Target deconvolution using *in vitro* evolution in haploid human cells

Juan Carlos Jado[1,2,*], Michelle Dow[2,3,4,*], Krypton Carolino[5], Gregory J. Fonseca[6], Trey Ideker[2,7,#], Hannah Carter[2,7,#], Elizabeth A. Winzeler[1,2,#]

Affiliations:

1Division of Host-Microbe Systems & Therapeutics, Department of Pediatrics, University of California, San Diego, Gilman Dr., La Jolla, CA 92093, USA
2Department of Medicine, Division of Medical Genetics, University of California San Diego, La Jolla, CA 92093, USA
3Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093, USA
4Health Science, Department of Biomedical Informatics, School of Medicine, University of California San Diego, La Jolla, CA 92093, USA
5Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA
6 Department of Medicine, Meakins-Christie Laboratories, McGill University Health Centre, 1001 Decaire Blvd, Montreal, QC H4A 3J1, McGill University, Montreal, Quebec, Canada
7Moores Cancer Center, University of California San Diego, La Jolla, CA 92093, USA

*These authors contributed equally

#Correspondence to Elizabeth Winzeler (ewinzeler@ucsd.edu) and Hannah Carter (hcarter@ucsd.edu)

Short title: *In vitro* evolution in human cells

ABBREVIATIONS

AF, allele frequency; CNV, Copy Number Variation; NGS, Next Generation Sequencing, WES, Whole Exome Sequencing; WGS, Whole Genome Sequencing; CML, Chronic Myelogenous Leukemia; IVIEWGA, *In Vitro* Evolution and Whole Genome Analysis; SNV, Single Nucleotide Variant; CNV, Copy Number Variation; TF, transcription factor; DOX, Doxorubicin; GEM, Gemcitabine; ETP, Etoposide; PTX, Paclitaxel; TPT, Topotecan; AML, Acute Myeloid Leukemia; TKIs, Tyrosine Kinase Inhibitors; MDR, Multi-Drug Resistance; gDNA, genomic DNA.

ABSTRACT

Robust methods are needed for determining the mechanism of action and mechanism of resistance for small molecules that are active against human cells. In microbes, in vitro evolution and whole genome analysis (IVIEWGA) is the preferred method for target deconvolution and for revealing resistance mechanisms. To investigate the method's applicability to human cells we evolved resistance to five different anticancer drugs (doxorubicin, gemcitabine, etoposide, topotecan, and paclitaxel) using a near-haploid cell line (HAP-1). Resultant clones (N=28) were compared to their isogenic parents via whole genome and whole exome sequencing (WES). High frequency alleles predicted to change protein sequence, or alleles which appeared in the same gene for multiple independent selections with the same compound were identified in only 21 genes: The set included clinically-relevant resistance genes or known drug targets (*TOP1, TOP2A*, *DCK*, *WDR33, SLCO3A1)*, as well as new genes (*SLC13A4*). In addition, some lines carried structural variants that encompassed additional known resistance genes (*ABCB1, WWOX and RRM1)*. Gene expression knockdown and knockout experiments (via shRNA and CRISPR-*Cas*9 respectively) of 10 validation targets showed a high degree of specificity and accuracy in our calls and demonstrates that the same drug resistance mechanisms found in diverse clinical samples can be evolved, identified and studied in an isogenic background. Our data show that in vitro evolution and whole genome analysis is a promising method for target identification as well as for identifying resistance mechanisms.

INTRODUCTION

In human cells, methods for discovering genes that play a role in drug resistance or which encode drug targets, especially for poorly characterized compounds, such as natural products, are limited. Genome-wide CRISPR-Cas9 knockdown experiments[1-3] in the presence of drug are useful but cannot readily reveal critical gain-of-function, single nucleotide alleles, such imatinib-resistance conferring mutations in BCR-Abl. Discovering common alleles in whole genome sequences of tumors from cohorts of patients that have relapsed after drug treatment requires very large datasets and is complicated by patient heterogeneity. The method cannot be used on experimental therapies.

Work in other organisms has shown that *in vitro* evolution and whole genome analysis (IVIEWGA) is a powerful method to discover both a comprehensive set of drug resistance alleles, as well as the targets of compounds with unknown mechanisms of action[4,5]. In this method, clonal or near clonal organisms are isolated and then clones are subjected to increasing levels of a drug that inhibits growth. After selection, the organism is cloned again. The genomes of resistant clones are then compared to the sensitive parent clone using next generation sequencing (NGS) methods. In organisms such as *Saccharomyces cerevisiae*[6], *Plasmodium falciparum*[4,5], Mycobacteria[7], Trypanosomes[6], and *Acinetobacter baumannii*[8] this method has been used to comprehensively discover resistance conferring variants. Surprisingly, the data shows that typically only a small number of *de novo* variants are detected after evolution. If multiple selections are performed on independent clones, the same resistance gene will appear repeatedly, although often appearing with different alleles, providing a high level of statistical confidence that the allele has not arisen by chance.

Many of the organisms on which IVIEWGA has been used with success have both haploid and diploid phases of their lifecycle, which means that selections can be performed in a haploid stage. Selecting for resistant clones in a haploid organism greatly simplifies analysis as a driver resistance allele will approach 100% frequency. In addition, for loss of function alleles, only one mutation is needed (versus both copies). Although metazoans are all diploid, haploid human cells lines are nevertheless available: HAP-1, is a human chronic mylogenous leukemia (CML)-derived cell line that is completely haploid except for a 30 megabase fragment of chromosome 15 [9]. HAP-1 has been used for genetic studies because mutated phenotypes are immediately exposed[10].

Using five different anticancer drugs (Doxorubicin, Gemcitabine, Etoposide, Topotecan, and Paclitaxel) as examples, we show that *in vitro* evolution in HAP-1 cells can be used to identify the molecular basis of drug resistance in human-derived cells. Through our unbiased analysis of evolved clones, we detect a limited number of genes that acquire SNVs or CNVs after prolonged, sublethal exposure to our selected xenobiotics. We further demonstrate the power of the approach by using shRNAs and CRISPR-Cas9 to downregulate or reintroduce selected alleles and demonstrate that this confers resistance or sensitivity to the drug which elicited the evolved genomic change. Our work has implications for clinical intervention strategies to prevent the emergence of drug resistance and tumor recurrence through gene mutations acquired through DNA damage from chemotherapeutics or natural variants which exist and persist from the heterogenous tumor cell environment.

RESULTS

Selection of compounds for resistance studies

To identify xenobiotics with the best efficacy against HAP-1 cells we first measured ATP levels (CellTiterGlo) treating HAP-1 cells with serial dilutions of 16 different drug for 48 hours. Five drugs showed EC_{50} values between 5 to 340 nM (Fig. 1A-B, Table S1). These included doxorubicin (DOX, $EC_{50} = 46.05 \pm 4.6$ nM), also known as adriamycin, an anthracycline antibiotic that works by inhibiting topoisomerase II alpha (*TOP2A*)[11,12]; gemcitabine (GEM, $EC_{50} = 8.7 \pm 0.7$ nM), a synthetic pyrimidine nucleoside prodrug that is used against a variety of hematopoietic malignancies[13-15]; etoposide (ETP, $EC_{50} = 338.6 \pm 39.7$ nM), a semisynthetic derivative of podophyllotoxin that also targets *TOP2A* and prevents re-ligation of the double-stranded DNA[16]; paclitaxel (PTX, $EC_{50} = 17.5 \pm 4.0$ nM) also known as taxol, an effective anticancer agent that targets tubulin, perturbing the cytoskeleton and causing M phase cell-cycle arrest[17], and topotecan (TPT, $EC_{50} = 5.6 \pm 0.1$ nM), a semisynthetic water-soluble derivative of camptothecin that inhibits topoisomerase I (*TOP1*)[18]. Our HAP-1 EC_{50} values were similar to those previously reported for other CML cell lines (www.cancerrxgene.org) with the exception of etoposide, which appeared more effective in HAP-1 cells ($EC_{50} = 338.6 \pm 39.7$ nM) relative to other CML cell lines (> 1 μM in BV-173, KU812, EM-2, MEG-01, JURL-MK1, KCL-22, RPMI-8866, LAMA-84, K-562).

Evolution of resistance is readily achieved for all compounds.

Our next objective was to create drug resistant lines. Although we have had difficulty creating resistant lines for some drugs in some species ("irresistibles"[19]), there is precedent for resistance to the drugs included here[20-22]. To reduce the possibility of selecting mutations that were present in a nonhomogenous population of HAP-1 cells and to facilitate later genomic analysis, we first cloned the cells. This was accomplished by diluting cells to an average density of ~0.5 cells per well in a poly-L-lysine treated 96-well plate (Fig. 1A) and then picking clones from wells that contained single colonies. Selections were initiated with different parent clones for the different drug-specific replicates (Fig.1C, Fig. S1).

To create drug resistant clones, cells were grown in tissue culture dishes (reaching 60-80% semi-confluence) in the presence of sublethal concentrations of each drug using one of two different methods. Most cell lines (DOX, GEM, TPT and PTX resistant clones) were subjected to a lethal concentration (~3-5X EC_{50} value), killing more than 95% of the cells. Then, treatment was removed until cells reached semi-confluence again (doubling every 22 hours[23]) whereupon drug at ~ the EC_{95} value was reapplied. Alternatively, for ETP-resistant clones a step-wise selection method was used whereby cells were repeatedly exposed to a concentration that killed around 50% of the cell population. Drug concentration was increased by 5-10% every 5 days while keeping the growth rate at 50% of untreated culture. Because mutations will arise randomly during long term cell culture, we attempted at least three independent selections for each drug, in each case starting with an identical parental isolate (Fig. 1C). In a few cases, independent selections could not be achieved and dependent clones with a shared lineage (DOX-R4a and DOX-R4b; PTX-R2a and PTX-R2b; TPT-R4a, TPT-R4b and TPT-R4c) were collected. Resistance emerged after several months depending on the drug and the method used (7-30 weeks approximately, 49-210 generations) (Fig. S1).

Once resistance was observed in the batch culture, we isolated clones from batch drug-selected cultures and the drug sensitivity of the clone was measured and compared to the isogenic parent clones (Fig. 1D). We observed an EC_{50} fold shift between 3.3 and 411.7 (Table S2) in paired comparisons. To demonstrate that the drug resistance phenotype was stable, drug pressure was removed for 8 weeks (approximately 56 generations) and clones were retested for sensitivity. We observed no changes in the EC_{50} values, indicating that resistance phenotypes were not due to transient adaptation.

Identification of putative resistance variants using next-generation sequencing

We next performed whole genome and exome paired-end read sequencing on the 35 cell lines (both drug-resistant clones and their matched drug-sensitive parent clones). Our IVIEWGA studies in *Plasmodium*[5], have shown that stable drug resistance is typically conferred by SNVs in coding regions and thus exome sequencing was an efficient mechanism to find causal variants. However, gene amplifications, which contribute to 1/3 of drug resistance events in *Plasmodium*[5], are more accurately detected with WGS because exact chromosomal recombination sites, which may fall in intergenic regions, can be reconstructed from WGS data. Because of falling costs over the course of the project, more samples (N=21) were ultimately whole genome sequenced than whole exome sequenced (N=14).

Sequencing quality was high for all samples: alignment showed that, on average, 99.9% of 700 million WGS (40 million WES) reads mapped to the hg19 reference genome with 86% of the bases covered by 20 or more reads (Table S3). By comparing sequences of evolved clones to their respective parental clones, we discovered a total of 26,625 unique SNVs (Table 1, Table S5, Methods). The majority of variants in all cell lines were non-coding (Table 1, Table S4, S5) and were evenly distributed with respect to chromosome length (Fig. S2). Of the 26,625 mutations most (26,468) were present at allele frequencies (AF) of less than 0.85 relative to their parent clone and would thus not be expected to be driver mutations, given that the parents were cloned (to the best of our ability) before selections were initiated.

We next developed a pipeline (Fig. S3A, Methods) to filter the 26,625 called mutations (Table S6) to a final list of potential variants conferring drug resistance (Fig. S2A, Table 1). Previous analyses in other species suggested that variants presented in coding regions are more likely to contribute to drug resistance even though this could exclude the variants associated with certain transcription factor (TF) binding sites. Therefore, our strategy focused on mutations that were in exonic regions, and were drug-specific (Fig. S3A,). We further considered only mutations likely to have a functional impact at the protein level (missense, nonsense, frameshift, start or stop gain or loss) which further reduced the number to an average of 35 and 23 nonsynonymous mutations for WGS and WES, respectively (Fig. S3A). Reasoning that resistance driver mutations (e.g. those actually causing resistance) would be present in 100% of the haploid cells in the sequenced culture, we selected only the variants with high allele frequency (AF > 0.85). Although

some with AF <0.85 could confer a beneficial advantage to the cell, most are likely to be random mutations that arose during long term culture. Finally, based on our experience with microbes whereupon genes with multiple, predicted independent amino acid changes (not expected by chance) are often found for the same drug, we added these genes to our final list of candidates (Table S8). The five drugs varied in the number of mutations with TPT having the highest overall numbers (Table 1).

Somatic Copy number variations (CNVs)

We next searched for CNVs (both amplifications and deletions) in our WGS and WES data using Control-FreeC[24]. Overall patterns of broad and focal alterations across the drugs and conditions varied, with copy number gains observed more often than losses (Fig. S3B, Table S7). Using a corrected p-value of less than 0.05, we identified 93 total amplification and 108 deletion events, with most appearing in the TPT-resistant samples (123) (Table S7). The CNVs had an average size of 8.5 Mbp (stdev 19 Mbp), ranged from 15,000 bp to 152 Mbp (Fig. S2B) and covered ~3% of the genome, on average. More CNVs were called in WES samples because of sequencing gaps—even for WGS samples, some CNVs were separated by short distances and were nearly contiguous (Fig. S3B). It is likely that some CNVs were also missed in WES data. The number of events was proportional to chromosome size with the exception of the Y chromosome, for which there were ~4X more events (47) per unit length.

Doxorubicin resistance is associated with mutations in *TOP2A* and a solute carrier transporter

To evaluate the approach, we next considered the set of SNVs and CNVs for each drug. For DOX, we had six available selections from two different starting clones (WT-1 (WGS) and WT-5 (WES)) that were analyzed by WGS (DOX-R1, DOX-R2, DOX-R3) and by WES (DOX-R4a, DOX-R4b and DOX-R5) (Fig. 2A). High allele frequency missense mutations were found in only 11 genes (Table S8). Of note, DOX-R2 and DOX-R3 carried mutations in *TOP2A* at frequencies of 0.89 and 0.87, respectively. This is the known target of DOX[16,25] and is known to play a role in drug resistance[25-27]. The amino acid mutation, Pro803Thr (Fig. 2B), sits within the principal DNA-binding locus, the DNA-gate, a region conserved in type II topoisomerases (*TOP2A* and *TOP2B*). It is also adjacent to the catalytic tyrosine (Tyr805), responsible for

nucleophilic attack on DNA[28]. While one explanation is that Pro803Thr creates steric hindrance and blocks DOX access to the site, a more likely explanation is that the mutation is a loss-of-function mutation, especially as knockdown of TOP2A activity has previously been shown to confer DOX resistance in a Eμ-Myc mouse lymphoma system[29]. To reproduce these results in our HAP-1 human cells, *TOP2A* was downregulated using a shRNA pool containing constructs encoding target-specific shRNA hairpins for *TOP2A*. Western blots further showed the expected down regulation of protein levels (Fig. 2C) and an EC_{50} analysis of the wildtype and the knockdown line revealed a 4.25-fold increase in DOX resistance compared to the isogenic parent (Fig. 2D, E).

We also found missense mutations present in 100% of the reads for several other attractive but less well characterized genes; *SLC13A4* (Gln165His, DOX-R4b), and *SPG7* (Lys593Asn, DOX-R5), as well as one uncharacterized gene (AC091801.1, His13Asn, DOX-R4a) in the three different clones that were subjected to WES and were derived from WT-5. *SLC13A4* is a solute carrier transport family member and members of this general solute carrier family have appeared in selections conducted in microbes (e.g. the UDP-galactose transporter and the AcetylCoA transporter[30]) and are also associated with cancer drug resistance[31]. The Gln165His mutation is located in the disordered region of the protein. To validate *SLC13A4* we performed a gene knockdown using a shRNA pool that targeted *SLC13A4*, containing three expression constructs each encoding target-specific 19-25 nucleotide shRNA hairpins. Protein expression levels of the knockdown line were measured by western blot followed by a dose-response assay to compare its EC_{50} value with the wildtype line (Fig. 2F). The 4X increase in resistance demonstrates that *SLC13A4* contributes to resistance, although it may not account completely for the level of resistance of the sequenced clones, which ranged from 4X to 11X (Fig. 2G, H).

Gemcitabine resistance is conferred by changes in DCK and RRM1 activity

Six selections were performed with GEM (starting from two different isogenic parents; WT-2 (WGS) and WT-3 (WES)). Among those, three GEM-resistant clones subjected to WGS (GEM-R1, GEM-R2 and GEM-R3) showed an average EC_{50} shift of 300 to 400-fold (Fig. 3A, Table S2), and the clones showed no change in HAP-1 sensitivity to other drugs (Fig. 3B). As there were no candidate alleles with AF > 0.85, we looked for genes that acquired mutations in multiple selections, identifying deoxycytidine kinase (*DCK*) as likely important for resistance.

Interestingly, across cell lines several distinct mutations were found in *DCK*, with varying effects (missense and frameshift) across several different positions (Table 2). In particular, the missense substitution Ser129Tyr, present in GEM-R1 and GEM-R3, not only alters the amino-acid size and charge also falls at the end of exon 3, within the ATP-binding pocket of a phosphorylation site, making it a strong causal candidate for GEM drug resistance (Fig. 3C). GEM only becomes pharmacologically active if it is phosphorylated and the first phosphorylation is catalyzed by DCK. shRNA knockdown of DCK was confirmed by western blot analysis (Fig. 3D). Downregulation of the gene showed a 36.5-fold increase in the EC_{50} value compared to both the isogenic parent line and the shRNA negative control (Fig. 3E, F; Table 2).

The three WT-3 derived GEM-resistant clones (GEM-R4, GEM-R5 and GEM-R6) subjected to WES were not as resistant as those used in WGS (~6X versus ~400X, Table S2). No high AF SNVs were evident in these lines and *DCK* exons were not mutated. On the other hand, the three WES clones contained 20 CNVs that could play a role in drug resistance. Most CNVs were not shared between lines but GEM-R4, GEM-R5 and GEM-R6 all bore overlapping CNVs of varying sizes on chromosome 11, with all three lines bearing 3-4 copies (p value = 1.38e-37 to 2.05e-142) (Fig. 2). The chromosome 11 CNV was only found in GEM resistant lines and not in any of the other evolved lines (in contrast to CNVs on chromosome 1 or 16, for example). While it is difficult to determine which of the 140 genes in the smallest interval contribute to resistance, a known resistance mediator or target of GEM, ribonucleotide reductase (*RRM1*), is found within the amplified region. RRM1 is the largest catalytic subunit of ribonucleotide reductase, a key enzyme catalyzing the transformation of ribonucleotide diphosphates to deoxyribonucleoside diphosphates that are required for DNA synthesis and repair, and GEM is known to inhibit DNA polymerase by inhibiting *RRM1*[32]. Furthermore, overexpression of RRM1 is associated with poorer prognosis after gemcitabine treatment in non-small cell lung cancer[33] and in bladder cancer[34].

Western blot analysis of the evolved lines showed that the amplification was indeed associated with increased protein levels (Fig. 3G). As an additional validation, we performed a single shRNA knockdown of *RRM1* to reduce protein expression (Fig. 3H), followed by a dose-response assay comparing EC_{50} values of both wildtype HAP-1 and *RRM1* knockdown lines, which showed that downregulation of *RRM1* made HAP-1 cells 31-fold more sensitive to GEM

than their isogenic parent (Fig. 3I, J). As expected *RRM1* downregulation had no effect on HAP-1 sensitivity to other drugs (Fig. S4).

Etoposide resistance is modulated by levels of *WDR33*

We created three independent ETP resistant clones, all of which were subjected to WES, and compared them to one isogenic parent clone (WT-3) (23, 13 and 9-fold increased resistance respectively (Fig. S5A, Table S2). A single gene, *WDR33* (ETP-R3), carried a SNV (Pro622Thr) with a 100% allele frequency. This gene encodes for a member of the *WD* repeat protein family and is one of the six subunits of a multiprotein complex called *CPSF* (cleavage and polyadenylation specific factor)[35] involved in cell cycle progression, signal transduction, apoptosis and gene regulation. Disruption of *WDR33* can lead to slowed DNA replication forks[36], which could potentially explain why its disruption protects against topoisomerase inhibitors that block DNA unwinding. Lines in which *WDR33* was knocked down via shRNA acquired an EC_{50} value 3.4 times greater than its parental line or the scrambled control (Fig. S5B-D; Table 2).

No clear candidate SNVs were evident for ETP-R1 and ETP-R2, which did not carry the *WDR33* mutation (Table S6, Table S8). All ETP lines carried multiple CNVs, however, including a large shared amplification on chromosome 15 (ETP-R1 and R3). Approximately 120 protein coding genes are found in this region, including *BUB1B*, the BUB1 mitotic checkpoint serine/threonine kinase B, *BMF*, a BCL-modifying factor, as well as the *RAD51* recombinase, whose overexpression has been previously shown to confer ETP resistance[37]. Overexpression of *RAD51* activity sensitizes cells to a variety of cancer drugs, including etoposide, doxorubicin and topotecan[38]. Notably, all ETP resistant lines were also cross-resistant to PTX, TPT and DOX, providing support for this general mechanism of resistance.

Paclitaxel resistance is mediated by transporters *SLCO3A1 and ABCB1*

Seven different paclitaxel lines were created with different resistance levels (PTX-R1, R2a, R2b and R3, ~10X to PTX-R4, R5, R6, 50X) (Fig. 4A). The first four were subjected to WGS and the latter three to WES. SNV analysis yielded no candidate genes (frameshift, indels, and missense mutations with an allele frequency >0.85). From genes with an allele frequency of less than 0.85, *SLCO3A1,* encoding another solute carrier transporter, was notable in that multiple missense alleles were identified (Ile587Asn, Ala263Thr). This class of transporter is known to play a role in the import of drugs as well as hormones such as prostaglandin[39]. Gene knockdown experiments

showed that clones with loss of *SLCO3A1* resulted in HAP-1 cells that were ~8 times more resistant than their isogenic parents to PTX (Fig. 4C-E).

Despite the lack of coding SNVs, PTX-R1, R2a, R2b and R3 had a combined number of 47 CNVs, while PTX-R4, R5 and R6 had 10 (the fact that more CNVs were found in WGS samples may reflect the ease with which CNVs are called with WGS versus WES data). Potentially significant genes with CNVs were *ABCB1* (MDR1) and *ABCB4* (*MDR3*) (Fig. 4B) on chromosome 7 (PTX-R2a, R2b). *ABCB1* amplifications are associated with clinical resistance to PTX[40]. PTX-R4 and R5 showed structural variants on chromosome 1, and PTX-R4 show an amplification event on chromosome 17 that encompassed a variety of ABC transporters (*ABCA5, 6, 8, 9, 10*). No compelling candidate genes were found in CNVs for PTX-R6. On the other hand, inspection with IGV showed that read coverage was poor and that CNVs might not have been detected with WES data.

To confirm the importance of transporters in PTX resistance, clones were treated with both PTX and verapamil, a calcium channel-blocker which can reverse ABC-transporter mediated resistance [41,42]. We observed a complete reversal of resistance in PTX lines (Fig. 4F). In contrast, we observed no reversal of resistance in GEM lines (Fig. 4G), suggesting the resistance role of ABC-transporters is PTX-specific.

Topotecan resistance is associated with complex alterations in *TOP1*, deletion of *WWOX* and SNVs in cytochrome p450s (*CYP1B1*).

The six TPT samples were derived from four independent selection events (TPT-R4a-c are clones from the same selection with levels of resistance ranging from 10-20X; Fig. 5A; Table S2) and all six clones were subjected to WGS together with their parent clones (WT-6 and WT-7). No clear coding SNVs with a high allele frequency were obvious in TPT-R1, R2 and R3 but we noted multiple SNVs (Asp217Glu from TPT-R4a,b,c and Val432Leu from TPT-R1) in *CYP1B1*, which encodes a cytochrome p450 isoform. Gene knockdown experiments showed a small (2X) increase in TPT resistance with shRNA knockdown (Fig. 5B-D). Overexpression of CYP1B1 has previously been associated with TPT resistance[41].

For TPT-R4a-c lines, 268 alleles were present with AF > 0.85, but of these, only six were coding mutations and the rest were intergenic. Three of the six coding mutations were frameshift

mutations (His81) with AF = 1 in *TOP1* (Fig. S6A), the known target of topotecan[18]. The His81 frameshift mutation, which introduces a premature stop codon, was confirmed by the absence of the full-length protein using N-terminal antibodies (Fig. 5E). Sequencing of the 5' cDNA through the His81 frameshift for all three lines and as well as the parent line also confirmed the two-base deletion in the mutant as well as homozygosity in TPT-R4a-c evolved lines (Fig. 5F). We did observe a decrease in mRNA expression with TPT-R4a-b showing a statistically significant decrease in TOP1 mRNA expression, relative to TPT-WT (Fig. S7). In fact, others have shown using a targeted RNAi approach, that suppression of Top1 produces resistance to camptothecin, a close analog of topotecan[29]. Interestingly, of the 22 TOP1 frameshift or nonsense mutations in the COSMIC tumor database, 6 were located within a 30 amino acid span (of 765 total) that includes His81 (exon 4), suggesting likely clinical relevance[43]. The probability of this distribution by chance is 9.65×10^{-5}).

TPT resistant lines (TPT-R1, R2 and R3) also showed large chromosomal abnormalities at *WWOX* (Fig. 5I) with a clear deletion of the *WWOX* gene region (chr16:78,569,166-78,792,736, exon7 and 8). The absence of WWOX protein was confirmed by western blot analysis (Fig. 5J). Knockdown of *WWOX* by shRNA resulted in marked resistance to TPT (Fig. 5K-M). WWOX bears a well-known fragile site (FRA16D) and encodes a putative oxidoreductase. WWOX acts as a tumor suppressor and plays a role in apoptosis. Its disruption may prevent TPT-induced apoptosis, promoting cell survival in the presence of TPT[44].

DISCUSSION and CONCLUSIONS

Here, we show for the first time that *in vitro* evolution and whole genome analysis (IVIEWGA) can readily lead to the identification of drug resistance mechanisms in human cells. Our results show *in vitro* resistance acquisition and provide a framework for the determination of chemotherapy resistance alleles that may arise in patients.

Using IVIEWGA to study drug resistance can be time-consuming but very rewarding if performed correctly and with exact attention to lineage. Unlike bacteria, where mobile genetic elements often confer resistance, eukaryotic pathogens often acquire drug resistance by point mutations and CNVs. *In vitro* evolution has been used in malaria parasites to either discover or rediscover most, if not all (to our knowledge), clinically relevant drug resistance genes including

the chloroquine resistance transporter[5], the artemisinin resistance gene, *Pfkelch13*[45], and well-known ABC transporters[5]. As with the HAP-1 cells used here, the process can be time-consuming with selection for *Pfkelch13* taking several years and requiring sequencing of intermediate clones to pinpoint genomic changes that that correlated with jumps in resistance levels. Despite the challenges of working with malaria parasites (which grow slowly in liquid culture within human red cells, seldom reaching parasitemias beyond 10%), hundreds of *in vitro* evolutions have been performed. Based on our results here, resistance emerges in HAP-1 cells readily but more work will need to be done to determine if this is because of the compounds that were used. The rate of at which resistance arises based on the number of population doublings and the number of starting cells is now used to prioritize drug candidates for development in malaria parasites.

Our work using IVIEWGA in pathogens (see [46] for a review) guided our pipeline development: We focused on protein coding alterations that arose in association with a single treatment condition. Overall, our results are similar to what we have observed in eukaryotic pathogens with a mix of CNVs and single nucleotide variants giving rise to resistance. In contrast to what we have observed in microbes, we find much more heterogeneity and a higher number of CNVs per cell line, potentially because of the much larger genome size. The number of observed CNVs may be an underestimation as well as some CNVs may not have achieved statistical significance, especially in our WES data. This could be addressed in the future by using different algorithms, as well sequencing to somewhat higher coverage.

Another difference is that in eukaryotic microbes three independent selection will often result in the repeated identification of the same gene, often with different alleles. Here, in our first set of GEM selections we created mutations in *DCK*, while in the second set of selections, we created amplifications of *RRM1*. This may because there are more routes to resistance in HAP-1 cells or with the drugs used. In addition, different protocols (single dose versus step-wise increase) that were used here may have led to different outcomes: Alternatively, it may be the result of using a different starting clone. More work and repetition will be needed to distinguish between these possibilities.

Our initial study presented here ignores noncoding genomic alterations. However, non-coding RNAs such as *EGRF-AS1* and activating cis elements such as enhancers have previously been implicated in evasion of drug response[47-49]. The intergenic mutations with high allele

frequency are present in our provided datasets and provide opportunity for reanalysis or for querying by those interested in a specific noncoding RNA or enhancer. It is feasible that even synonymous mutations could confer resistance if they altered the rate of protein folding.

As expected from our work in eukaryotic pathogens, variants found at high allele frequencies in resistant cell lines validated with a high rate of success, suggesting that high allele frequency events often contribute to resistance. Given that we frequently observed more than one attractive genetic change per cell line, the final phenotype of most drug resistant lines can be attributed to a complex combination of multiple resistance alleles. Without the painstaking task of introducing the exact selected allele into the genome via CRISPR-Cas9 it is difficult to determine the exact contribution of a given allele: shRNA knockdowns are often incomplete and thus it is difficult to reciprocate the effect of resistance alleles or compare the resulting EC_{50}s in a quantitative way. For clearer results, fewer generations of selection should be performed in the future. In addition, collecting intermediate clones throughout the selection process for sequencing might allow better detection of driver mutations.

Our repeated unbiased (at least in the case of SNVs) identification of genes that are known to confer resistance in both patient isolates as well as in diploid cancer cells lines suggests that our results will be clinically relevant and that the use of a haploid cell line has little bearing on the results. Nevertheless, there may be differences between haploid and diploid cells as well as differences in the set of transcribed gene in HAP-1 cells, relative to other cancer cell lines. While modeling with diploid cells may provide more clinically relevant information, the haploid cells allow us to understand how chemoresistance arises from each mutation without the effect being masked by the other alleles.

Although the mutations that were studied here only appeared with drug pressure, for some genes there could be natural variants within the human population that could affect the response to chemotherapy. Knowing the complete set of genes associated with resistance could allow better therapies to be designed that match a given patients genetic makeup. For example, we identified the cytochrome p540, CYP1B1, and there is substantial natural variation in this gene. The studies described here could allow more personalized therapies and better therapeutic responses.

Finally, it is important to keep in mind that the compounds examined here are not modern cancer therapies and while still used clinically, they were developed without consideration of

resistance rates. In microbes, it is commonplace to evaluate compounds which will be developed into anti-infective therapies for their potential to give resistance using IVIEWGA. We anticipate that this method will be useful for predicting resistance mechanisms for new drugs in clinical development, as well as for new drug combinations and may lead to better classes of drugs for chemotherapy.

MATERIALS AND METHODS

Compounds

All chemotherapeutic agents used in this study were obtained from Sigma-Aldrich, dissolved in DMSO at 10mM concentration and stored at -20°C.

Cell cultures

The human chronic myelogenous leukemia cell line, HAP-1, was purchased as authenticated at passage 7 from Horizon Discovery and cultured in tissue culture dishes (Genesee Scientific, Cat# 25-202) as a monolayer at 37°C in a humidified atmosphere with 5% CO_2 using Iscove's Modified Dulbecco's Medium (IMDM) (Life Technologies, CA) supplemented with 10% fetal bovine serum (FBS), 0.29mg/mL L-glutamine, 25mM HEPES, 100U/mL Penicillin and 100µg/mL Streptomycin (1% pencillin/streptomycin). Monoclonal and polyclonal stocks were made and stored in IMDM + 10% DMSO in liquid nitrogen.

In vitro evolution of resistant HAP-1 clones.

Prior to selection, an aliquot of the parental line was stocked as a reference for subsequent whole genome sequencing analysis. Three independent clones of HAP-1 cells were cultured in tissue culture dishes exposed to increasing sublethal concentrations of each chemotherapeutic agent at a starting concentration previously determined by the EC_{50} value for around 7-30 weeks depending on the drug, its speed of action and the method used as two methods were considered: high-pressure intermittent selection method and a step-wise selection method. For high pressure selection, cells were treated at a concentration 3-10 X EC_{50} value until more than 95% of the cells died. Then treatment was removed and cells were allowed to recover. After reaching around 60% semi-confluence, treatment was reinstalled and EC_{50} value monitored. For step-wise selection

method, drug concentration used was at the EC_{50} which implied reduced growth rate of approximately 50% and drug pressure was increased in intervals of around 5-10% keeping growth inhibition around 50%. Once the EC_{50} values of the resistant lines were at least 5 times greater than the one used as control, cells were again cloned by limiting dilution and further reconfirmed for drug resistance and subsequent DNA extraction for whole genome sequencing analysis.

Dose-response assay by EC_{50} determination and bioluminescence quantification

Drug sensitivity and cell viability were assessed by a bioluminescence measurement as follows: twenty-four hours prior to addition of the drugs, $2x10^4$ cells/well for every replicate were seeded in a 96-well plate. Once attached, media was removed and 10 different concentrations of drug were added in serial dilutions 1:3 with a starting concentration of 10µM or one of which the EC_{50} value of the replicates fell within an intermediate drug concentration. When drug-resistant lines were co-treated in combination with verapamil, a fixed concentration of verapamil (10µM) was added to every concentration of the drug. After a 48-hour incubation period at 37°C and 5% CO_2 with the drug, cells were treated with CellTiterGlo (Promega) reagent (diluted 1:2 with deionized water) for quantification of HAP-1 cell viability. Immediately after addition of the luminescence reagent, luminescence was measured using the Synergy HT Microplate Reader Siafrtd (BioTek). The data was normalized to 100% cell survival and 100% cell death and EC_{50} values were obtained using the average normalized luminescence intensity of 8 wells per concentration and a non-linear variable slope four-parameter regression curve fitting model in Prism 8 (GraphPad Software Inc.)

Isolation of total DNA from drug resistant lines

Genomic DNA (gDNA) was extracted from drug-specific resistant cell lines together with their isogenic parental lines using the DNeasy® Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. Samples were assessed for quantity with the Qubit™ dsDNA BR Assay Kit (Life Technologies, Carlsbad, CA, USA). All samples (>2.0µg, >50ng/µL, >20µL) were prepared for quality control by testing gDNA degradation or potential contamination using agarose gel electrophoresis (1% Agarose, TAE, ~100 Voltage). Then gDNA concentration was again measured using the Qubit® DNA Assay Kit with the Qubit® 2.0 Fluorometer (Life Technologies). Finally, fragment distribution of the gDNA library was measured using the DNA 1000 Assay Kit with the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). DNA

libraries were sequenced with 150 base pair (bp) paired single end reads on an Illumina HiSeq 4000 (PE150).

Genome Sequencing and Data Analysis

The quality of the raw FASTQ files was checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Whole genome sequencing (WGS) reads were mapped to GRCh37 (hg19) using BWA (v.0.7.17), specifically with the hs37d5 reference genome from 1000 Genomes project (Phase II). Whole exome sequencing (WES) samples were captured using Agilent SureSelect Human All Exon V6 (58 M), and the reads were also mapped to GRCh37 using BWA (v.0.7.17) with the same reference genome as WGS. Duplicate reads were removed using Picard (v.1.94); paired resistant and parent (WT) BAM files were used as input for The Genome Analysis Toolkit (GATK, v3.8-1). Local realignment and base quality recalibration were performed using default parameters. Somatic single nucleotide variants (SNVs) and small insertion and deletion (indels) were called using GATK MuTect2 following the state-of-the-art GATK Best Practices pipeline (https://ccbr.github.io/Pipeliner/Tools/MuTect2.html). In this project, the input to MuTect2 consisted of alignments for the parent and resistant clone in order. to call mutations with statistically significant differences in read support in the setting of resistance. Only the variants with PASS status, suggesting confident somatic mutations, were considered for further analysis. Variant allelic fraction was determined as the fraction of reads supporting the variant allele relative to total read depth at the variant position. Minimum callable depth was set to 10 and base quality score threshold was set to 18, following the default from MuTect2. All sequences have been deposited in SRA BioProject PRJNA603390.

Small-Variant Annotation for SNVs and Indels

Somatic variants were annotated using snpEff (v 4.3q)[50]. The annotation was performed with parameters including (1) canonical transcripts and (2) protein coding to enable identification of different functional classes of variant and their impact on protein coding genes (Table 1 showing finalized and consolidated annotations; Table S4 shows the raw annotation from snpEff and consolidated classification used in Table 1; Table S7 shows all the SNVs with their raw annotations). The snpEff sequence ontology designation was used in the filtering steps to classify variants generally as noncoding or coding (Table S4).

Identification of Drug Specific Genes

First, we excluded all variants in non-coding regions. Second, we excluded all non-functional variants, retaining only variants with a snpEff definition of HIGH or MODERATE impact (missense, stop lost, stop gain, and structural interaction variants). Finally, we selected only the variants with high allele frequency (AF > 0.85) and genes with multiple independent amino acid changes found in the same drug as the final list of candidates. The potential candidate variants were evaluated through Integrative Genomics Viewer (IGV)[51] to ensure coverage and allele fractions of the mutation positions. The top genes for each drug were included in Table 2 and Table S8.

Somatic Copy Number Variations (CNVs) Analysis

Copy number regions for WGS and WES were called by ControlFreeC[47] using the default settings for WGS and WES data. Background parental clone samples for each drug served as the control. Recurrent CNV regions were defined as regions observed in more than 1 sample, but not in all of clones from the tested drugs (as these are more likely to indicate potential sequencing artifacts).

Gene knockdowns

shRNAs targeting *TOP2A* (Cat# sc-36695-V), *DCK* (Cat# sc-60509-V), *SLCO3A1* (Cat# sc-62713-V), *SLC13A4* (Cat# sc-89760-V), *KLF-1* (Cat# sc-37831-V), *WWOX* (Cat# sc-44193-V), *WDR33* (Cat# sc-94735-V) and the non-coding control (Cat# sc-108080) were obtained in pLKO.1-Puro from Santa Cruz Biotechnology. *RRM1* (clone ID NM_001033.2-476s1c1) and *CYP1B1* (clone ID NM_000104.2-1176s1c1) were obtained in pLKO.1-Puro-CMV-tGFP from Sigma Aldrich.

Gene expression was knocked down using either a shRNA pool (Santa Cruz Biotechnology) containing between three and five expression constructs each encoding target-specific 19-25 shRNAs or a single shRNA (Sigma Aldrich). HAP-1 cells were plated at 120,000 cells per well (~40% confluency) in a 24-well plate 24 hours prior to viral transduction. On the day of transduction, complete media was replaced with serum-free media and 7µg/mL Polybrene® (Cat# sc-134220) and virus was added to the cells at a multiplicity of infection of 0.5 and cells were incubated overnight at 37°C. The following day, media was replaced with complete media

without Polybrene and cells were incubated at 37°C overnight. Cells were then split 1:3 and incubated for 24 hours more and finally stable clones expressing the shRNA were selected using complete media with 2µg/mL puromycin. After 7 days of selection with puromycin, knockdown efficiency was confirmed by western blot. Cells transduced with shRNAs containing fluorescent tags, were trypsinized (TrypLE™ Express; Cat# 12605-010, Gibco) after puromycin selection, washed twice with DPBS (1X) (Gibco) and sorted by flow cytometry.

Knockout of USP47

USP47 was knocked out (Cat# HSPD0000092816) using a single plasmid CRISPR-Cas9 system, using as lentivirus backbone the LV01 U6-gRNA:ef1a-puro-2A-Cas9-2A-tGFP targeting *USP47* (Sigma Aldrich). The target sequence (5'-3') was CAATGGGGCTTCTACTAGG. Transduction was as described above. HAP-1 cells were plated at 120,000 cells per well (~40% confluency) in a 24-well plate 24 hours prior to viral transduction. On the day of transduction, complete media was replaced with serum-free media and 7µg/mL Polybrene® (Cat# sc-134220), virus was added to the cells at a multiplicity of infection of 0.5 and cells were incubated overnight at 37°C. The following day, media was replaced with complete media without Polybrene and cells were incubated at 37°C overnight. Cells were then split 1:3 for 24 hours more and stable clones expressing the CRISPR/Cas9 sequence were selected using complete media with 2µg/mL puromycin. After 14 days of selection with puromycin and propagation as required, cells were trypsinized (TrypLE™ Express; Cat# 12605-010, Gibco), washed twice with DPBS (1X) (Gibco) and sorted by flow cytometry using the GFP fluorochrome which is expressed with Cas9. GFP positive cells were plated at an average density of 0.5 cells per well in a 96-well plate (previously treated with poly-L-Lysine (Sigma #P4707-50ml) to improve cell adhesion) in presence of 2µg/mL puromycin (limiting dilution cloning). Cell growth was monitored via microscopy during 25 days to select those wells which were observed to contain single colonies and *USP47* knockout was confirmed in those monoclonal HAP-1 cell lines first via PCR and then reconfirmed by western blot using the *USP47* rabbit polyclonal antibody (Abcam, Cat# ab97835).

Immunoblotting

HAP-1 cells (at least $5x10^6$) were trypsinized, washed twice with cold 1X DPBS and then lysed in 500µL Pierce™ RIPA Buffer (Thermo Scientific) containing 1:100 protease inhibitor (Halt™ Protease & Phosphatase Inhibitor Cocktail, Thermo Scientific) and 1:100 0.5M EDTA

Solution (Thermo Scientific). Total protein concentration was measured using the DC Protein Assay (Bio-Rad). Equal amounts of proteins were resolved by SDS-PAGE and transferred to nitrocellulose membranes (Bio-Rad #1704271), blocked in PBS with 5% (w/v) Blotting-Grade Blocker (Bio-Rad #170-6404) and 0.1% (v/v) Tween20 for 1h and probed. As secondary antibodies, HRP-linked anti-mouse or anti-rabbit (Cell Signaling Technology) were used and the HRP signal was visualized with SuperSignal®West Pico Chemiluminescent Substrate (Thermo Scientific #34080) using Syngene G-Box imager. Protein enrichment was calculated relative to *vinculin*, *γ-tubulin* or *β-actin*. Primary antibodies are listed below. Full size western blots are shown in Fig. S8.

Antibodies

TOP2A (Sigma #SAB4502997), *USP47* (Abcam #ab97835), *WDR33* (Abcam #ab72115), *DCK* (Abcam #ab151966), *β-actin* (Cell Signaling #3700S), *γ-tubulin* (Cell Signaling #4285S), *Vinculin* (Invitrogen #700062), *SLC13A4/SUT-1* (Abcam #ab236619), *WWOX* (Abcam #ab137726), *EKLF/KLF-1* (Abcam #175372), *SLCO3A1/OATP-A* (Santa Cruz #sc-365007), *TOP1* (Proteintech #20705-1-AP), CRISPR/Cas9 (Sigma #SAB4200701), *RRM1* (Abcam #ab133690), *CYP1B1* (Abcam #ab137562), *SPG7* (Sigma #SAB1406470 and Abcam #ab96213), goat anti-mouse (Invitrogen #G21040), goat anti-rabbit (Invitrogen #G21234).

RNA isolation, RT-PCR analysis and sequencing of TOP1 (*His81*)

TPT-resistant cells and TPT-WT ($1x10^6$ cells) were dissociated from plates by the addition of 2mL of TrypLE (Cat #12605-010, Gibco), washed and total RNA was isolated and purified using a Qiagen RNeasy® Mini Kit (Cat #74104, Qiagen) according to manufacturer's instructions. cDNA was synthesized from 1μg of total RNA using the Superscript™ First-Strand Synthesis System for RT-PCR Kit (Invitrogen #11904-018) and random hexamers. The primers used to amplify the region containing *His81* were FWD: GATCGAGAACACCGGCAC and REV: TCAGCATCATCCTCATCTCGAG. DNA from PCR product was extracted, using the QIAquick® Gel Extraction Kit (Qiagen #28706) following the manufacturer's instruction, measured using the Qubit® DNA Assay Kit with the Qubit® 2.0 Fluorometer (Life Technologies), and sequenced. The cDNA was sent to Eton Biosciences for Sanger sequencing. Quantification of TOP1 expression was performed using PerfeCTa® Sybr Green Fast Mix (Quanta #95072-250)

the following primers: FWD: CGAATCATGCCCGAGGATATAA; REV: CCAGGAAACCAGCCAAGTAA, following the manufacturer's instruction.

Acknowledgments: This work was supported by the National Institute of Health (NIH) to EAW, HC, and TI (GM085764), the San Diego Center for Systems Biology and UC San Diego Health Science fellowship to JCJ, and NIH National Library of Medicine Training Grant T15LM011271 to MD. EAW is also supported by grants from the Bill and Melinda Gates Foundation and the Medicines for Malaria Venture. The authors declare no conflicts of interest.

Author contributions: JCJ performed selections, validation experiments and wrote the manuscript. MD performed all computational analyses, and assembled figures and tables and wrote the manuscript. GF assisted with CRISPR-Cas9 experiments. KC performed RNA extraction and RT-qPCR experiments. HC provided advice and management and wrote the manuscript. EAW performed data analysis, provided advice, obtained funding and wrote the manuscript. TI obtained funding and provided advice.

Selection of drugs with activity against HAP-1

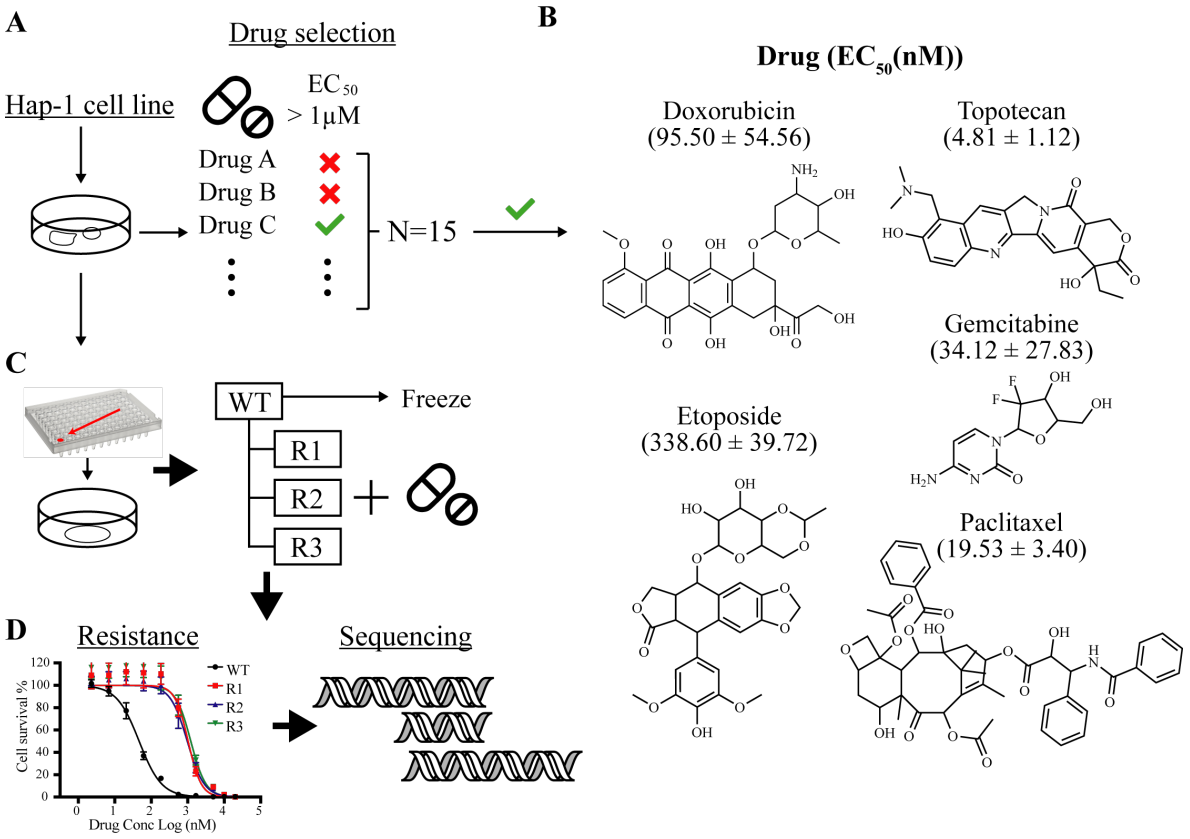

Fig. 1. Workflow for experimental design. A. Drug selection step for chemotherapy drugs. Dose-response assays by EC_{50} determination and bioluminescence quantification were first performed on 15 different chemotherapeutic agents (Table S1). Only drugs to which HAP-1 were sensitive (EC_{50} value below 1µM) were considered for IVIEWGA. **B.** Chemical structures of the sensitive chemotherapy agents used in the study. They all present diverse structures with different mechanisms of action. EC_{50} values are presented as mean ± s.e.m., n=3 with individual biological replicates overlaid and n=8 with individual technical replicates per concentration point. **C.** Clone selection. To ensure the background of the cell line, limiting dilution cloning was used to isolate individual cells and grow the colonies. For each drug selected, three independent selections and one parental line were tested in dose-response drug sensitivity assays to quantify level of drug resistance. **D.** Drug resistance development - cell resistance was achieved in 7 to 30 weeks approximately (49 and 210 generations). The parental cell line and the drug resistant lines were then sequenced.

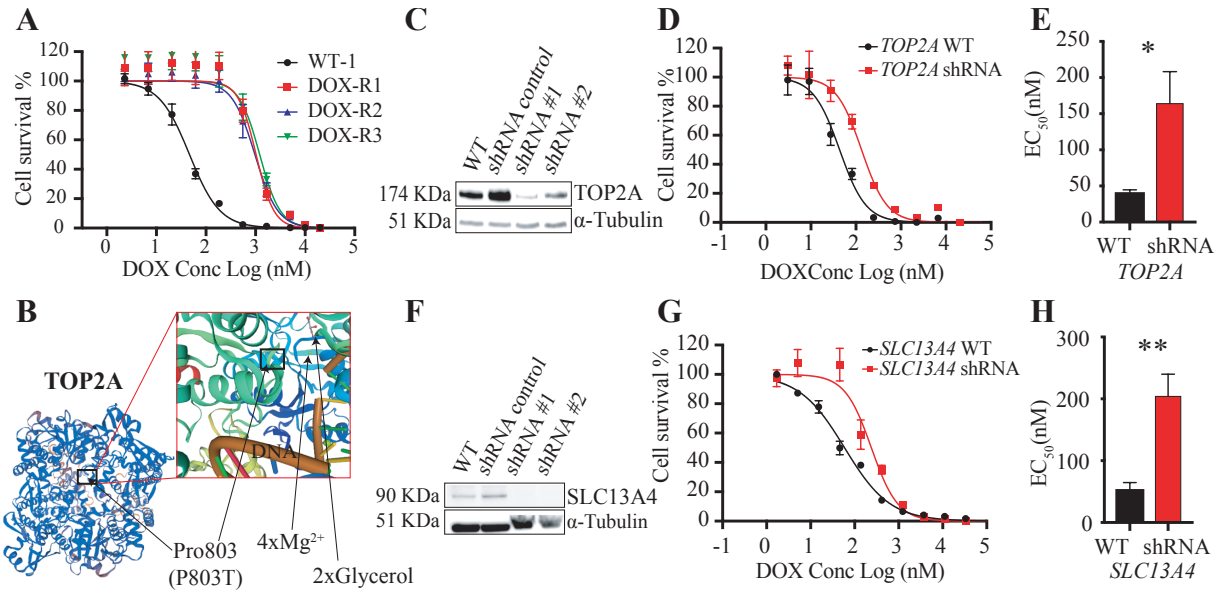

Fig. 2. A. EC_{50} curves for initial screening of DOX resistance **B.** Crystal structure of *TOP2A* highlighting (black box) the allele Pro803Thr next to the principal drug-binding DNA locus, Tyr805 (structure taken from PDBe 4FM9[28]). **C.** Western blot confirming that shRNA gene depletion downregulates protein levels for *TOP2A*. **D.** EC_{50} curves of the WT and shRNA knockdown cell lines for *TOP2A*. **E.** Barplot of the WT and shRNA knockdown cell lines for *TOP2A*. **F.** Western blot confirming that shRNA gene depletion downregulates protein levels for *SLC13A4*. **G.** EC_{50} curves of the WT and shRNA knockdown cell lines for *SLC13A4*. **H.** Barplot of the WT and shRNA knockdown cell lines for *SLC13A4*. Data is represented by mean ± s.e.m. with n=3 biological replicates; and n=8 (*TOP2A*) and n=4 (*SLC13A4*) technical replicates for every concentration point. * = p value < 0.05, ** = p < 0.01. p values determined by two-tailed *t* test.

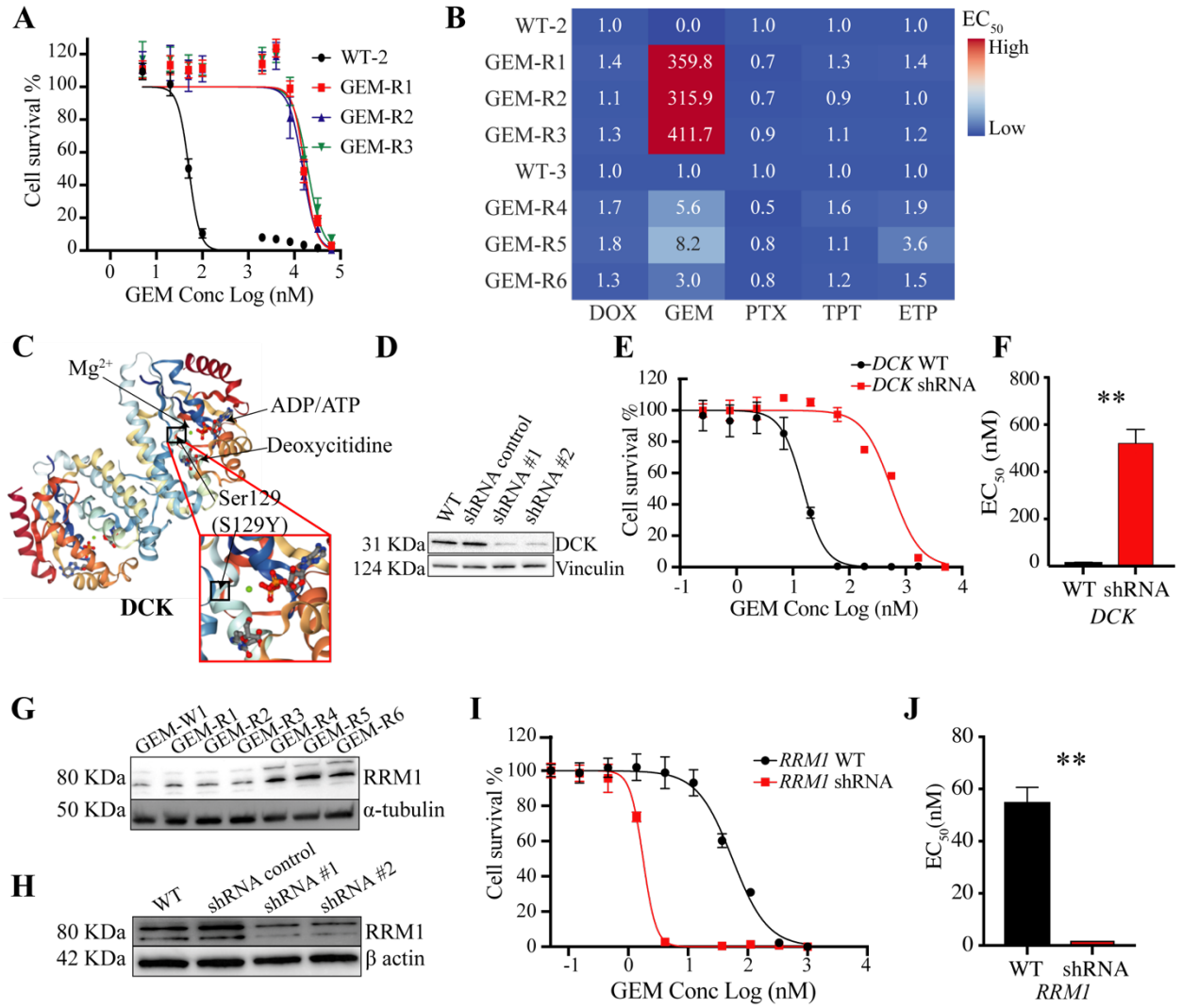

Fig. 3. A. EC_{50} curves for initial screening of GEM resistance using n=8 technical replicates per concentration point **B.** EC_{50} ratio matrix showing absence of multidrug resistance pathways of GEM resistant lines. **C.** Crystal structure of deoxycytidine kinase (*DCK*) showing the location of the Ser129Tyr (marked by a black box) allele near the deoxycytidine and ATP/ADP binding pocket (structure taken from PDBe 2A7Q[52]). **D.** Western blot confirming that shRNA gene depletion downregulates protein levels for *DCK*. **E.** EC_{50} curves of the WT and shRNA knockdown cell lines for *DCK*. n=8 with individual technical replicates overlaid for every concentration point. **F.** Barplot of the WT and shRNA knockdown cell lines for *DCK*. **G.** Western blot for *RRM1* across all GEM samples showing overexpression pattern of *RRM1* in GEM-R4-6 resistant clones. *γ-tubulin* is used as a loading control. **H.** Western blot confirming that shRNA gene depletion downregulates protein levels for *RRM1*. *β-actin* is used as a loading control. **I.** EC_{50} curves of the WT and shRNA knockdown cell lines for *RRM1*. n=4 with individual technical replicates overlaid for every concentration point. **J.** Barplot of the WT and shRNA knockdown cell lines for *RRM1*. All data is represented by mean ± s.e.m. with n=3 with individual biological replicates overlaid. ** = p value < 0.01. p values determined by two-tailed *t* test.

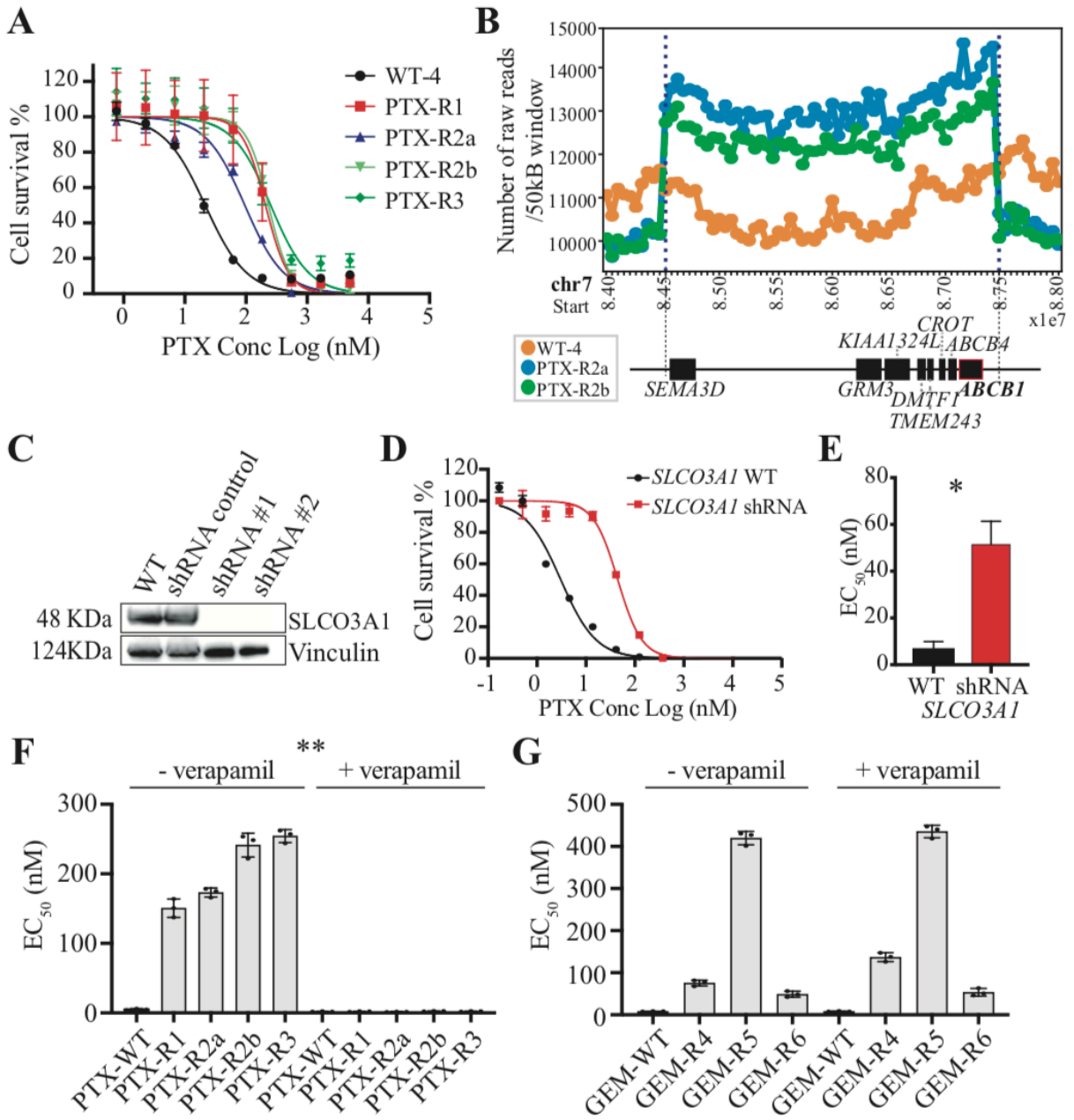

Fig. 4. A. EC_{50} curves for initial screening of PTX resistance with n=8 technical replicates. Values are given in Table S2. **B.** Raw copy number profile for the amplification event containing protein coding genes including ABC transporters (*ABCB1/ABCB4)* at PTX cell lines. The amplification region (chr7:84,500,000-87,300,000) had a higher number of raw reads (labeled with blue dash lines) with default window size of 50K bp. Genes associated with the CNV event are depicted by black boxes underneath according to their position and sizes. *ABCB1* is highlighted with red outline. **C**. Western blot confirming that shRNA gene depletion downregulates protein levels for *SLCO3A1*. **D.** EC_{50} curves of the WT and shRNA knock-down cell lines. **E.** Barplot of the WT and shRNA knockdown cell lines for *SLCO3A1*. **F.** Barplot of EC_{50} of the PTX treated cell lines with and without verapamil and verapamil alone showing sensitization in presence of verapamil

as ABC inhibitor (n=4 technical replicates). **G.** Barplot of EC_{50} of the GEM cell lines ± verapamil showing no EC_{50} shift for GEM cell lines when co-treated with verapamil with n=4 technical replicates. All data is represented by mean ± s.e.m. with n=3 with individual biological replicates overlaid. ** = p value < 0.01. p values determined by two-tailed *t* test.

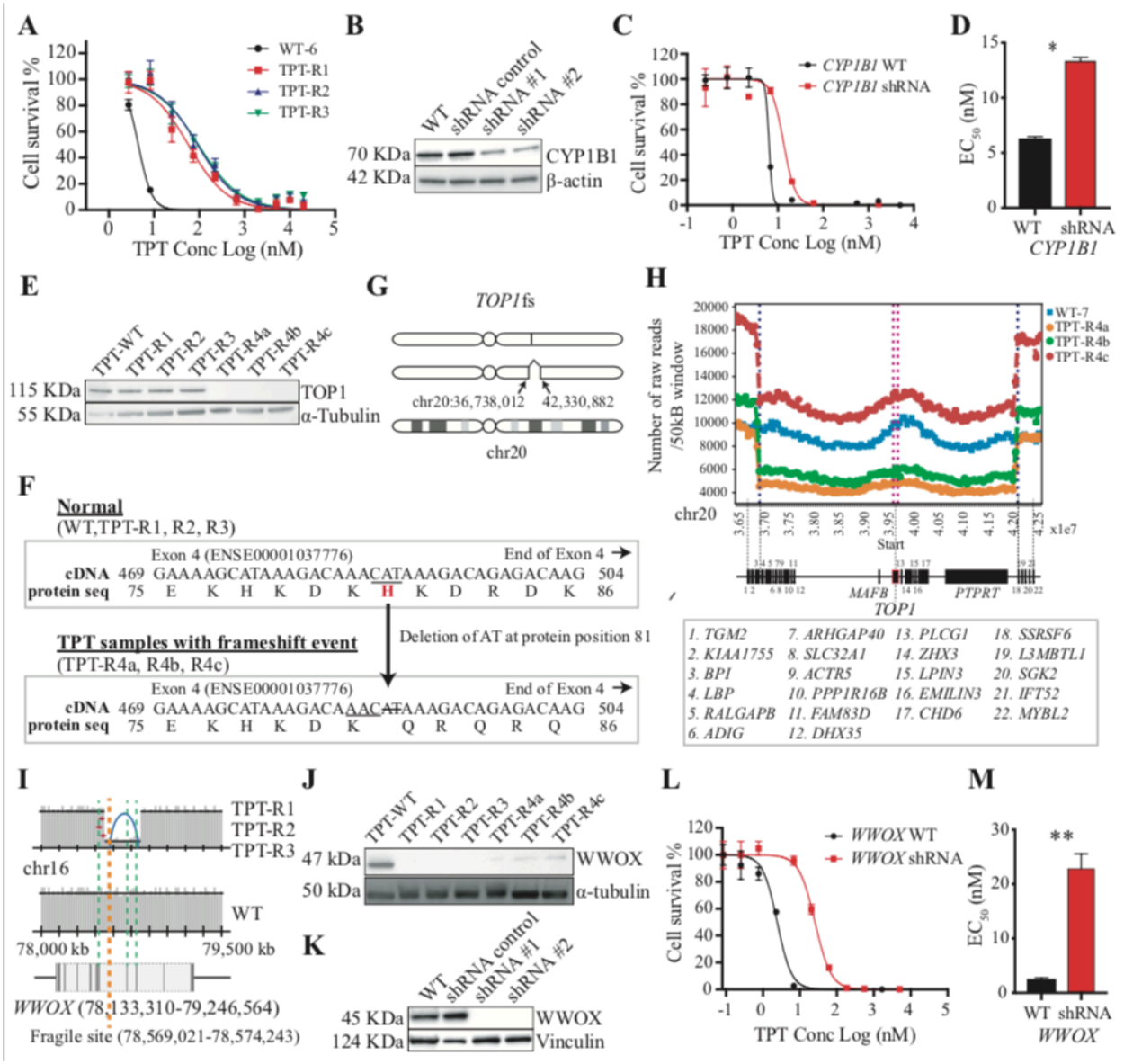

Fig. 5. A. EC_{50} curves for initial screening of TPT resistance with n=8 technical replicates. **B.** Western blot confirming that shRNA gene depletion downregulates protein levels for *CYP1B1*. **C.** EC_{50} curves of the WT and shRNA knock-down cell lines. **D.** Barplot of the WT and shRNA knockdown cell lines for *CYP1B1*. **E.** Western blot with their gene depletion downregulates protein levels for *TOP1*. **F.** Transcript sequence (cDNA) and protein sequence for TOP1 transcript for normal (top) and altered (bottom) cases. The figure shows only part of the cDNA (position 469-504) and protein sequence (position 75-86) for TOP1 at the affected exon (Exon 4, ENSE00001037776). The frameshift deletion of nucleotides 'CAT' to 'C' observed in TPT samples (TPT-R4a, R4b, and R4c) is predicted to give a frameshift at amino acid 81 (His, red highlight in normal). Amino acids affected by the frameshift deletion are highlighted in red. **G.** Complex alteration for regions surrounding *TOP1* gene. Schematic of chr20 reads around *TOP1* showing the frameshift event (top), deletion (middle), and the WT with characteristic banding. **H.**

Raw copy number profiles of *TOP1* deletion at TPT cell lines. The deleted regions (hg19, chr20: 36,738,012-42,330,882) have lower numbers of raw reads (labeled with blue dash lines) than WT-7 with default window size of 50K bp. *TOP1* is labeled with magenta dashed lines. **I.** Schematic of chr16 reads around *WWOX* for TPT-R1, R2, and R3 compares to the WT chromosome 16 parental cell line. Blue arch represents a deleted region and red arch represents an amplified region. Red arrows show insertion events. *WWOX* below shows the exonic (black lines) and intronic (white box) regions of the gene. The start of the deletion event is also close to a known fragile site (orange dashed line). **J.** Western blot showing WWOX protein levels in TPT resistant clones. **K.** Western blot shows downregulation of protein levels for *WWOX* in shRNA samples compared to WT and scrambled control. **L.** EC_{50} curves of the WT and shRNA knock-down cell lines with n=4 technical replicates per concentration point. **M.** Barplot of the WT and shRNA knockdown cell lines for *WWOX*. Data is represented by mean ± s.e.m. with n=3 biological replicates. ** = $p < 0.01$. p values determined by two-tailed *t* test.

TABLES

			WGS				***WES***		
	DOX (n=3)	GEM (n=3)	PTX (n=4)	TPT (n=3)	TPT (n=3)	DOX (n=3)	ETP (n=3)	GEM (n=3)	PTX (n=3)
Indels									
Disruptive inframe ins.	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00
Frameshift	0.00	1.00	1.33	2.33	0.00	1.00	1.00	1.00	1.00
Frameshift plus stop-gained	0.00	0.00	0.00	0.33	0.00	0.00	1.00	0.00	0.00
Inframe insertion	0.00	0.00	1.00	0.33	0.00	0.00	1.00	0.00	0.00
Intergenic	27.67	43.00	26.75	24.67	47.67	2.00	3.50	1.00	1.00
Intragenic	10.00	5.67	9.00	9.00	14.33	1.00	2.00	1.00	1.00
Intron	12.00	20.33	16.25	15.67	32.33	1.00	1.00	4.00	1.50
Splice region plus intron	0.00	0.00	0.00	103.3	0.00	0.00	0.00	1.00	0.00
SNVs									
Disruptive inframe del.	0.00	1.00	1.00	0.67	0.00	1.00	0.00	1.00	1.33
Frameshift	1.00	2.00	1.00	22.33	2.00	3.00	1.67	1.00	3.00
Inframe deletion	0.00	0.00	1.00	0.00	0.00	0.00	2.00	0.00	0.00
Intergenic	898.33	1303.3	1416.6	258.67	2635.7	25.00	19.00	17.00	21.33
Intragenic	272.00	403.33	389.25	77.67	834.67	15.33	12.33	10.33	11.33
Intron	448.67	701.33	764.00	128.3	1358.33	28.00	27.00	33.00	22.67
Missense	16.00	14.00	12.75	7.00	34.33	15.00	19.67	21.67	15.67
Others	1.00	1.00	0.00	0.00	1.33	1.00	1.00	1.50	1.00
Splice region plus intron	1.00	1.67	2.00	30.00	3.33	1.67	1.50	1.00	1.33
Start lost	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
Stop gained	0.00	2.50	1.00	0.00	1.67	1.50	2.00	2.67	0.00
Stop lost	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Synonymous	3.00	7.33	4.00	6.33	7.67	6.67	5.33	7.67	5.33

Table 1. Summary of average number of mutations. Number of selections performed with the drug is given by n. SNVs and Indels were grouped according to snpEff sequence ontology annotations (Methods, Table S4), and detailed counts per clone can be found in Table S5.

Drug	*Sample*	*Gene*	*Amino Acid Change*	*Type*	*AF*	*EC_{50} (nM) WT*	*EC_{50} (nM) KD/KO/CI*
DOX	R1, R2, R3	TOP2A	Pro803Thr	MS	0.89, 0.87	38.6 ± 4.3	164.3 ± 43.9
	R4b	SLC13A4	Gly165His	MS	1	52.9 ± 11.6	204.3 ± 35.7
	R5	SPG7	Lys593Asn	MS	1	NE*	
PTX	R1, R2a, R3	WWOX	16q23.1	CNV	-	5.8 ± 2.5	42.9 ± 11.7
	R2a, R2b	ABCB1	7q21.12	CNV	-	252 ± 38 218 ± 14	1.3 ± 0.1 1.3 ± 0.1
	R2b, R6	SLCO3A1	Ile587Asn (R2b), Ala263Thr (R6)	MS		6.6 ± 3.3	51.5 ± 9.9
GEM	R1, R2, R3	DCK	Ser129Tyr (R1, R2), Asn80fs (R1. R3); Asn113fs (R2); Thr184fs (R3)	MS, FS		14.3 ± 1.7	521.7 ± 58.3
	R4, R5, R6	RRM1	11p15.4	CNV	-	54.9 ± 5.8	1.8 ± 0.1
ETP	R2, R3	WNT3A	1q42.13 (R2)	CNV	-	ND	-
	R3	WDR33	P622T	MS	1	241.5 ± 31.0	821.6 ± 226.9
TPT	R1, R4a, R4b, R4c	CYP1B1	Val432Leu; Asp217Glu (R4a,b,c)	MS	0.13, 0.40, 0.43, 0.42	6.3 ± 0.2	13.3 ± 0.3
	R1, R2, R3	WWOX	16q23.1	CNV	-	2.4 ± 0.3	22.8 ± 2.7
	R4a, R4b, R4c	TOP1	His81fs; 20q12	FS; CNV	1	ND	-
	R4a, R4b, R4c	USP47	Arg408*	Stop	0.38, 0.57, 0.58	3.05± 0.2	1.07± 0.07

Table 2. Validation (knockdown) results for selected genes. CNV, copy-number variant. MS, missense. FS, frameshift variant, KO/KD/CI, knockout, knockdown, chemical inhibition (verapamil, ABCB1). ND: No data: gene knockdowns were attempted but could not be achieved. NE: Not expressed (protein not detected by Western blot, preventing validation). *EC_{50} WT and EC_{50} KO/KD/CI* are from matched pairs for the given drug and represent the mean ± s.e.m. (n=3 biological replicates).

References

1 Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87, doi:10.1126/science.1247005 (2014).

2 Kanarek, N. *et al.* Histidine catabolism is a major determinant of methotrexate sensitivity. *Nature* **559**, 632-636, doi:10.1038/s41586-018-0316-7 (2018).

3 Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84, doi:10.1126/science.1246981 (2014).

4 Antonova-Koch, Y. *et al.* Open-source discovery of chemical leads for next-generation chemoprotective antimalarials. *Science* **362**, doi:10.1126/science.aat9446 (2018).

5 Cowell, A. N. *et al.* Mapping the malaria parasite druggable genome by using in vitro evolution and chemogenomics. *Science* **359**, 191-199, doi:10.1126/science.aan4472 (2018).
6 Ottilie, S. *et al.* Rapid Chagas Disease Drug Target Discovery Using Directed Evolution in Drug-Sensitive Yeast. *ACS Chem Biol* **12**, 422-434, doi:10.1021/acschembio.6b01037 (2017).
7 Andries, K. *et al.* A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science* **307**, 223-227, doi:1106753 [pii]
10.1126/science.1106753 (2005).
8 Santos-Lopez, A., Marshall, C. W., Scribner, M. R., Snyder, D. J. & Cooper, V. S. Evolutionary pathways to antibiotic resistance are dependent upon environmental structure and bacterial lifestyle. *Elife* **8**, doi:10.7554/eLife.47612 (2019).
9 Burckstummer, T. *et al.* A reversible gene trap collection empowers haploid genetics in human cells. *Nat Methods* **10**, 965-971, doi:10.1038/nmeth.2609 (2013).
10 Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340-343, doi:10.1038/nature10348 (2011).
11 Mandell, J. B. *et al.* Combination Therapy with Disulfiram, Copper, and Doxorubicin for Osteosarcoma: In Vitro Support for a Novel Drug Repurposing Strategy. *Sarcoma* **2019**, 1320201, doi:10.1155/2019/1320201 (2019).
12 Jiang, C. *et al.* Pyrroloquinoline quinine ameliorates doxorubicin-induced autophagy-dependent apoptosis via lysosomal-mitochondrial axis in vascular endothelial cells. *Toxicology* **425**, 152238, doi:10.1016/j.tox.2019.152238 (2019).
13 Drenberg, C. D. *et al.* A high-throughput screen indicates gemcitabine and JAK inhibitors may be useful for treating pediatric AML. *Nat Commun* **10**, 2189, doi:10.1038/s41467-019-09917-0 (2019).
14 Alvarellos, M. L. *et al.* PharmGKB summary: gemcitabine pathway. *Pharmacogenet Genomics* **24**, 564-574, doi:10.1097/FPC.0000000000000086 (2014).
15 Mini, E., Nobili, S., Caciagli, B., Landini, I. & Mazzei, T. Cellular pharmacology of gemcitabine. *Ann Oncol* **17 Suppl 5**, v7-12, doi:10.1093/annonc/mdj941 (2006).
16 Pommier, Y., Leo, E., Zhang, H. & Marchand, C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* **17**, 421-433, doi:10.1016/j.chembiol.2010.04.012 (2010).
17 Hayashi, M. T. & Karlseder, J. DNA damage associated with mitosis and cytokinesis failure. *Oncogene* **32**, 4593-4601, doi:10.1038/onc.2012.615 (2013).
18 Rasheed, Z. A. & Rubin, E. H. Mechanisms of resistance to topoisomerase I-targeting drugs. *Oncogene* **22**, 7296-7304, doi:10.1038/sj.onc.1206935 (2003).
19 Corey, V. C. *et al.* A broad analysis of resistance development in the malaria parasite. *Nat Commun* **7**, 11901, doi:10.1038/ncomms11901 (2016).
20 Christowitz, C. *et al.* Mechanisms of doxorubicin-induced drug resistance and drug resistant tumour growth in a murine breast tumour model. *BMC Cancer* **19**, 757, doi:10.1186/s12885-019-5939-z (2019).
21 Amrutkar, M. & Gladhaug, I. P. Pancreatic Cancer Chemoresistance to Gemcitabine. *Cancers (Basel)* **9**, doi:10.3390/cancers9110157 (2017).
22 Pan, P., Li, Y., Yu, H., Sun, H. & Hou, T. Molecular principle of topotecan resistance by topoisomerase I mutations through molecular modeling approaches. *Journal of chemical information and modeling* **53**, 997-1006, doi:10.1021/ci400066x (2013).

23 Andersson, B. S. *et al.* KBM-7, a human myeloid leukemia cell line with double Philadelphia chromosomes lacking normal c-ABL and BCR transcripts. *Leukemia* **9**, 2100-2108 (1995).

24 Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423-425, doi:10.1093/bioinformatics/btr670 (2012).

25 Jeon, K. H., Yu, H. V. & Kwon, Y. Hyperactivated m-calpain affects acquisition of doxorubicin resistance in breast cancer cells. *Biochim Biophys Acta Gen Subj* **1862**, 1126-1133, doi:10.1016/j.bbagen.2018.02.002 (2018).

26 Turner, J. G. *et al.* Treatment of acquired drug resistance in multiple myeloma by combination therapy with XPO1 and topoisomerase II inhibitors. *J Hematol Oncol* **9**, 73, doi:10.1186/s13045-016-0304-z (2016).

27 Ghisoni, E. *et al.* TOP2A as marker of response to pegylated lyposomal doxorubicin (PLD) in epithelial ovarian cancers. *J Ovarian Res* **12**, 17, doi:10.1186/s13048-019-0492-6 (2019).

28 Wendorff, T. J., Schmidt, B. H., Heslop, P., Austin, C. A. & Berger, J. M. The structure of DNA-bound human topoisomerase II alpha: conformational mechanisms for coordinating inter-subunit interactions with DNA cleavage. *J Mol Biol* **424**, 109-124, doi:10.1016/j.jmb.2012.07.014 (2012).

29 Burgess, D. J. *et al.* Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proc Natl Acad Sci U S A* **105**, 9053-9058, doi:10.1073/pnas.0803513105 (2008).

30 Lim, M. Y. *et al.* UDP-galactose and acetyl-CoA transporters as Plasmodium multidrug resistance genes. *Nat Microbiol* **1**, 16166, doi:10.1038/nmicrobiol.2016.166 (2016).

31 Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S. & Baradaran, B. The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv Pharm Bull* **7**, 339-348, doi:10.15171/apb.2017.041 (2017).

32 Bepler, G. *et al.* RRM1 modulated in vitro and in vivo efficacy of gemcitabine and platinum in non-small-cell lung cancer. *J Clin Oncol* **24**, 4731-4737, doi:10.1200/JCO.2006.06.1101 (2006).

33 Zeng, C., Fan, W. & Zhang, X. RRM1 expression is associated with the outcome of gemcitabine-based treatment of non-small cell lung cancer patients--a short report. *Cellular oncology (Dordrecht)* **38**, 319-325, doi:10.1007/s13402-015-0225-9 (2015).

34 Yang, Z. *et al.* RRM1 predicts clinical outcome of high-and intermediate-risk non-muscle-invasive bladder cancer patients treated with intravesical gemcitabine monotherapy. *BMC urology* **19**, 69, doi:10.1186/s12894-019-0497-x (2019).

35 Chan, S. L. *et al.* CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev* **28**, 2370-2380, doi:10.1101/gad.250993.114 (2014).

36 Teloni, F. *et al.* Efficient Pre-mRNA Cleavage Prevents Replication-Stress-Associated Genome Instability. *Mol Cell* **73**, 670-683 e612, doi:10.1016/j.molcel.2018.11.036 (2019).

37 Hansen, L. T., Lundin, C., Spang-Thomsen, M., Petersen, L. N. & Helleday, T. The role of RAD51 in etoposide (VP16) resistance in small cell lung cancer. *Int J Cancer* **105**, 472-479, doi:10.1002/ijc.11106 (2003).

38 Huang, F. & Mazin, A. V. A small molecule inhibitor of human RAD51 potentiates breast cancer cell killing by therapeutic agents in mouse xenografts. *PLoS ONE* **9**, e100993, doi:10.1371/journal.pone.0100993 (2014).

39 Thakkar, N., Lockhart, A. C. & Lee, W. Role of Organic Anion-Transporting Polypeptides (OATPs) in Cancer Therapy. *The AAPS journal* **17**, 535-545, doi:10.1208/s12248-015-9740-x (2015).

40 Vaidyanathan, A. *et al.* ABCB1 (MDR1) induction defines a common resistance mechanism in paclitaxel- and olaparib-resistant ovarian cancer cells. *Br J Cancer* **115**, 431-441, doi:10.1038/bjc.2016.203 (2016).

41 Cruz-Munoz, W. *et al.* Analysis of acquired resistance to metronomic oral topotecan chemotherapy plus pazopanib after prolonged preclinical potent responsiveness in advanced ovarian cancer. *Angiogenesis* **17**, 661-673, doi:10.1007/s10456-014-9422-9 (2014).

42 Wu, C. P., Calcagno, A. M. & Ambudkar, S. V. Reversal of ABC drug transporter-mediated multidrug resistance in cancer cells: evaluation of current strategies. *Current molecular pharmacology* **1**, 93-105, doi:10.2174/1874467210801020093 (2008).

43 Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* **45**, D777-d783, doi:10.1093/nar/gkw1121 (2017).

44 Del Mare, S., Salah, Z. & Aqeilan, R. I. WWOX: its genomics, partners, and functions. *J Cell Biochem* **108**, 737-745, doi:10.1002/jcb.22298 (2009).

45 Ariey, F. *et al.* A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature* **505**, 50-55, doi:10.1038/nature12876 (2014).

46 Luth, M. R., Gupta, P., Ottilie, S. & Winzeler, E. A. Using in Vitro Evolution and Whole Genome Analysis To Discover Next Generation Targets for Antimalarial Drug Discovery. *ACS Infect Dis* **4**, 301-314, doi:10.1021/acsinfecdis.7b00276 (2018).

47 Si, W., Shen, J., Zheng, H. & Fan, W. The role and mechanisms of action of microRNAs in cancer drug resistance. *Clin Epigenetics* **11**, 25, doi:10.1186/s13148-018-0587-8 (2019).

48 Sanchez Calle, A., Kawamura, Y., Yamamoto, Y., Takeshita, F. & Ochiya, T. Emerging roles of long non-coding RNA in cancer. *Cancer Sci* **109**, 2093-2100, doi:10.1111/cas.13642 (2018).

49 Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat Rev Cancer* **16**, 483-493, doi:10.1038/nrc.2016.62 (2016).

50 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92, doi:10.4161/fly.19695 (2012).

51 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

52 Zhang, Y., Secrist, J. A., 3rd & Ealick, S. E. The structure of human deoxycytidine kinase in complex with clofarabine reveals key interactions for prodrug activation. *Acta Crystallogr D Biol Crystallogr* **62**, 133-139, doi:10.1107/S0907444905034293 (2006).