarian cancers. *Onkologie* **32**, 405-410, doi:10.1159/000219368 (2009).

90   Hanken, H. *et al.* CCND1 amplification and cyclin D1 immunohistochemical expression in head and neck squamous cell carcinomas. *Clinical oral investigations*, doi:10.1007/s00784-013-0967-6 (2013).

91   Bizari, L. *et al.* Alterations of the CCND1 and HER-2/neu (ERBB2) proteins in esophageal and gastric cancers. *Cancer genetics and cytogenetics* **165**, 41-50, doi:10.1016/j.cancergencyto.2005.08.031 (2006).

92   Feng, Z. *et al.* CCND1 as a predictive biomarker of neoadjuvant chemotherapy in patients with locally advanced head and neck squamous cell carcinoma. *PloS one* **6**, e26399, doi:10.1371/journal.pone.0026399 (2011).

93   Geradts, J. & Wilson, P. A. High frequency of aberrant p16(INK4A) expression in human breast cancer. *The American journal of pathology* **149**, 15-20 (1996).

94   Miracca, E. C., Kowalski, L. P. & Nagai, M. A. High prevalence of p16 genetic alterations in head and neck tumours. *British journal of cancer* **81**, 677-683, doi:10.1038/sj.bjc.6690747 (1999).

95   Xing, X. *et al.* The prognostic value of CDKN2A hypermethylation in colorectal cancer: a meta-analysis. *British journal of cancer*, doi:10.1038/bjc.2013.251 (2013).

96   Kim, M. *et al.* Homozygous deletion of CDKN2A (p16, p14) and CDKN2B (p15) genes is a poor prognostic factor in adult but not in childhood B-lineage acute lymphoblastic leukemia: a comparative deletion and hypermethylation study. *Cancer genetics and cytogenetics* **195**, 59-65, doi:10.1016/j.cancergencyto.2009.06.013 (2009).

97   Center, J. C. C. *A Open Label Phase II Study of the Efficacy and Safety of PD0332991 a Selective Inhibitor of the Cyclin Dependent Kinases 4 and 6 in Patients With Recurrent Ovarian Cancer Demonstrating Rb-proficiency and Low p16 Expression*, <http://www.clinicaltrials.gov/ct2/show/NCT01536743> (2013).

98   Valverde, J. R., Alonso, J., Palacios, I. & Pestana, A. RB1 gene mutation up-date, a meta-analysis based on 932 reported mutations available in a searchable database. *BMC genetics* **6**, 53, doi:10.1186/1471-2156-6-53 (2005).

99   Network, T. *The Cancer Genome Atlas Data Portal*, <http://www.nature.com/ncomms/2013/130709/ncomms3126/pdf/ncomms3126.pdf> (2013).

100   Ormandy, C. J., Musgrove, E. A., Hui, R., Daly, R. J. & Sutherland, R. L. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. *Breast cancer research and treatment* **78**, 323-335 (2003).

101   Karlsson, E. *et al.* High-resolution genomic analysis of the 11q13 amplicon in breast cancers identifies synergy with 8p12 amplification, involving the mTOR targets S6K2 and 4EBP1. *Genes, chromosomes & cancer* **50**, 775-787, doi:10.1002/gcc.20900 (2011).

102   Huang, X., Godfrey, T. E., Gooding, W. E., McCarty, K. S., Jr. & Gollin, S. M. Comprehensive genome and transcriptome analysis of the 11q13 amplicon in human oral cancer and synteny to the 7F5 amplicon in murine oral carcinoma. *Genes, chromosomes & cancer* **45**, 1058-1069, doi:10.1002/gcc.20371 (2006).

103   Meredith, S. D. *et al.* Chromosome 11q13 amplification in head and neck squamous cell carcinoma. Association with poor prognosis. *Archives of otolaryngology--head & neck surgery* **121**, 790-794 (1995).

104   Gray-Bablin, J. *et al.* Cyclin E, a redundant cyclin in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 15215-15220 (1996).

105   Courjal, F. *et al.* Cyclin gene amplification and overexpression in breast and ovarian cancers: evidence for the selection of cyclin D1 in breast and cyclin E in ovarian tumors. *International journal of cancer. Journal international du cancer* **69**, 247-253, doi:10.1002/(SICI)1097-0215(19960822)69:4<247::AID-IJC1>3.0.CO;2-X (1996).

106   Natrajan, R. *et al.* Functional characterization of the 19q12 amplicon in grade III breast cancers. *Breast cancer research : BCR* **14**, R53, doi:10.1186/bcr3154 (2012).

107   Nakayama, N. *et al.* Gene amplification CCNE1 is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* **116**, 2621-2634, doi:10.1002/cncr.24987 (2010).

108   Cassia, R. *et al.* Cyclin E gene (CCNE) amplification and hCDC4 mutations in endometrial carcinoma. *The Journal of pathology* **201**, 589-595, doi:10.1002/path.1474 (2003).

109   Hayashi, H. *et al.* High cyclin E and low p27/Kip1 expressions are potentially poor prognostic factors in lung adenocarcinoma patients. *Lung cancer* **34**, 59-65 (2001).

110   Richter, J. *et al.* High-Throughput Tissue Microarray Analysis of Cyclin E Gene Amplification and Overexpression in Urinary Bladder Cancer. *The American journal of pathology* **157**, 787-794, doi:10.1016/s0002-9440(10)64592-0 (2000).

111   Lin, L. *et al.* Identification and characterization of a 19q12 amplicon in esophageal adenocarcinomas reveals Cyclin E as the best candidate gene for this amplicon. *Cancer research* **60**, 7021-7027 (2000).

112   Leung, S. Y. *et al.* Comprehensive analysis of 19q12 amplicon in human gastric cancers. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **19**, 854-863, doi:10.1038/modpathol.3800593 (2006).

113   Bortner, D. M. & Rosenberg, M. P. Induction of mammary gland hyperplasia and carcinomas in transgenic mice expressing human cyclin E. *Molecular and cellular biology* **17**, 453-459 (1997).

114   Hunt, K. K. & Keyomarsi, K. Cyclin E as a prognostic and predictive marker in breast cancer. *Seminars in cancer biology* **15**, 319-326, doi:10.1016/j.semcancer.2005.04.007 (2005).

115   Donnellan, R. & Chetty, R. Cyclin E in human cancers. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **13**, 773-780 (1999).

116   Sala, A. *et al.* Activation of human B-MYB by cyclins. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 532-536 (1997).

117   Zhao, J., Dynlacht, B., Imai, T., Hori, T. & Harlow, E. Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. *Genes & development* **12**, 456-461 (1998).

118   Ross, J. S. *et al.* The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy. *The oncologist* **8**, 307-325 (2003).

119   Serrano-Olvera, A., Duenas-Gonzalez, A., Gallardo-Rincon, D., Candelaria, M. & De la Garza-Salazar, J. Prognostic, predictive and therapeutic implications of HER2 in invasive epithelial ovarian cancer. *Cancer treatment reviews* **32**, 180-190, doi:10.1016/j.ctrv.2006.01.001 (2006).

120   Livasy, C. A., Reading, F. C., Moore, D. T., Boggess, J. F. & Lininger, R. A. EGFR expression and HER2/neu overexpression/amplification in endometrial carcinosarcoma. *Gynecologic oncology* **100**, 101-106, doi:10.1016/j.ygyno.2005.07.124 (2006).

121   Yano, T. *et al.* Comparison of HER2 gene amplification assessed by fluorescence in situ hybridization and HER2 protein expression assessed by immunohistochemistry in gastric cancer. *Oncology reports* **15**, 65-71 (2006).

122   Mimura, K. *et al.* Frequencies of HER-2/neu expression and gene amplification in patients with oesophageal squamous cell carcinoma. *British journal of cancer* **92**, 1253-1260, doi:10.1038/sj.bjc.6602499 (2005).

123   Cox, G. *et al.* Herceptest: HER2 expression and gene amplification in non-small cell lung cancer. *International journal of cancer. Journal international du cancer* **92**, 480-483 (2001).

124   Mano, M. S., Rosa, D. D., De Azambuja, E., Ismael, G. F. & Durbecq, V. The 17q12-q21 amplicon: Her2 and topoisomerase-IIalpha and their importance to the biology of solid tumours. *Cancer treatment reviews* **33**, 64-77, doi:10.1016/j.ctrv.2006.10.001 (2007).

125   Nakamura, H., Kawasaki, N., Taguchi, M. & Kabasawa, K. Association of HER-2 overexpression with prognosis in nonsmall cell lung carcinoma: a metaanalysis. *Cancer* **103**, 1865-1873, doi:10.1002/cncr.20957 (2005).

126   Tanner, M. *et al.* Amplification of HER-2 in gastric carcinoma: association with Topoisomerase IIalpha gene amplification, intestinal type, poor prognosis and sensitivity to trastuzumab. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **16**, 273-278, doi:10.1093/annonc/mdi064 (2005).

127   Morrison, C. *et al.* HER-2 is an independent prognostic factor in endometrial cancer: association with outcome in a large cohort of surgically staged patients. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**, 2376-2385, doi:10.1200/JCO.2005.03.4827 (2006).

128   Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-182 (1987).

129   Olayioye, M. A. Update on HER-2 as a target for cancer therapy - Intracellular signaling pathways of ErbB2/HER-2 and family members. *Breast Cancer Research* **3**, 385-389, doi:Doi 10.1186/Bcr327 (2001).

130   Ursini-Siegel, J., Schade, B., Cardiff, R. D. & Muller, W. J. Insights from transgenic mouse models of ERBB2-induced breast cancer. *Nature reviews. Cancer* **7**, 389-397, doi:10.1038/nrc2127 (2007).

131   Lee, A. *et al.* Expression of c-erbB2, cyclin D1 and estrogen receptor and their clinical implications in the invasive ductal carcinoma of the breast. *Japanese journal of clinical oncology* **37**, 708-714, doi:10.1093/jjco/hym082 (2007).

132   Timms, J. F., White, S. L., O'Hare, M. J. & Waterfield, M. D. Effects of ErbB-2 overexpression on mitogenic signalling and cell cycle progression in human breast luminal epithelial cells. *Oncogene* **21**, 6573-6586, doi:10.1038/sj.onc.1205847 (2002).

133   Yang, C. *et al.* The role of the cyclin D1-dependent kinases in ErbB2-mediated breast cancer. *The American journal of pathology* **164**, 1031-1038, doi:10.1016/S0002-9440(10)63190-2 (2004).

134   Pittman, A. M. *et al.* The CDH1-160C>A polymorphism is a risk factor for colorectal cancer. *International journal of cancer. Journal international du cancer* **125**, 1622-1625, doi:10.1002/ijc.24542 (2009).

135   Zhan, Z. *et al.* CDH1 gene polymorphisms, plasma CDH1 levels and risk of gastric cancer in a Chinese population. *Molecular biology reports* **39**, 8107-8113, doi:10.1007/s11033-012-1658-0 (2012).

136   Berx, G., Becker, K. F., Hofler, H. & van Roy, F. Mutations of the human E-cadherin (CDH1) gene. *Human mutation* **12**, 226-237, doi:10.1002/(SICI)1098-1004(1998)12:4<226::AID-HUMU2>3.0.CO;2-D (1998).

137   Benusiglio, P. R. *et al.* CDH1 germline mutations and the hereditary diffuse gastric and lobular breast cancer syndrome: a multicentre study. *Journal of medical genetics*, doi:10.1136/jmedgenet-2012-101472 (2013).

138   Corso, G. *et al.* Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**, 868-875, doi:10.1200/JCO.2012.44.4612 (2013).

139   Ghaffari, S. R., Dastan, J., Rafati, M. & Sabokbar, T. Novel human pathological mutations. Gene symbol: CDH1. Disease: gastric cancer. *Human genetics* **125**, 337 (2009).

140   Li, M. & Zhang, P. The function of APC/CCdh1 in cell cycle and beyond. *Cell division* **4**, 2, doi:10.1186/1747-1028-4-2 (2009).

141   Liu, K., Lin, F. T., Ruppert, J. M. & Lin, W. C. Regulation of E2F1 by BRCT domain-containing protein TopBP1. *Molecular and cellular biology* **23**, 3287-3304 (2003).

142   Karppinen, S. M. *et al.* Identification of a common polymorphism in the TopBP1 gene associated with hereditary susceptibility to breast and ovarian cancer. *European journal of cancer* **42**, 2647-2652, doi:10.1016/j.ejca.2006.05.030 (2006).

143   Jeon, Y. *et al.* TopBP1 deficiency causes an early embryonic lethality and induces cellular senescence in primary cells. *The Journal of biological chemistry* **286**, 5414-5422, doi:10.1074/jbc.M110.189704 (2011).

144   Going, J. J. *et al.* Aberrant expression of TopBP1 in breast cancer. *Histopathology* **50**, 418-424, doi:10.1111/j.1365-2559.2007.02622.x (2007).

145   Jeon, Y. *et al.* Human TopBP1 participates in cyclin E/CDK2 activation and preinitiation complex assembly during G1/S transition. *The Journal of biological chemistry* **282**, 14882-14890, doi:10.1074/jbc.M609116200 (2007).

146   Rubin, L. L. & de Sauvage, F. J. Targeting the Hedgehog pathway in cancer. *Nature reviews. Drug discovery* **5**, 1026-1033, doi:10.1038/nrd2086 (2006).

147   Roy, S. & Ingham, P. W. Hedgehogs tryst with the cell cycle. *Journal of cell science* **115**, 4393-4397, doi:10.1242/jcs (2002).

148   Lam, C. W. *et al.* A frequent activated smoothened mutation in sporadic basal cell carcinomas. *Oncogene* **18**, 833-836, doi:10.1038/sj.onc.1202360 (1999).

149   Yauch, R. L. *et al.* Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326**, 572-574, doi:10.1126/science.1179386 (2009).

150   Wang, X. D. *et al.* Mutations in the hedgehog pathway genes SMO and PTCH1 in human gastric tumors. *PloS one* **8**, e54415, doi:10.1371/journal.pone.0054415 (2013).

151   Xie, J. *et al.* Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90-92, doi:10.1038/34201 (1998).

152   Pais, A. *et al.* Identification of various MLL gene aberrations that lead to MLL gene mutation in patients with acute lymphoblastic leukemia (ALL) and infants with acute leukemia. *Leukemia research* **29**, 517-526, doi:10.1016/j.leukres.2004.11.016 (2005).

153   Agilent. *Agilent SureSelect Human All Exon 50Mb Kit Data Sheet*, <http://www.chem.agilent.com/library/datasheets/Public/5990-6319en_lo.pdf> (2013).