



OPEN In vitro evolution and whole genome analysis to study chemotherapy drug resistance in haploid human cells

Juan Carlos Jado^{1,2,8}, Michelle Dow^{2,3,4,8}, Krypton Carolino⁵, Adam Klie^{2,3}, Gregory J. Fonseca⁶, Trey Ideker^{2,7}✉, Hannah Carter^{2,7}✉ & Elizabeth A. Winzeler^{1,2}✉

In vitro evolution and whole genome analysis has proven to be a powerful method for studying the mechanism of action of small molecules in many haploid microbes but has generally not been applied to human cell lines in part because their diploid state complicates the identification of variants that confer drug resistance. To determine if haploid human cells could be used in MOA studies, we evolved resistance to five different anticancer drugs (doxorubicin, gemcitabine, etoposide, topotecan, and paclitaxel) using a near-haploid cell line (HAP1) and then analyzed the genomes of the drug resistant clones, developing a bioinformatic pipeline that involved filtering for high frequency alleles predicted to change protein sequence, or alleles which appeared in the same gene for multiple independent selections with the same compound. Applying the filter to sequences from 28 drug resistant clones identified a set of 21 genes which was strongly enriched for known resistance genes or known drug targets (*TOP1*, *TOP2A*, *DCK*, *WDR33*, *SLCO3A1*). In addition, some lines carried structural variants that encompassed additional known resistance genes (*ABCB1*, *WWOX* and *RRM1*). Gene expression knockdown and knockout experiments 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, discovered and studied in an isogenic background.

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
TF	Transcription factor
DOX	Doxorubicin
GEM	Gemcitabine
ETP	Etoposide
PTX	Paclitaxel
TPT	Topotecan
AML	Acute myeloid leukemia

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. ⁶Department of Medicine, Meakins-Christie Laboratories, McGill University Health Centre, 1001 Decaire Blvd, Montreal, QC H4A 3J1, Canada. ⁷Moore's Cancer Center, University of California San Diego, La Jolla, CA 92093, USA. ⁸These authors contributed equally: Juan Carlos Jado and Michelle Dow. ✉email: tideker@ucsd.edu; hcarter@ucsd.edu; ewinzeler@ucsd.edu

TKIs	Tyrosine kinase inhibitors
MDR	Multi-drug resistance
gDNA	Genomic DNA

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 a drug are useful to broadly implicate relevant genes, but cannot readily reveal critical gain-of-function, single nucleotide alleles, such as 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. Furthermore, such studies also 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 cell lines are nevertheless available: HAP1, is a human chronic myelogenous leukemia (CML)-derived cell line that is completely haploid except for a 30 megabase fragment of chromosome 15⁹. HAP1 has been used for genetic studies because mutated phenotypes are immediately exposed^{10–15}.

Using five different anticancer drugs (Doxorubicin, Gemcitabine, Etoposide, Topotecan, and Paclitaxel) as examples, we show that in vitro evolution in HAP1 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 heterogeneous tumor cell environment.

Results

Selection of compounds for resistance studies

Our previous studies have shown that evolution works best with potent inhibitors of cell growth. We therefore sought to identify proof-of-concept compounds which would show 50% growth inhibition at concentrations of less than 10 μ M and ideally less than 1 μ M. To do this, 16 readily-available compounds were tested for the ability to reduce ATP levels (measured using CellTiterGlo) in HAP1 cells using a 48-h dose response assay. Although sometimes longer incubation times can be used in dose response assays, especially for slow growing cell lines, 48 h provided a good signal to noise ratio for our purposes and longer incubation times were not considered. The 16 compounds we tested had previously been tested against *S. cerevisiae*. Some were older drugs derived from natural products (e.g. rapamycin, topotecan) known to have activity against many different eukaryotic cell lines. Others were modern cancer drugs, including the kinase inhibitors, lisitinib, imatinib and sorafenib, which are predicted to act only against specific targets in specific cell lines. We include a separate section on the types of drugs applicable to this method in the discussion. Only five drugs showed EC₅₀ values between 5 and 340 nM (Fig. 1A,B, Table S1). These included doxorubicin (DOX, EC₅₀ = 95.50 \pm 54.56 nM), also known as adriamycin, an anthracycline antibiotic that works by inhibiting topoisomerase II alpha (*TOP2A*)^{16,17}; gemcitabine (GEM, EC₅₀ = 34.12 \pm 27.83 nM), a synthetic pyrimidine nucleoside prodrug that is used against a variety of hematopoietic malignancies^{18–20}; etoposide (ETP, EC₅₀ = 338.60 \pm 39.72 nM), a semisynthetic derivative of podophyllotoxin that also targets *TOP2A* and prevents re-ligation of the double-stranded DNA²¹; paclitaxel (PTX, EC₅₀ = 19.43 \pm 3.40 nM) also known as taxol, an effective anticancer agent that targets tubulin, perturbing the cytoskeleton and causing M phase cell-cycle arrest²², and topotecan (TPT, EC₅₀ = 4.81 \pm 1.12 nM), a semisynthetic water-soluble derivative of camptothecin that inhibits topoisomerase I (*TOP1*)²³. Our HAP1 EC₅₀ values were similar to those previously reported for other CML cell lines (www.cancerrxgene.org^{24,25}) with the exception of etoposide, which appeared more effective in HAP1 cells (EC₅₀ = 338.6 \pm 39.72 nM) relative to other CML cell lines (> 1 μ M in BV-173, KU812, EM-2, MEG-01, JURL-MK1, KCL-22, RPMI-8866, LAMA-84, K-562). As described below, ETP resistance seemed to be associated with CNVs which could potentially explain some of the difference between published values for diploid lines and these haploid cells.

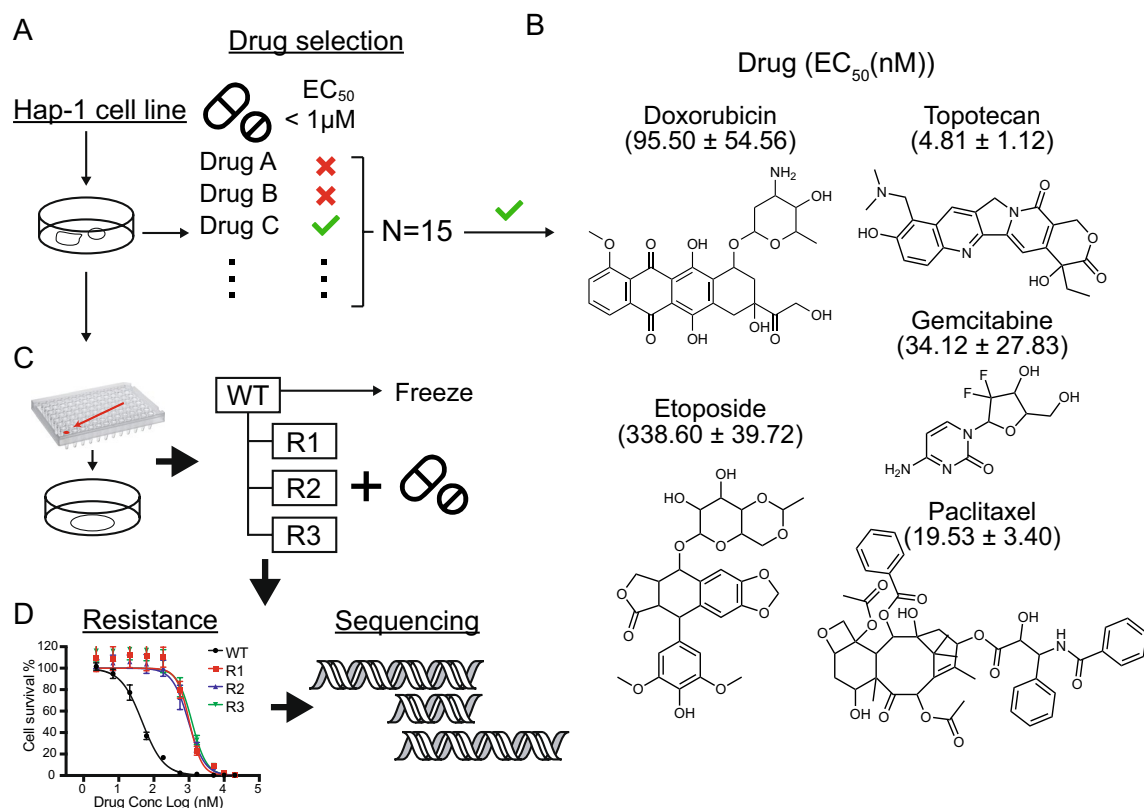


Figure 1. Experimental workflow. (A) Chemotherapy drug evaluation. EC₅₀ dose response assays were performed on 16 different chemotherapeutic agents (Table S1). Only drugs to which HAP1 cells were sensitive (EC₅₀ value below 1 μM) were considered for IVIEWGA. (B) Chemical structures of the chemotherapy agents ultimately used for IVIEWGA. EC₅₀ values are presented as the mean ± s.e.m., for n = 3 biological replicates and n = 8 with technical replicates per concentration point. (C) Clone selection. To ensure a homogenous genetic background limiting dilution cloning was used to isolate individual cells prior to drug selection. For each drug three independent selections were performed. Resistance was confirmed using dose–response assays (D) Drug resistance was achieved in 7–30 weeks approximately (49 and 210 generations). The parental cell line and the drug resistant lines were then sequenced.

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”²⁶), there is precedent for resistance to the drugs included here^{27–29}. To reduce the possibility of selecting mutations that were present in a non-homogenous population of HAP1 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. 1C) 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 first grown in tissue culture dishes (reaching 60–80% semi-confluence) in the presence of sublethal concentrations of each drug. Most cell lines (DOX, GEM, TPT and PTX resistant clones) were then subjected to a lethal concentration (~3–5 × EC₅₀ value), killing more than 95% of the cells. Treatment was removed until cells reached semi-confluence again (doubling every 22 h³⁰) whereupon drug at ~the EC₉₅ value was reapplied. Alternatively, for ETP-resistant clones, a stepwise 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. Although others have used mutagenesis³¹, we have found that this can increase the rate of background mutations, which would complicate an already difficult analysis. 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 EC₅₀ 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 EC₅₀ 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 comprehensive IVIEWGA studies in *Plasmodium*⁵, have shown that stable drug resistance is most frequently conferred by SNVs in coding regions and thus exome sequencing seemed initially the most cost-effective 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 41,259 SNVs (Table S4), of which 26,625 were unique (Table S5, “Methods”). The majority of variants in all cell lines were non-coding (Tables S4, S5) and were evenly distributed with respect to chromosome length (Fig. S2). Of the 26,625 mutations almost all (26,468) were present at allele frequencies (AF) of less than 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. The five drugs varied in the number of mutations, with TPT having the highest overall numbers (Table 1).

We next developed a pipeline (Fig. S3A, “Methods”) to filter the 26,625 “called” mutations (Table S5) to a final list of potential variants conferring drug resistance (Table S6). Our 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

	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 S9), and detailed counts per clone can be found in Table S4.

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, as determined by sequencing read count). The top 2.5% of highest AF mutations corresponded to an AF > 0.85 (Fig. S3B). At this cutoff, the majority of cell lines harbored a candidate resistance mutation. While selecting a cutoff represents a tradeoff with potentially missed relevant mutations, the full list of mutations is provided in the supplement (Table S5). We did not note any strong correlation between read depth and allelic fraction in our study ($R^2 = 0.06$; Fig. S3C) and all of the final mutations selected for further analysis had a read depth > 10 reads, with the majority supported by over 20 (Fig. S3D). 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, our experience with microbes has shown that genes which are not from gene families that are known to be excessively variable and which show multiple, independently derived amino acid changes will have high rates of confirmation using independent methods. Considering this allowed us to add in 4 genes (*STARD9*, *CYP1B1*, *SLCO3A1* and *DCK*) to our final list of 21 candidates (Table S6) despite these having slightly lower AF numbers (e.g. *DCK* AFs were 0.78 or less). We also identified a group of 298 genes with lower AF values that were either recurrent with disruptive alleles or were found in CNVs (Table S7).

Although they may theoretically play a role in resistance, we did not consider intergenic mutations nor intron variants because their high numbers suggested they could more likely arise by chance. Altogether only 465 (~ 1%) of the 41,259 total SNVs were predicted missense mutations in our complete dataset (Tables S4, S5) while 11,613 and 19,127 variants encoded a non-splice site intronic variant (underestimation due to WES samples) or intergenic mutation, respectively. Despite this, we did find several intronic mutations in genes that have a high probability of conferring resistance, including in *SLCO3A1* (3) and *TOP1* (1). We found two intra-genic mutations upstream of *DCK* but one was in an unexpected line (PTX not GEM) and both were more than 100,000 kb away from *DCK*.

Somatic copy number variations (CNVs)

We next searched for CNVs (both amplifications and deletions) in our WGS and WES data using Control-FreeC³². Overall patterns of broad and focal alterations across the drugs and conditions varied (Fig. S4A, Table S8). 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 S8). The CNVs had an average size of 8.5 Mbp (stdev 19 Mbp), ranged from 15,000 bp to 152 Mbp (Fig. S4A) 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. S4A). It is likely that some CNVs were also missed in the WES data. The number of events was proportional to chromosome size, with the exception of the Y chromosome, for which there were ~ 4 × more events (47) per unit length. Some CNV calls were supported by paired end red data, for example, the one near *WWOX* (Fig. S4B,C).

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 (Fig. 2A)) and by WES (DOX-R4a, DOX-R4b and DOX-R5). High allele frequency missense mutations were found in only 11 genes (Table S5). Of note, DOX-R2 and DOX-R3 carried mutations in *TOP2A* at frequencies of 0.89 and 0.87, respectively. *TOP2A* is the known target of DOX^{21,33} and is known to play a role in drug resistance^{33–35}. The amino acid mutation, Pro803Thr (Fig. 2B,C), 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 (Fig. 2B,C)³⁶. The Pro803 site is highly conserved in metazoans as is *TOP2A* generally (Fig. 2D). Examination of the human *TOP2A* DNA-bound crystal structure suggests that Pro803Thr is a gain of function mutation that may block DOX-intercalated DNA from binding, thus preventing the *TOP2A*/DNA/DOX adduct from poisoning the cell. Alternatively, it could be 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 system³⁷. Downregulation of *TOP2A* using a pool of target-specific shRNA hairpins (confirmed by Westerns, Fig. 2E) resulted in a modest resistance (~ 4X, Table 2) that was not equivalent to the resistance seen in the evolved cell lines (Fig. 2E,G). Although these knockdown results recapitulated those of others, we favor the hypothesis that the mutation is a gain-of-function allele. Indeed, others have identified and validated the *TOP2A* Pro803 residue as important for drug resistance using recombinant protein expression and a yeast model³⁸.

We also observed some lines bore 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, AF = 1), as well as one uncharacterized gene (AC091801.1, His13Asn, DOX-R4a) in the three different clones that were subjected to WES and which 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 transporter³⁹) and are also associated with cancer drug resistance⁴⁰. The Gln165His mutation is located in a relatively disordered region of the protein making it less attractive as a candidate. *SPG7* (also called paraplegin) is a gene associated with spastic paraplegia in humans and would initially not seem to be a plausible drug resistance gene. It encodes one subunit of the mAAA (matrix ATPase associated with diverse cellular activities) protease, which is located in the mitochondrial matrix⁴¹. In particular, mutations to mitochondrial genes in *C. elegans* have been shown to confer resistance to hemiasterlin, an antimitotic agent that functions by disrupting microtubule dynamics⁴². Human variants have been shown to elevate mitochondrial

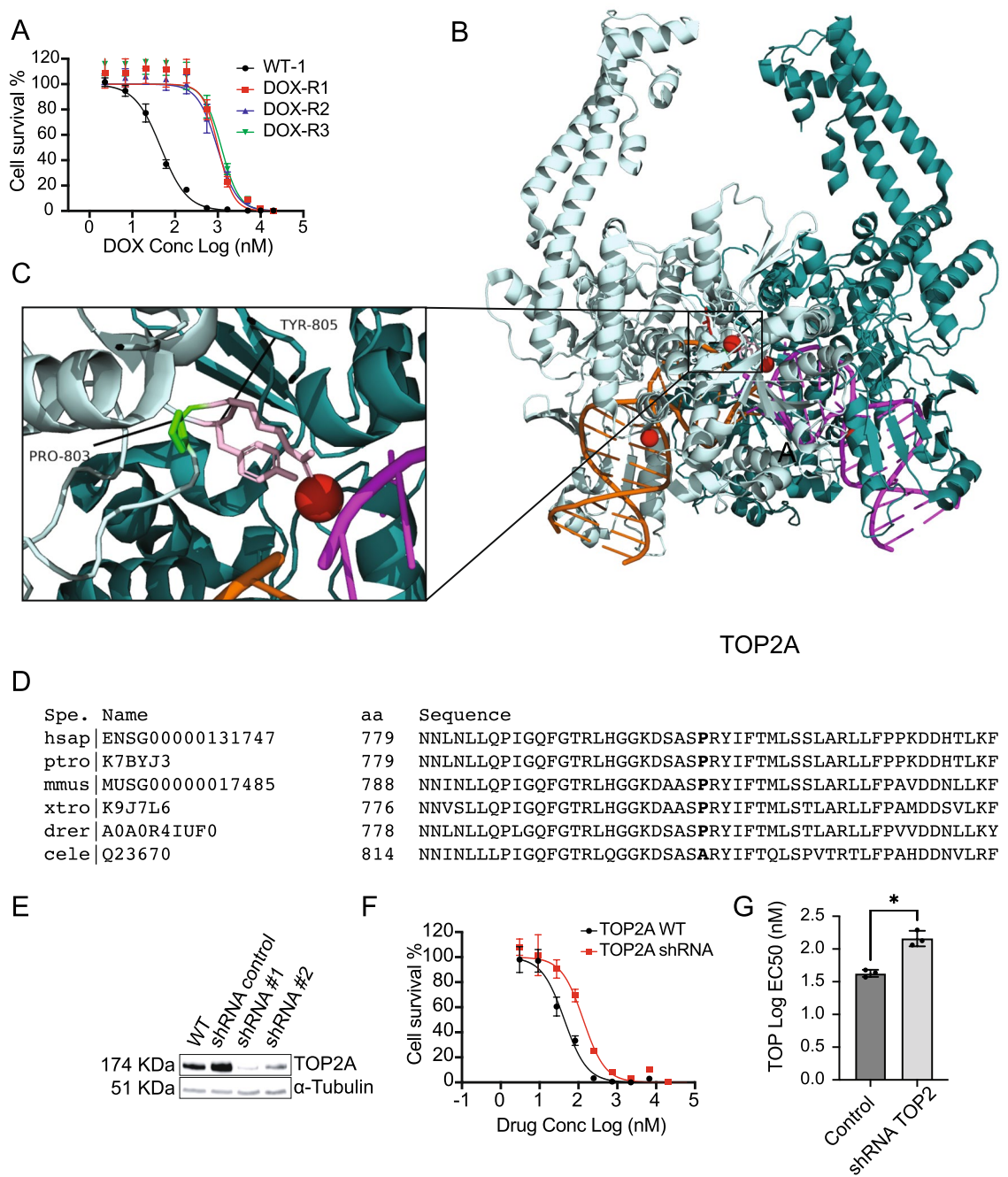


Figure 2. (A) DOX EC₅₀ curves for DOX evolved clones using 8 technical replicates for each concentration (B) Crystal structure of human TOP2A dimer (subunits shown in teal and pale cyan) in complex with DNA (4FM9) and inset (C) showing location of Pro803 (green) in relationship to catalytic arginine and tyrosine (804, 805, pink) and Mg²⁺ (red). (D) Alignments of the TOP2A sequence subsets showing conservation of the Pro803 residue (bold) in all vertebrates but not *Caenorhabditis elegans* (cele). (E) Western blot confirming that shRNA gene depletion downregulates TOP2A protein level. shRNA #1 and shRNA #2 indicate independent biological replicates with same pool. (F) EC₅₀ curves of the WT and shRNA (shRNA #1) knockdown cell lines for TOP2A. (G) Barplot of the WT and shRNA knockdown cell line (shRNA #1) for TOP2A. *p value < 0.05. P values determined by a paired ratio *t* test with EC₅₀s determined by three independent biological replicates.

reactive oxygen species⁴¹. shRNA knockout experiments were not successful with *SPG7* and it seems likely that CRISPR-Cas9 editing would be needed to confirm a predicted gain of function, but it remains an attractive candidate based on work in other species. *SPG7* is one of 170 genes included in a set of drug metabolizing enzymes and transporters, that also includes *DCK* (see below), *CYP1B1*, *SLCO3A1* and *ABCB1* and shows an association with resistance to docetaxel⁴³ and as with *TOP2A*, the mutated region is highly conserved across vertebrates.

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	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). EC₅₀ WT and EC₅₀ KO/KD/CI are from matched pairs for the given drug and represent the mean ± s.e.m. (n = 3 biological replicates).

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 EC₅₀ shift of 300 to 400-fold (Fig. 3A, Table S2), and the clones showed no change in HAP1 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 but also falls within the known GEM binding pocket in a *DCK*/GEM crystal structure (1P62⁴⁴), making it an exceptionally strong causal candidate for GEM drug resistance (Fig. 3C). GEM is a prodrug that only becomes pharmacologically active if it is phosphorylated by *DCK*⁴⁴. A shRNA knockdown of *DCK* was performed and confirmed by western blot analysis (Fig. 3D). Although the mutation appears to be a gain of function allele, downregulation of the *DCK* nevertheless resulted in a 36.5-fold increase in the EC₅₀ 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 (~6× versus ~400×, Fig. 3G, Table S2). Our work in other species with well characterized compounds suggests this is not surprising and that even single nucleotide changes in the same gene can yield different levels of resistance. For example, repeated selections with dihydroorotate dehydrogenase (DHOD) inhibitors in a mouse model and in vitro culture gave rise to 13 different point mutations in parasite DHODH with levels of resistance ranging from 2- to ~400-fold⁴⁵. 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. S4). 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*), was 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*⁴⁶. Furthermore, overexpression of *RRM1* is associated with poorer prognosis after gemcitabine treatment in non-small cell lung cancer⁴⁷ and in bladder cancer⁴⁸.

Western blot analysis of the evolved lines showed that the amplification was indeed associated with increased protein levels (Fig. 3H). As an additional validation, we performed shRNA knockdown of *RRM1* to reduce protein expression (Fig. 3I), followed by a dose–response assay comparing EC₅₀ values of both wildtype HAP1 and *RRM1* knockdown lines, which showed that downregulation of *RRM1* made HAP1 cells 31-fold more sensitive to GEM than their isogenic parent (Fig. 3J,K). As expected *RRM1* downregulation had no effect on HAP1 sensitivity to other drugs (Fig. S5).

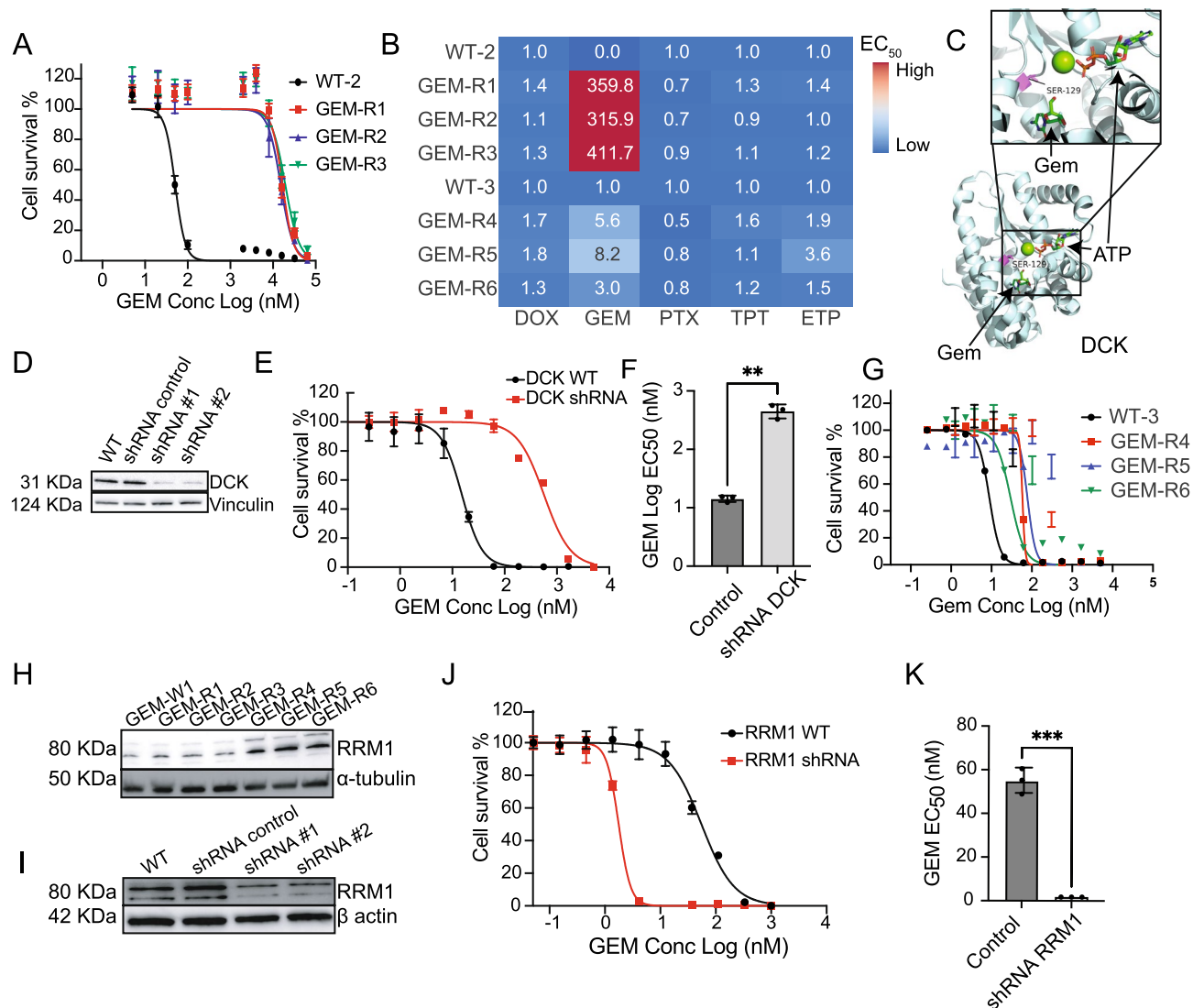


Figure 3. (A) GEM EC₅₀ curves for first set of GEM evolved lines using n=8 technical replicates per concentration point (B). EC₅₀ ratio matrix showing absence of multidrug resistance pathways in all GEM resistant lines. (C) Crystal structure of human DCK (1P62) co-crystallized with GEM showing the position of Ser129, ~11 Å from GEM. D. Western blot confirming that shRNA gene depletion downregulates protein levels for DCK. shRNA #1 and shRNA #2 indicate independent biological replicates with same pool. (E) EC₅₀ 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 EC₅₀ WT and shRNA knockdown cell line for DCK in presence of GEM. Significance (**p<0.05) determined by a paired ratio t test for three biological replicates. (G) GEM EC₅₀ curves for second set of GEM evolved lines. (H) 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. I. Western blot confirming that shRNA gene depletion downregulates protein levels for RRM1. shRNA #1 and shRNA #2 indicate independent biological replicates with same pool. β -actin is used as a loading control. (J) EC₅₀ curves of the WT and shRNA knockdown cell lines for RRM1. (K) Barplot of the EC₅₀ control and shRNA RRM1 knockdowns in presence of GEM. Significance (***p<0.01) determined by a paired ratio t test for three biological replicates.

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, ~10× to PTX-R4, R5, R6, 50X) (Table S2). The first four (Fig. 4A) were subjected to WGS and the latter three to WES (Fig. 4B). 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 prostaglandin⁴⁹. Gene knockdown experiments showed that clones with loss of *SLCO3A1* (Fig. 4C) resulted in HAP1 cells that were ~8 times more resistant than their isogenic parents to PTX (Fig. 4D,E).

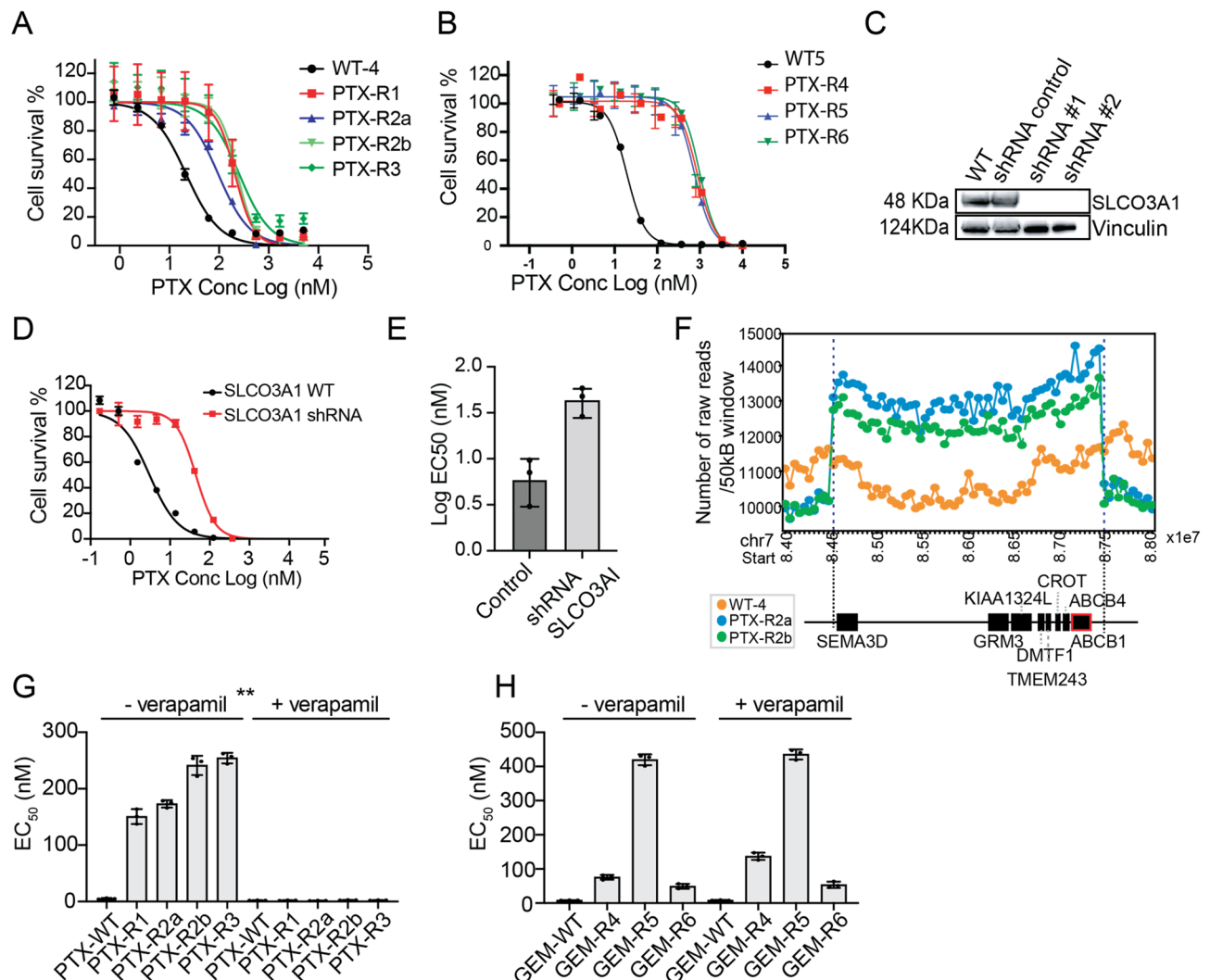


Figure 4. (A) PTX EC_{50} curves for evolved lines with $n=8$ technical replicates (values are in Table S2) (B) PTX EC_{50} curves for a second independent set of evolved lines with $n=8$ technical replicates (C) Western blot confirming shRNAs downregulate protein levels for *SLCO3A1*. shRNA #1 and shRNA #2 indicate independent biological replicates with same pool. (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* for three independent biological replicates. (F) Raw copy number profile for the amplification event containing protein coding genes including ABC transporters (*ABCB1/ABCB4*) PTX cell lines. The amplification region (chr7:84500000–87300000) had a higher number of raw reads (labeled with blue dash lines) with default window size of 50 K 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. (G) Barplot of EC_{50} of the PTX treated cell lines with and without verapamil and verapamil alone showing sensitization in presence of verapamil an ABC inhibitor ($n=4$ technical replicates). (H) Control barplot of EC_{50} of the GEM cell lines \pm verapamil showing no EC_{50} shift for GEM cell lines when co-treated with verapamil. Unless otherwise noted, all data is represented by mean \pm s.e.m. with $n=3$ with individual biological replicates overlaid. ** p value < 0.01 . p values determined by two-tailed t test.

Despite the lack of obvious 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. 4F) on chromosome 7 (PTX-R2a, R2b). *ABCB1* amplifications are associated with clinical resistance to PTX⁵⁰. 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 ABC transporters in PTX resistance, clones were treated with both PTX and verapamil, a calcium channel-blocker which can reverse ABC-transporter mediated resistance^{51,52}. We observed a complete reversal of resistance in PTX lines (Fig. 4G). In contrast, we observed no reversal of resistance in GEM lines (Fig. 4H), suggesting the resistance role of ABC-transporters is PTX-specific.

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 ninefold increased resistance respectively (Fig. S6A, 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)⁵³ involved in cell cycle progression, signal transduction, apoptosis and gene regulation. Members of the CPSF complex have appeared repeatedly in *in vitro* evolution experiments in parasites (both in *Toxoplasma gondii* and *P. falciparum*) and have been validated as resistance mechanisms^{54,55}. While CPSF subunits may represent targets of specific drugs, point mutations in members of the complex may also have a role in overcoming cellular stress that blocks cellular proliferation. Disruption of *WDR33* can lead to slowed DNA replication forks⁵⁶, 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 EC₅₀ value 3.4 times greater than its parental line or the scrambled control (Fig. S6B–D; Table 2), despite an incomplete disruption of the gene by shRNA silencing. ETP-R3 also has a Cys56Phe WNT3A mutation that is found in a highly conserved cysteine that participates in disulfide bond formation and whose disruption abolishes WNT3A activity in recombinant assays⁵⁷. WNT3A has been linked to Etoposide resistance as well as resistance to multiple chemotherapies⁵⁸ (reviewed in Ref.⁵⁹).

No clear candidate SNVs were evident for ETP-R1 and ETP-R2, which did not carry the *WDR33* or *WNT3A* mutations (Table S5). We also noted that all the three ETP resistant lines were cross-resistant to PTX, and DOX, slightly cross-resistant with TPT (Fig. S6E,F) but showed no sensitivity to GEM (Fig. S6G). Despite the cross-resistance, we did not observe any overrepresented genes with missense mutations or shared CNVs that were found in the ETP, PTX and DOX lines but not GEM (Table S7). Because automated analysis of WES could miss smaller CNVs and because *ABCB1* has been linked to both ETP, DOX and PTX resistance⁶⁰ but not GEM, we examined the *ABCB1* region manually using IGV. These data showed a large increase in read coverage for ETP-R1 and ETP-R2 relative to WT-3 parent line (Fig. S6H) at the same chromosome 7 *ABCB1* region amplified in the PTX lines (Fig. 4F). As a final step, we tested PTX lines (PTX-R1, R2a, R2b and R3) that contain amplifications of *ABCB1*, and they were indeed cross-resistant to ETP (Fig. S6I). These data show the limitations of WES data and current automated algorithms. Deciphering the contributions of amplified genes in the absence of a rich literature may require other technologies such as genome-wide knockdown or overexpression libraries.

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 to 20×; Table S2) and all six clones were subjected to WGS together with their parent clones (WT-6 and WT-7).

For TPT-R4a–c lines (Fig. 5A), 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. 5B, Fig. S7A), the known target of topotecan²³. The His81 frameshift mutation, which introduces a premature stop codon, was confirmed by examining the read alignments (Fig. S7A) and by the absence of the full-length protein using N-terminal antibodies (Fig. 5C). Because there were also complex structural variants in the region (Fig. 5D, Fig. S7B) we also sequenced the 5' cDNA through the His81 frameshift for all three lines and as well as the parent line and confirmed the two-base deletion in the mutant as well as homozygosity in TPT-R4a–c evolved lines. We also observed a decrease in mRNA expression with TPT-R4a–b showing a statistically significant decrease in *TOP1* mRNA expression, relative to TPT-WT (Fig. S8). It has been previously shown that a targeted RNAi suppression of *Top1* produces resistance to camptothecin, a close analog of topotecan³⁷. 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 relevance⁶¹. The probability of this distribution by chance is 9.65×10^{-5} .

No clear coding SNVs with a high allele frequency were obvious in TPT-R1, R2 and R3 (Fig. 5E) but we noted multiple SNVs (Asp217Glu from TPT-R4a, b, c and Val432Leu from TPT-R1) in *CYP1B1*, which encodes a cytochrome p450 isoform. Overexpression of *CYP1B1* has previously been associated with TPT resistance⁵¹. TPT resistant lines (TPT-R1, R2 and R3) also showed large chromosomal abnormalities at *WWOX* (Fig. 5F) with a clear deletion of the *WWOX* gene region (chr16:78569166–78792736, exon7 and 8). *WWOX* bears a well-known fragile site (FRA16D) and encodes a putative oxidoreductase. The complete absence of *WWOX* protein was confirmed by Western in TPT-R1, 2 and 3 (Fig. 5G). Interestingly, lower levels of *WWOX* were also observed in TPT-R4a–c, which could be a consequence of other *cis* or *trans* variants in this cell line and might also contribute to this level of resistance. Knockdown of *WWOX* by shRNA resulted in marked resistance to TPT (Fig. 5H,I). *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 TPT⁶². *WWOX* disruption also resulted in resistance to PTX (Fig. 5J), and as reported by others who examined *WWOX*-transfected epithelial ovarian cancer cells⁶³.

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. This method is designed to identify the targets of drugs that block cellular proliferation and could be used on any compound that has an antiproliferative effect, including chemotherapies such as asparaginase. The method could also theoretically be used with biologics such as monoclonal antibodies or antisense oligonucleotides. In drug development, it could be used to demonstrate on-target activity. It would not be as useful on compounds that only change cellular signaling or those that alter

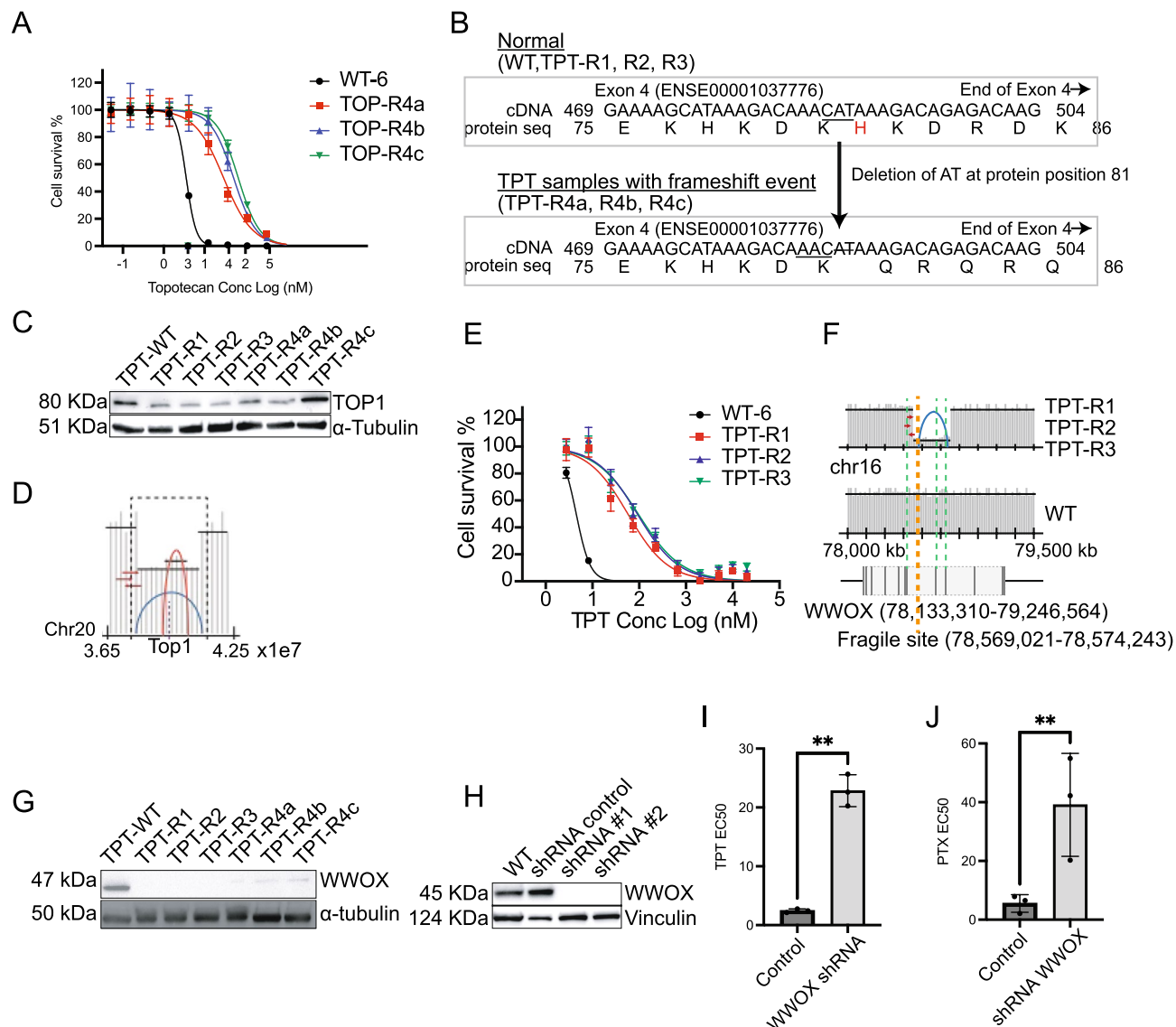


Figure 5. (A) TPT EC₅₀ curves for TOP-R4a,b and c. (B) Transcript sequence (cDNA) and protein sequence for TOP1 transcript for normal (top) and TPT 4a, b and c. The figure shows only part of the cDNA (position 469–504) and protein sequence (position 75–86) for TOP1 at the affected exon (Exon 4, ENSE00001037776). 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. (C) Western blot TOP1 protein depletion in evolved lines. (D) Schematic showing complex read depth patterns around *TOP1*. (E) TPT EC₅₀ curves for evolved TPT-R1, R2 and R3. (F) Schematic of chr16 reads around *WWOX* for TPT-R1, R2, and R3 compared to the WT chromosome 16 parental cell line. Blue arch represents a deleted region. *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). (G) Western blot showing *WWOX* protein levels in TPT resistant clones. (H) Western blot shows downregulation of protein levels for *WWOX* in shRNA samples compared to WT and scrambled control. shRNA #1 and shRNA #2 indicate independent biological replicates with same pool. Barplot of EC₅₀ control (wt) and *WWOX* shRNA knockdowns for TPT (I) and PTX (J). EC₅₀ data is represented by mean \pm s.e.m. with $n=3$ biological replicates and 4–8 technical replicates. ** = $p < 0.05$. p values determined by ratio, paired two-tailed t test.

the immune function, like prednisolone. Our results show in vitro resistance acquisition and provide a framework for the determination of chemotherapy resistance alleles that may arise in patients.

Our work using IVIEWGA in pathogens (see Ref.⁶⁴ for a review) guided our pipeline development: We focused on protein coding alterations that arose in association with a single treatment condition, that were nonsynonymous, occurring repeatedly and were high allele frequency. We also removed alleles for genes that are known to mutate frequently, like odorant receptors. 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.

Because of the substantially greater costs associated with WGS, here we did evaluate both WES and WGS sequencing methods. Despite a higher likelihood of discovering all changes by WGS, the disadvantage of WGS is cost and computational time. While human WES data can be analyzed on a laptop, human WGS data files are large and difficult to handle, computationally. One potential strategy is to perform low pass WGS for CNV detection and WES for SNV detection on all samples.

The biggest disadvantage of using WES is that CNVs will be harder to call. This is partly for statistical reasons with many reads that support CNV calls located outside of coding regions in WGS samples. In addition, if one sequences over the exact location of the recombination event (or the start or end of the deletion) one can obtain additional support for location calls via split read analysis of paired-end libraries. In addition, one can extract the sequence of the short read and reconstruct the exact recombination breakpoint, as shown in Fig. S3. This would not be feasible with whole exome sequencing. Recently it was shown that CNV detection tools perform poorly on WES cancer genome samples. Comparative analysis showed a low consensus in CNV calling tools with moderate sensitivity (~50% to ~80%), fair specificity (~70% to ~94%) and poor FDRs (~27% to ~60%). Also, using simulated data these authors observed that increasing the coverage more than 10× in exonic regions did not improve the detection power⁶⁵. Of course, detecting CNVs is likely to be more challenging in diploid genomes, than haploid genomes. In support of this, we were able to identify and validate the RRM1 amplification event in GEM-R4, 5 and 6, which were only subjected to WES. In addition, in yeast, it appears CNVs are much less important than SNVs in driving drug resistance as well: in a more comprehensive *in vitro* evolution study in yeast⁶⁶ with 80 different and 355 whole genome sequences we observed only 24 CNVs, including apparent aneuploidy (11 times, occurring in 10 clones) and small, intrachromosomal amplifications (13 times, occurring in 13 clones) in our set of 355 whole genome sequences⁶⁷.

A lesser disadvantage of WES is the rare possibility that resistance is conferred by an intergenic mutation, which would be missed by WES data. GWAS studies have frequently identified noncoding variants as important for phenotypes in the human population. It is likely that most of the mutations that are evolved here would have a negative impact on fitness and would not be tolerated in the germline and thus would not appear in human GWAS studies⁶⁸. Our work in other organisms has shown that almost all resistance conferring SNVs or small indels are nonsynonymous changes that would be detected by both WES and WGS. In our comprehensive yeast study⁶⁶, 271 mutations of the 1405 detected mutations in the 355 evolved lines were intergenic. Of these, only five were directly upstream or downstream of one of the 137 genes that were repeatedly identified in the study. In contrast to coding mutations, most intergenic mutations lacked any statistical support suggesting relevance and were likely to be background mutations⁶⁷. Despite the lower probability that intergenic or other noncoding mutations may have functional effect, we recognize that there are examples from the literature where intergenic mutations have contributed to drug resistance. Non-coding RNAs such as *EGRF-AS1* and activating cis elements such as enhancers have previously been implicated in evasion of drug response^{69–71}. 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. In addition, humans clearly have much more complex splicing patterns than microbes and indeed we found evidence for multiple intronic (and intergenic mutations) in genes with higher-than-expected rates of disruptive mutations and some level of validation (e.g. *SLCO31A*, *TOP1*, *DCK*) contributing to resistance in our dataset.

A limitation of our HAP1 study, as presented, and in contrast to our work in other species, is that despite some level of repetition, we seldom achieved strong statistical confidence by just performing selections and sequencing. This may not be unexpected. Evolution is, unfortunately, a relatively stochastic process even when working with the exact same starting clone. In the yeast study⁶⁶ we only obtained the same allelic change in the same critical drug resistance gene a few times despite > 3 repetitions per each of the 80 compounds. For example, two independent selections with hectochlorin both yielded an Arg116Lys in ACT1, the target of hectochlorin⁶⁷. Similarly, a Leu671Phe change in YRM1 was observed 5 times for 4 different compounds.

Another disadvantage of using human cells is the challenge of validation of SNVs; we were not able to engineer any SNVs into HAP1 cells to demonstrate their importance. On the other hand, with the statistical confidence that comes from identifying the same gene repeatedly, CRISPR-Cas9 validation becomes less important. In the same yeast study described above, YRM1, a gene encoding a transcription factor involved in drug resistance in yeast was independently identified 52 times with 27 different alleles⁶⁶. The likelihood of 355 selections yielding the same gene by chance is roughly 3.53×10^{-116} . This enrichment analysis becomes an attractive method for teasing apart driver and passenger mutations and may become possible with more repetitions despite the larger genome size of HAP1 cells. However, performing enough repetitions to achieve statistical confidence would require substantial resources with WGS, even with a thousand-dollar human genome. WES is thus likely to be more useful.

While HAP1 cells may not be considered a perfect model for human cancer biology, for the purposes of target identification, they are likely very useful. As with pathogens, our use of well-studied drugs, largely uncovered genes that were mostly already well known to confer resistance such as RRM1^{46,72}, DCK^{73,74}, TOP2A³⁷, and TOP1³⁷ in a variety of different cancer cell lines. Although it was initially argued that the *in vitro* evolution system might be artificial, in malaria parasites it has been used to discover or rediscover most, if not all (to our knowledge), clinically relevant drug resistance genes including the chloroquine resistance transporter⁵, the artemisinin resistance gene, *Pfkelch13*⁷⁵, and well-known ABC transporters⁵.

Despite questions about how much they mimic human cells, the value of using haploid cell lines is evident from our allele frequency data. If our lines had been diploid, we would have needed to consider allele frequency data of up to 0.4. There are 205 missense mutations with an AF of > 0.4, making pinpointing the causative allele much more difficult without candidate genes or without many repetitions. Although *in vitro* evolution has been used repeatedly for discovering the mechanism of action of completely uncharacterized compounds in malaria parasites (reviewed in Ref.⁶⁴), there are fewer examples of *in vitro* evolution being used for *de novo*

target discovery in diploid eukaryotic pathogens. Although there are some examples in trypanosomes^{76–78}, some hypothesis about the mechanism of action was already present before evolution studies were attempted. Despite this, low allele frequency data should not necessarily be discarded. There have been multiple examples from haploid *Plasmodium* where a resistance-conferring allele was located within an amplification region and thus showed an $AF < 0.5$. It is possible that the incomplete penetrance that we observed here with respect to AF is due to duplication events that were not readily detected. This may be particularly true for *WNT3A* ($AF = 0.571$) which lies in the chromosome 1 amplified region.

Although the HAP1 cells could be considered unnatural, it is likely that similar evolution experiments in other types of human cells will largely give the same genes. This is because conservation of drug targets and drug resistance mechanisms across phyla is often observed, although a given compound or inhibitor may show differences in selectivity and specificity. Resistance to topotecan/camptothecin in yeast is also provided by mutations in *Top1*⁷⁹. Our IVIEWGA studies in yeast also identified *Top1* as the target in yeast⁶⁶. Evolution studies with cladosporin in yeast and plasmodium both give the same resistance mechanism for cladosporin, lysyl tRNA synthetase⁸⁰.

Our studies were not meant to study the process of evolution. Within the field of laboratory-based evolution, there are two broad areas of study. The first are those that fall under the heading of “experimental evolution” and which try to mimic evolution in natural conditions. Here, growth rates are often recorded, and experimental conditions may be varied in a controlled manner (carbon sources, temperature, etc.). Such studies include long term studies of *E. coli* or other bacteria (reviewed in Ref.⁸¹) and have also been performed with small molecules^{82,83}, primarily with known mechanism of action. Alternatively, there are also studies in which evolution has been used as a tool to discover targets and resistance genes for therapeutic purposes^{7,64,84}. In many cases^{84,85}, although not in all cases the term “in vitro evolution” is used instead of “experimental evolution.” Based on our results here, resistance readily emerges in HAP1 cells but more work will need to be done to determine if this is because of the compounds that were used. Here we used in vitro evolution (versus experimental evolution) to select for mutant lines that could withstand treatment with the selected drugs. Although it may be possible to use HAP1 cells for experimental evolution, at present sequencing costs are so high that whole genome studies with whole genome analysis are not practical but this may change in the future. Questions that might be investigated include the fitness of different mutations, reproducibility of the process, impact of the starting clone, carbon sources or growth rate and whether one resistance mechanism predominates or if a variety are found.

Finally, it is important to keep in mind that the compounds examined here are not modern cancer therapies and while still used clinically, they are imperfect. Newer molecules include bortezomib, a small molecule proteasome inhibitor, imatinib, a small molecule tyrosine kinase inhibitor or seliciclib, small molecule cyclin-dependent kinase inhibitor or even small molecule cancer immunotherapies. A limitation of the cell line that was used here is that it is already resistant to some of these modern drugs, including imatinib (Fig. S1), which could necessitate the use/development of other haploid lines, or potentially the use of diploid lines, although data analysis is expected to be more difficult. We anticipate mutations in the drug's targets will be identified sometimes, as is observed in microbes. In fact, unbiased IVIEWGA studies with bortezomib in *P. falciparum* have demonstrated mutations in the proteasome subunit, $Pf20S\ \beta5^{86}$, confer resistance, and similar resistance-conferring mutations have been discovered after using in vitro evolution in human cells, although whole genome sequencing was not performed and the mutations were identified using a candidate gene approach⁸⁷. On the other hand, some targeted therapies may not work well against HAP1 cells because they do not bear the sensitizing mutation (e.g. the *BRAF/EGFR* mutations for vemurafenib, gefitinib or erlotinib, respectively⁸⁸). Alternatively, the HAP1 cells may be intrinsically resistant because they harbor other resistance conferring mutations, which may be the case for imatinib which targets the BCR-Abl fusion protein encoded by HAP1 cells. Nevertheless, if they do show sensitivity or can be engineered to sensitivity HAP1 cells may prove useful for predicting resistance mechanisms for new drugs in clinical development, for determining on-target activity and for studying the many other drugs that were not included in our small screening set.

Materials and methods

Compounds

All chemotherapeutic agents used in this study were obtained from Sigma-Aldrich, dissolved in DMSO at 10 mM concentration and stored at $-20\ ^\circ\text{C}$.

Cell cultures

The human chronic myelogenous leukemia cell line, HAP1, 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 $37\ ^\circ\text{C}$ 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), 0.29 mg/mL L-glutamine, 25 mM HEPES, 100U/mL Penicillin and 100 $\mu\text{g}/\text{mL}$ Streptomycin (1% penicillin/streptomycin). Monoclonal and polyclonal stocks were made and stored in IMDM + 10% DMSO in liquid nitrogen.

In vitro evolution of resistant HAP1 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 HAP1 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 stepwise selection method. For high pressure selection, cells were treated at a concentration $3\text{--}10 \times \text{EC}_{50}$ 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 stepwise 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: 24 h prior to addition of the drugs, 2×10^4 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 10 μ M 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 (10 μ M) was added to every concentration of the drug. After a 48-h 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 HAP1 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.). Unless otherwise noted, dose response experiments consisted of 4–8 technical replicates and 3 biological replicates.

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 ($> 2.0 \mu$ g, $> 50 \text{ ng}/\mu\text{L}$, $> 20 \mu\text{L}$) 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 Bio-analyzer 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/Pipeline/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)⁸⁹. 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 S9 shows the raw annotation from snpEff and consolidated classification used in Table 1; Table S5 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 S9).

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)⁹⁰ to ensure coverage and allele fractions of the mutation positions. The top genes for each drug were included in Table 2 and Table S7.

Somatic copy number variations (CNVs) analysis

Copy number regions for WGS and WES were called by ControlFreeC⁴⁷ 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 pools 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), *WVOX* (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). HAP1 cells were plated at 120,000 cells per well (~40% confluency) in a 24-well plate, 24 h prior to viral transduction. On the day of transduction, complete media was replaced with serum-free media and 7 µg/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 h more and finally stable clones expressing the shRNA were selected using complete media with 2 µg/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 lenti-virus 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. HAP1 cells were plated at 120,000 cells per well (~40% confluency) in a 24-well plate 24 h prior to viral transduction. On the day of transduction, complete media was replaced with serum-free media and 7 µg/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 h more and stable clones expressing the CRISPR-Cas9 sequence were selected using complete media with 2 µg/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 #P4707-50 ml) to improve cell adhesion) in presence of 2 µg/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 HAP1 cell lines first via PCR and then reconfirmed by western blot using the *USP47* rabbit polyclonal antibody (Abcam, Cat# ab97835).

Immunoblotting

HAP1 cells (at least 5×10^6) were trypsinized, washed twice with cold 1 × DPBS and then lysed in 500 µL Pierce[™] RIPA Buffer (Thermo Scientific) containing 1:100 protease inhibitor (Halt[™] Protease & Phosphatase Inhibitor Cocktail, Thermo Scientific) and 1:100 0.5 M 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 1 h 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. S9.

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), *WVOX* (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 (1×10^6 cells) were dissociated from plates by the addition of 2 mL 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 1 µg 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: CGAATCATGCCCCGAGGATATAA; REV: CCAGGAAAC CAGCCAAGTAA, following the manufacturer's instruction.

Data availability

All whole genome sequences have been deposited in SRA BioProject PRJNA603390.

Received: 23 February 2024; Accepted: 3 June 2024

Published online: 18 June 2024

References

- Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)* **343**, 84–87. <https://doi.org/10.1126/science.1247005> (2014).
- Kanarek, N. *et al.* Histidine catabolism is a major determinant of methotrexate sensitivity. *Nature* **559**, 632–636. <https://doi.org/10.1038/s41586-018-0316-7> (2018).
- Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science (New York, N.Y.)* **343**, 80–84. <https://doi.org/10.1126/science.1246981> (2014).
- Antonova-Koch, Y. *et al.* Open-source discovery of chemical leads for next-generation chemoprotective antimalarials. *Science (New York, N.Y.)* **359**, 191–199. <https://doi.org/10.1126/science.aan4472> (2018).
- Cowell, A. N. *et al.* Mapping the malaria parasite druggable genome by using in vitro evolution and chemogenomics. *Science (New York, N.Y.)* **359**, 191–199. <https://doi.org/10.1126/science.aan4472> (2018).
- Ottillie, S. *et al.* Rapid Chagas disease drug target discovery using directed evolution in drug-sensitive yeast. *ACS Chem. Biol.* **12**, 422–434. <https://doi.org/10.1021/acscchembio.6b01037> (2017).
- Andries, K. *et al.* A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science (New York, N.Y.)* **307**, 223–227. <https://doi.org/10.1126/science.1106753> (2005).
- 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*. <https://doi.org/10.7554/eLife.47612> (2019).
- Burckstummer, T. *et al.* A reversible gene trap collection empowers haploid genetics in human cells. *Nat. Methods* **10**, 965–971. <https://doi.org/10.1038/nmeth.2609> (2013).
- Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–343. <https://doi.org/10.1038/nature10348> (2011).
- Gerhards, N. M. *et al.* Haploid genetic screens identify genetic vulnerabilities to microtubule-targeting agents. *Mol. Oncol.* **12**, 953–971. <https://doi.org/10.1002/1878-0261.12307> (2018).
- Findlay, G. M. *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222. <https://doi.org/10.1038/s41586-018-0461-z> (2018).
- Behnisch-Cornwell, S., Wolff, L. & Bednarski, P. J. The effect of glutathione peroxidase-1 knockout on anticancer drug sensitivities and reactive oxygen species in haploid HAP-1 cells. *Antioxidants (Basel, Switzerland)*. <https://doi.org/10.3390/antiox9121300> (2020).
- Smits, A. H. *et al.* Biological plasticity rescues target activity in CRISPR knock outs. *Nat. Methods* **16**, 1087–1093. <https://doi.org/10.1038/s41592-019-0614-5> (2019).
- Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065. <https://doi.org/10.1101/gr.177220.114> (2014).
- Mandell, J. B. *et al.* (2019) Combination therapy with disulfiram, copper, and doxorubicin for osteosarcoma: In vitro support for a novel drug repurposing strategy. *Sarcoma*. **2019**, 1320201. <https://doi.org/10.1155/2019/1320201>.
- Jiang, C. *et al.* Pyrroloquinoline quinine ameliorates doxorubicin-induced autophagy-dependent apoptosis via lysosomal-mitochondrial axis in vascular endothelial cells. *Toxicology* **425**, 152238. <https://doi.org/10.1016/j.tox.2019.152238> (2019).
- 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. <https://doi.org/10.1038/s41467-019-09917-0> (2019).
- Alvarellos, M. L. *et al.* PharmGKB summary: Gemcitabine pathway. *Pharmacogenet. Genom.* **24**, 564–574. <https://doi.org/10.1097/FPC.000000000000086> (2014).
- Mini, E., Nobili, S., Caciagli, B., Landini, I. & Mazzei, T. Cellular pharmacology of gemcitabine. *Ann. Oncol.* **17**(Suppl 5), v7-12. <https://doi.org/10.1093/annonc/mdj941> (2006).
- Pommier, Y., Leo, E., Zhang, H. & Marchand, C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.* **17**, 421–433. <https://doi.org/10.1016/j.chembiol.2010.04.012> (2010).
- Hayashi, M. T. & Karlseder, J. DNA damage associated with mitosis and cytokinesis failure. *Oncogene* **32**, 4593–4601. <https://doi.org/10.1038/ncr.2012.615> (2013).
- Rasheed, Z. A. & Rubin, E. H. Mechanisms of resistance to topoisomerase I-targeting drugs. *Oncogene* **22**, 7296–7304. <https://doi.org/10.1038/sj.onc.1206935> (2003).
- Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–961. <https://doi.org/10.1093/nar/gks1111> (2013).
- Iorio, F. *et al.* A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754. <https://doi.org/10.1016/j.cell.2016.06.017> (2016).
- Corey, V. C. *et al.* A broad analysis of resistance development in the malaria parasite. *Nat. Commun.* **7**, 11901. <https://doi.org/10.1038/ncomms11901> (2016).
- 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. <https://doi.org/10.1186/s12885-019-5939-z> (2019).
- Amrutkar, M. & Gladhaug, I. P. Pancreatic cancer chemoresistance to gemcitabine. *Cancers (Basel)*. <https://doi.org/10.3390/cancers9110157> (2017).
- Pan, P., Li, Y., Yu, H., Sun, H. & Hou, T. Molecular principle of topotecan resistance by topoisomerase I mutations through molecular modeling approaches. *J. Chem. Inf. Model* **53**, 997–1006. <https://doi.org/10.1021/ci400066x> (2013).
- 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).
- Brammell, J. S. *et al.* Genome-wide chemical mutagenesis screens allow unbiased saturation of the cancer genome and identification of drug resistance mutations. *Genome Res.* **27**, 613–625. <https://doi.org/10.1101/gr.213546.116> (2017).

32. Boeva, V. *et al.* Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425. <https://doi.org/10.1093/bioinformatics/btr670> (2012).
33. 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.* **1126**–1133, 2018. <https://doi.org/10.1016/j.bbagen.2018.02.002> (1862).
34. 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. <https://doi.org/10.1186/s13045-016-0304-z> (2016).
35. Ghisoni, E. *et al.* TOP2A as marker of response to pegylated liposomal doxorubicin (PLD) in epithelial ovarian cancers. *J. Ovarian Res.* **12**, 17. <https://doi.org/10.1186/s13048-019-0492-6> (2019).
36. 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. <https://doi.org/10.1016/j.jmb.2012.07.014> (2012).
37. Burgess, D. J. *et al.* Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* **105**, 9053–9058. <https://doi.org/10.1073/pnas.0803513105> (2008).
38. Hsiung, Y. *et al.* Functional expression of human topoisomerase II alpha in yeast: Mutations at amino acids 450 or 803 of topoisomerase II alpha result in enzymes that can confer resistance to anti-topoisomerase II agents. *Cancer Res.* **56**, 91–99 (1996).
39. Lim, M. Y. *et al.* UDP-galactose and acetyl-CoA transporters as Plasmodium multidrug resistance genes. *Nat. Microbiol.* **1**, 16166. <https://doi.org/10.1038/nmicrobiol.2016.166> (2016).
40. Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S. & Baradaran, B. (2017) The Different mechanisms of cancer drug resistance: A brief review. *Adv. Pharm. Bull.* **7**, 339–348. <https://doi.org/10.1517/apb.2017.041>.
41. Almontashiri, N. A. *et al.* SPG7 variant escapes phosphorylation-regulated processing by AFG3L2, elevates mitochondrial ROS, and is associated with multiple clinical phenotypes. *Cell Rep.* **7**, 834–847. <https://doi.org/10.1016/j.celrep.2014.03.051> (2014).
42. Zubovych, I. O., Straud, S. & Roth, M. G. Mitochondrial dysfunction confers resistance to multiple drugs in *Caenorhabditis elegans*. *Mol. Biol. Cell* **21**, 956–969. <https://doi.org/10.1091/mbc.e09-08-0673> (2010).
43. Deeken, J. F. *et al.* A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform. *Pharmacogenomics J* **10**, 191–199. <https://doi.org/10.1038/tpj.2009.57> (2010).
44. Sabini, E., Ort, S., Monnerjahn, C., Konrad, M. & Lavie, A. Structure of human dCK suggests strategies to improve anticancer and antiviral therapy. *Nat. Struct. Biol.* **10**, 513–519. <https://doi.org/10.1038/nsb942> (2003).
45. Mandt, R. E. K. *et al.* In vitro selection predicts malaria parasite resistance to dihydroorotate dehydrogenase inhibitors in a mouse infection model. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aav1636> (2019).
46. Beppler, 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. <https://doi.org/10.1200/JCO.2006.06.1101> (2006).
47. 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. *Cell. Oncol. (Dordrecht)* **38**, 319–325. <https://doi.org/10.1007/s13402-015-0225-9> (2015).
48. 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 Urol.* **19**, 69. <https://doi.org/10.1186/s12894-019-0497-x> (2019).
49. Thakkar, N., Lockhart, A. C. & Lee, W. Role of organic anion-transporting polypeptides (OATPs) in cancer therapy. *AAPS J.* **17**, 535–545. <https://doi.org/10.1208/s12248-015-9740-x> (2015).
50. 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. <https://doi.org/10.1038/bjc.2016.203> (2016).
51. 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. <https://doi.org/10.1007/s10456-014-9422-9> (2014).
52. Wu, C. P., Calcagno, A. M. & Ambudkar, S. V. Reversal of ABC drug transporter-mediated multidrug resistance in cancer cells: Evaluation of current strategies. *Curr. Mol. Pharmacol.* **1**, 93–105. <https://doi.org/10.2174/1874467210801020093> (2008).
53. Chan, S. L. *et al.* CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev.* **28**, 2370–2380. <https://doi.org/10.1101/gad.250993.114> (2014).
54. Sonoiki, E. *et al.* A potent antimalarial benzoxaborole targets a Plasmodium falciparum cleavage and polyadenylation specificity factor homologue. *Nat. Commun.* **8**, 14574. <https://doi.org/10.1038/ncomms14574> (2017).
55. Palencia, A. *et al.* (2017) Targeting Toxoplasma gondii CPSF3 as a new approach to control toxoplasmosis. *EMBO Mol. Med.* **9**, 385–394. <https://doi.org/10.15252/emmm.201607370>.
56. Teloni, F. *et al.* Efficient pre-mRNA cleavage prevents replication-stress-associated genome instability. *Mol. Cell* **73**, 670–683.e612. <https://doi.org/10.1016/j.molcel.2018.11.036> (2019).
57. MacDonald, B. T. *et al.* Disulfide bond requirements for active Wnt ligands. *J. Biol. Chem.* **289**, 18122–18136. <https://doi.org/10.1074/jbc.M114.575027> (2014).
58. Thiago, L. S. *et al.* The Wnt signaling pathway regulates Nalm-16 b-cell precursor acute lymphoblastic leukemic cell line survival and etoposide resistance. *Biomed. Pharmacother.* **64**, 63–72. <https://doi.org/10.1016/j.biopha.2009.09.005> (2010).
59. Zhong, Z. & Virshup, D. M. Wnt signaling and drug resistance in cancer. *Mol. Pharmacol.* **97**, 72–89. <https://doi.org/10.1124/mol.119.117978> (2020).
60. Skinner, K. T., Palkar, A. M. & Hong, A. L. Genetics of ABCB1 in cancer. *Cancers (Basel)*. <https://doi.org/10.3390/cancers15174236> (2023).
61. Forbes, S. A. *et al.* COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–d783. <https://doi.org/10.1093/nar/gkw1121> (2017).
62. Del Mare, S., Salah, Z. & Aqeilan, R. I. WWOX: Its genomics, partners, and functions. *J. Cell. Biochem.* **108**, 737–745. <https://doi.org/10.1002/jcb.22298> (2009).
63. Janczar, S. *et al.* WWOX sensitises ovarian cancer cells to paclitaxel via modulation of the ER stress response. *Cell Death Disease* **8**, e2955. <https://doi.org/10.1038/cddis.2017.346> (2017).
64. Luth, M. R., Gupta, P., Otilie, S. & Winzeler, E. A. Using in vitro evolution and whole genome analysis to discover next generation targets for antimalarial drug discovery. *ACS Infect. Diseases* **4**, 301–314. <https://doi.org/10.1021/acsinfecdis.7b00276> (2018).
65. Zare, F., Dow, M., Monteleone, N., Hosny, A. & Nabavi, S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC Bioinform.* **18**, 286. <https://doi.org/10.1186/s12859-017-1705-x> (2017).
66. Otilie, S. *et al.* (2022) Adaptive laboratory evolution in *S. cerevisiae* highlights role of transcription factors in fungal xenobiotic resistance. *Commun. Biol.* **5**, 128. <https://doi.org/10.1038/s42003-022-03076-7>.
67. Otilie, S. *et al.* Defining the yeast resistome through in vitro evolution and whole genome sequencing. *bioRxiv*, 2021.2002.2017.430112. <https://doi.org/10.1101/2021.02.17.430112> (2021).
68. Tak, Y. G. & Farnham, P. J. Making sense of GWAS: Using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenet. Chromatin* **8**, 57. <https://doi.org/10.1186/s13072-015-0050-4> (2015).
69. Si, W., Shen, J., Zheng, H. & Fan, W. The role and mechanisms of action of microRNAs in cancer drug resistance. *Clin. Epigenet.* **11**, 25. <https://doi.org/10.1186/s13148-018-0587-8> (2019).
70. 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. <https://doi.org/10.1111/cas.13642> (2018).

71. Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat. Rev. Cancer* **16**, 483–493. <https://doi.org/10.1038/nrc.2016.62> (2016).
72. Minami, K. *et al.* Ribonucleotide reductase is an effective target to overcome gemcitabine resistance in gemcitabine-resistant pancreatic cancer cells with dual resistant factors. *J. Pharmacol. Sci.* **127**, 319–325. <https://doi.org/10.1016/j.jphs.2015.01.006> (2015).
73. Saiki, Y. *et al.* DCK is frequently inactivated in acquired gemcitabine-resistant human cancer cells. *Biochem. Biophys. Res. Commun.* **421**, 98–104. <https://doi.org/10.1016/j.bbrc.2012.03.122> (2012).
74. Ohhashi, S. *et al.* Down-regulation of deoxycytidine kinase enhances acquired resistance to gemcitabine in pancreatic cancer. *Anticancer Res.* **28**, 2205–2212 (2008).
75. Ariey, F. *et al.* A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* **505**, 50–55. <https://doi.org/10.1038/nature12876> (2014).
76. Khare, S. *et al.* Utilizing chemical genomics to identify cytochrome b as a novel drug target for chagas disease. *PLoS Pathogens* **11**, e1005058. <https://doi.org/10.1371/journal.ppat.1005058> (2015).
77. Khare, S. *et al.* Proteasome inhibition for treatment of leishmaniasis. Chagas disease and sleeping sickness. *Nature* **537**, 229–233. <https://doi.org/10.1038/nature19339> (2016).
78. Wyllie, S. *et al.* Cyclin-dependent kinase 12 is a drug target for visceral leishmaniasis. *Nature* **560**, 192–197. <https://doi.org/10.1038/s41586-018-0356-z> (2018).
79. Bjornsti, M.-A., Knab, A. M. & Benedetti, P. Yeast *Saccharomyces cerevisiae* as a model system to study the cytotoxic activity of the antitumor drug camptothecin. *Cancer Chemother. Pharmacol.* **34**, S1–S5. <https://doi.org/10.1007/BF00684856> (1994).
80. Hoepfner, D. *et al.* Selective and specific inhibition of the plasmodium falciparum lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. *Cell Host Microbe* **11**, 654–663. <https://doi.org/10.1016/j.chom.2012.04.015> (2012).
81. Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J.* **11**, 2181–2194. <https://doi.org/10.1038/ismej.2017.69> (2017).
82. Maeda, T. *et al.* High-throughput laboratory evolution reveals evolutionary constraints in *Escherichia coli*. *Nat. Commun.* **11**, 5970. <https://doi.org/10.1038/s41467-020-19713-w> (2020).
83. Card, K. J., Thomas, M. D., Graves, J. L. Jr., Barrick, J. E. & Lenski, R. E. Genomic evolution of antibiotic resistance is contingent on genetic background following a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. <https://doi.org/10.1073/pnas.2016886118> (2021).
84. Rock, F. L. *et al.* An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science (New York, N.Y.)* **316**, 1759–1761. <https://doi.org/10.1126/science.1142189> (2007).
85. Xu, Y. *et al.* In vivo evolution of drug-resistant *Mycobacterium tuberculosis* in patients during long-term treatment. *BMC Genom.* **19**, 640. <https://doi.org/10.1186/s12864-018-5010-5> (2018).
86. Xie, S. C. *et al.* Target validation and identification of novel boronate inhibitors of the plasmodium falciparum proteasome. *J. Med. Chem.* **61**, 10053–10066. <https://doi.org/10.1021/acs.jmedchem.8b01161> (2018).
87. Allmeroth, K. *et al.* Bortezomib resistance mutations in PSMB5 determine response to second-generation proteasome inhibitors in multiple myeloma. *Leukemia* **35**, 887–892. <https://doi.org/10.1038/s41375-020-0989-4> (2021).
88. Ellis, L. M. & Hicklin, D. J. Resistance to targeted therapies: Refining anticancer therapy in the era of molecular oncology. *Clin. Cancer Res.* **15**, 7471–7478. <https://doi.org/10.1158/1078-0432.Ccr-09-1070> (2009).
89. 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. <https://doi.org/10.4161/fly.19695> (2012).
90. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26. <https://doi.org/10.1038/nbt.1754> (2011).

Acknowledgements

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, AK performed 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. AK performed sequence analysis. HC provided advice and management and wrote the manuscript. EAW performed data analysis, provided advice, obtained funding, wrote the manuscript and created figures and tables. TI obtained funding and provided advice.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-63943-7>.

Correspondence and requests for materials should be addressed to T.I., H.C. or E.A.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024