Revisiting the Effects of MDR1 Variants Using Computational Approaches

Tal Gutman[1] and Tamir Tuller[1,2(✉)]

[1] Department of Biomedical Engineering, The Engineering Faculty, Tel-Aviv University, 69978 Tel Aviv, Israel
tamirtul@tauex.tau.ac.il
[2] The Sagol School of Neuroscience, Tel-Aviv University, 6997801 Tel Aviv, Israel

Abstract. P-glycoprotein, encoded by the MDR1 gene, is an ATP-dependent pump that exports various substances out of cells. Its overexpression is related to multi drug resistance in many cancers. Numerous studies explored the effects of MDR1 variants on p-glycoprotein expression and function, and on patient survivability. T1236C, T2677G and T3435C are prevalent MDR1 variants that are the most widely studied, typically in-vitro and in-vivo, with remarkably inconsistent results. In this paper we perform computational, data-driven analyses to assess the effects of these variants using a different approach. We use knowledge of gene expression regulation to elucidate the variants' mechanism of action. Results indicate that T1236C is correlated with worse patient prognosis. Additionally, examination of MDR1 folding strength suggests that T3435C potentially modifies co-translational folding. Furthermore, all three variants reside in potential translation bottlenecks and likely cause increased translation rates. These results support several hypotheses suggested by previous studies. To the best of our knowledge, this study is the first to apply a computational approach to examine the effects of MDR1 variants.

Keywords: MDR1 · synonymous variants · gene expression · cancer evolution · mRNA folding selection

1 Introduction

P-glycoprotein (p-gp) was first discovered in 1976, where it was shown to alter drug permeability in Chinese hamster ovary cells [1]. Ensuing extensive studies unveiled ample information about the structure and function of this transmembrane protein; P-gp is a member of the ATP-binding cassette transporter superfamily [2]. It is a 170 KDa molecule that is comprised of two halves with high homology, each containing six transmembrane domains (TMD) and a single nucleotide-binding domain (NBD) [3]. P-gp is interleaved in the cell's membrane and functions as an ATP-dependent efflux pump, exporting diverse substances out of the cell. Lipids, steroids, xenobiotics, various drugs and chemotherapy agents are some of the molecules transferred by this protein [4]. P-gp is normally highly expressed in the adrenal glands, kidneys, liver, the gastro-intestinal

C. Scornavacca and M. Hernández-Rosales (Eds.): RECOMB-CG 2024, LNBI 14616, pp. 226–247, 2024.
https://doi.org/10.1007/978-3-031-58072-7_12

tract and the blood-brain barrier [5]. It protects the body from deleterious substances by excreting them to the gut lumen, urine and bile, and by reducing permeability of sensitive tissues [6]. P-gp is also highly expressed in many cancer cells. Its ability to export substrates that vary greatly in chemical structure and size makes it fundamental for the multi-drug resistance (MDR) mechanism [7] of cancer cells.

P-gp is encoded by the MDR1 gene, also frequently called ABCB1. It is located on chromosome 7 and encodes for a protein of 1,280 amino acids [8]. MDR1 is highly polymorphic, with over 50 single nucleotide polymorphisms (SNP) currently discovered in its coding region [9]. Three variants are the most extensively studied - T1236C (rs1128503), T2677G (rs2032582) and T3435C (rs1045642). T1236C is a synonymous variant that occurs in the first NBD (at position 1236 of the coding sequence), changing the codon GGT to GGC (both encode for Glycine). T2677G is a non-synonymous variant that occurs between the tenth and eleventh TMDs. It changes the codon TCT to GCT, resulting in Serine being replaced by Alanine. Finally, T3435C is a synonymous variant that occurs in the second NBD, causing the codon ATT to be changed to ATC but keeping the encoded amino acid Isoleucine [10]. All three variants occur in intracellular parts of the protein [8]. Though the frequencies vary for different ethnicities, they are common in the general population; In fact, the alternative allele is present in 57%, 55% and 49% of genotypes for T1236C, T2677G and T3435C respectively [11]. The three also make a common haplotype, occurring in 33% of individuals [12].

Variants, particularly ones defined as “silent” such as synonymous alterations, can modify the gene expression process. They can affect transcription through changing transcription-factor binding sites (TFBS) [13]; they can impact translation and co-translational folding (CTF) through changing codon usage [14, 15]; they can also accelerate mRNA degradation by changing affinity to miRNA binding sites [16] or lead to mis-spliced proteins by modifying the nucleotides in the vicinity of donor or acceptor sites [17]. Altogether, they can impact all phases of gene expression. Though these three variants are widely studied, conclusions regarding their effects have been exceptionally inconsistent [18]. Some studies suggest they change mRNA or protein expression [19–23] while others claim they do not [24–29] and associate their presence with modifications in protein structure [24, 29]. Some find them to significantly impact patients’ response to chemotherapy and survivability while others do not [30–46]. Remarkably, controversy is also found among those who claim that the variants do affect survivability [47]; whether the variants are detrimental or advantageous for patient survival remains a matter of disagreement.

The effects of these MDR1 variants on p-gp expression, function and on chemotherapy resistance have been studied mostly in clinical trials and in-vitro cell line experiments. Each method has inherent disadvantages; results of clinical trials are affected by factors such as cohort size, patient ethnicity, tumor type and conducted chemotherapy regimen. In-vitro experiments cannot fully mimic the symbiosis of a cell with its natural environment and substantially differ from the corresponding cell type in-vivo [48]. Even though the effects of the variants may vary under different conditions, it is possible that some of the contradictions emerge not due to biological complexity but rather due to impediments in scientific methods. The objective of this paper is to examine, for the first time, the effects of T1236C, T2677G and T3435C on all phases of the gene expression

process using various computational approaches. We aim to gain a better understanding of the mechanism of action of this three MDR1 variants, endeavoring to propose plausible explanations and validate some of the previously proposed effects.

2 Methods

2.1 Data Sources

For performing the various analyses, we utilized data of several known databases. Single Nucleotide Variants (SNV) data, mRNA expression data and clinical data of TCGA [49, 50] projects was downloaded from the GDC [51] (https://gdc.cancer.gov/) on November 2021. The complete human CDS was downloaded from Ensembl [52, 53] (https://ftp.ensembl.org/pub/release-109/fasta/homo_sapiens/cds/) on 2020. Human protein expression measurements were downloaded from PaxDb [54] on 2020 (https://pax-db.org/dataset/9606/1502934799/).

2.2 MDR1 Expression

MDR1 Expression Change of TCGA Patients. SNVs and expression data of TCGA patients were used to examine the effects of T1236C, T2677G, T3435C and the haplotypes on MDR1 expression. For each mutation or haplotype, the cohort was split to carriers and non-carriers groups. The non-carriers group was randomly sampled 100,000 times such that each sampled group contained the same amount of patients as the carriers group. The average MDR1 expression was calculated for each sampled group and used to create a distribution of the MDR1 expression for non-carriers. The average MDR1 expression of the carriers group was compared to the distribution and an empirical p-value was calculated.

MDR1 Expression Change According to Enformer. Enformer [55] is a state-of-the-art model that predicts gene expression and transcription regulation based on an extremely long input sequence (hundreds of thousands of nucleotides). Enformer was used to predict MDR1 expression levels both for the wild type and the mutated sequence, for all three variants. As the output of Enformer is made of predictions for multiple genomic tracks that differ in tissues and measure of gene expression, we used several different approaches to obtain a single score from the output. For example, some of the scores only considered tracks related to tissues where MDR1 is highly expressed, others only considered CAGE tracks, some accounted all tracks. The same post-processing was performed for random variants with similar characteristics (explained in the "Empirical p-values" section) and the scores of the original and random variants were compared. If the score of the variant was larger than the scores of 95% of the random variants it was deemed to significantly change MDR1 expression.

2.3 Splicing Events

To examine the effect of T1236C, T2677G, T3435C on the occurrence of splicing events we used SpliceAI [56], a model that, given a genomic sequence, predicts the probabilities

of each site to be a donor or acceptor site. We perform the prediction for both the wildtype and mutated sequence, centered around each of the three variants. For each position in the prediction range, we calculated the difference in the probability of it being a donor/acceptor site that was caused by the variant. We searched for positions for which the probability changed by more than 50%. Positions exhibiting an increment of more than 50% are considered new prospective donor/acceptor sites, whereas positions for which the probability decreases by more than 50% are regarded as potentially abolished donor/acceptor sites.

2.4 Translation Rates

To examine the effect of T1236C, T2677G and T3435C on translation rates we utilized several measures positively correlated with it - MFE, CAI, FPTC and tAI. We calculated these measures for both the wildtype and mutated MDR1 CDS sequences and examined the difference in these measures at the position of the variant.

MFE. A per-position MFE score was computed using ViennaRNA [57]; first, a sliding window (length = 39 nucleotides, stride = 1 nucleotide) was used to obtain a per-window MFE score. Then, the MFE score of a specific position in the CDS was set as the average of all MFE scores of the windows that the position is in.

CAI. Human CAI weights were computed as suggested in the original paper [58]. The set of highly expressed genes (15% most highly expressed) was curated using human protein expression levels from PAXdb [54].

FPTC. Human FPTC weights were downloaded from the Kazusa website [59].

tAI. tAI tissue-specific weights were taken from Hernandez-Alias et al. [60] and the s weights were optimized as depicted in Sabi et al. [61].

2.5 Co-translational Folding

To examine the effect of T1236C, T2677G and T3435C on the CTF of p-gp, a computational model that assesses which positions in the CDS are important for correct protein folding was deployed. The model is yet to be published and therefore we will provide much detail about the model's methodology. The analysis is based on the basic assumption that CTF is governed by local translation rates, and that this rate is evolutionary conserved for a position that is crucial for correct folding [62]. Thus, it searches for positions with both evolutionarily conserved low and evolutionarily conserved high MFE (a measure correlated with translation rate) across orthologous versions of a gene. All MDR1 orthologous CDSs (n = 383) were downloaded from Ensembl (https://rest.ensembl.org/documentation/info/homology_ensemblgene) and were aligned using Clustal Omega [63]. The MFE score per nucleotide position was calculated for all sequences in the multiple sequence alignment (MSA). Then, we calculated the average MFE at each CDS position across the different organisms. To find the positions were the MFE score is conserved as significantly lower or higher than expected by chance, we created two kinds of permuted versions of the MSA. In the first method (named "vertical

permutation"), we shuffle synonymous codons within the same column of the MSA. In the second method (named "horizontal" permutation), we horizontally swap between synonymous codons of pairs of columns. Both methods affect the MFE scores while keeping basic characteristics of the MSA such as the amino-acid, codon and nucleotide content. One hundred permutations of the MSA are created for each kind. We calculate the per-position MFE score averaged across orthologs for each permuted MSA in the same manner as was done for the original one. At this point, for each CDS position we have a single true MFE score and one hundred scores from each of the permuted versions. We utilized the permutations to calculate a z-score for each position. Finally, we intersect the results to get positions that had significantly low or high MFE when compared to both kinds of permuted versions of the MSA.

2.6 Survivability of TCGA Patients

To examine the effect of T1236C, T2677G, T3435C and their haplotypes on the survival of cancer patients we used TCGA SNV and clinical data. We used the vital status of the patients and a matching time-stamp in order to create Kaplan-Meier survival curves [64] for the carriers and non-carriers groups. The time-stamp was derived from the maximum value of the following attributes- "Days from the initial diagnosis to current follow-up", "Days from the initial diagnosis to the current confirmation of vital status", "Days from the initial diagnosis to patient death". The logrank test [65] was used to assess whether the survival curves of the carriers and non-carriers group significantly differ.

To assert that the results are not caused by differences in tumor mutational burden (TMB), we repeated the analysis when controlling for this confounder; the balance of the TMB between the mutated and control groups was tested using a two sample KS test [66].

2.7 Stratification of TCGA Patients

For several analyses TCGA patients were stratified according to their cancer types as suggested by Gao et al. [67] (see Table 1). The patients were assigned to one of three groups–metabolic cancers, which are associated with altered metabolic pathways, proliferative cancers, which are associated with dysregulated cell proliferation, and inflammatory cancers, which are associated with immune system dysregulation. Because cancers in these categories dysregulate different pathways, it is likely that drug resistance patterns also differ between these categories. For example, metabolic cancers could dysregulate drug-metabolizing enzymes or drug efflux pumps, while proliferative cancers have rapid division rates and could lead to emergence of drug resistant clones through acquisition of new mutations.

2.8 Empirical p-Values

To infer the significance of the changes caused by T1236C, T2677G and T3435C in several of the analyses, they were compared to changes caused by random variants with similar characteristics. T2677G is a non-synonymous variant and therefore its effect was

compared to other, randomly sampled, T -> G non-synonymous variants in the MDR1 CDS. T1236C and T3435C are synonymous variants and therefore were both compared to randomly sampled synonymous T -> C variants in the MDR1 CDS. Each original variant was compared to one-hundred randomized sequences.

2.9 TCGA Variants Correlated with T1236C

T1236C was found as correlated with worse survival. To investigate the possibility of a causal relationship, we examined mutations that are highly correlated with T1236C and their effect on patient survival, searching for other potential effects. Highly correlated variants were defined as variants that are detected in the genomes of more than 75% of T1236C positive patients.

3 Results

Diverse computational models were deployed aiming to assess whether T1236C, T2677G and T3435C could modify different phases of the gene expression process. Additionally, genetic, clinical and expression data from The Cancer Genome Atlas (TCGA) were analyzed for this purpose. It is important to mind the difference between a germline variant, a genetic variation that is inherited and may be prevalent in the population, and a somatic variant which is a genetic variation that is acquired during one's lifetime. While most of the previous research regarding MDR1 analyses germline variants, our study examines both germline and somatic ones; when analyzing TCGA data we strictly evaluate the effect of somatic variants in cancerous tissues, whereas the rest of the computational analyses that do not utilized data from TCGA can be used to interpret the effects of both somatic and germline variants, as they simply analyze the effects caused by the nucleotide change and are "blind" to the type of variant.

3.1 T1236C, T2677G and T3435C Do not Seem to Significantly Affect MDR1 Expression Levels

MDR1 abundance was compared between TCGA patients that acquired any of the three somatic variants and patients who did not (see Methods). Results, shown in Fig. 1, indicate a marginal association between T1236C and MDR1 expression ($p = 0.06$) and no association between T2677G and T3435C with MDR1 expression levels. Differences in mRNA abundance between carriers of the haplotypes and non-carriers were also found non-significant ($0.11 < p < 0.24$). When stratifying patients according to cancer types (Table 1), no significant changes in expression were found as well.

Additionally, we used Enformer [55] to computationally assess whether the variants are likely to cause a change in MDR1 expression. Enformer is a transformer-based neural network that is trained on thousands of epigenetic and transcriptional datasets to predict gene expression. It identifies complicated genomic patterns related to gene expression regulation such as TSSs, TFBSs, histone modifications sites and miRNA binding sites. With an input sequence of 393,216 nucleotides, it is currently the gene expression predictor with the widest receptive field. For each variant we ran Enformer for the reference sequence and for the mutated sequence (see Methods). The predicted change in MDR1 expression was not significant for any of the variants, including T1236C.

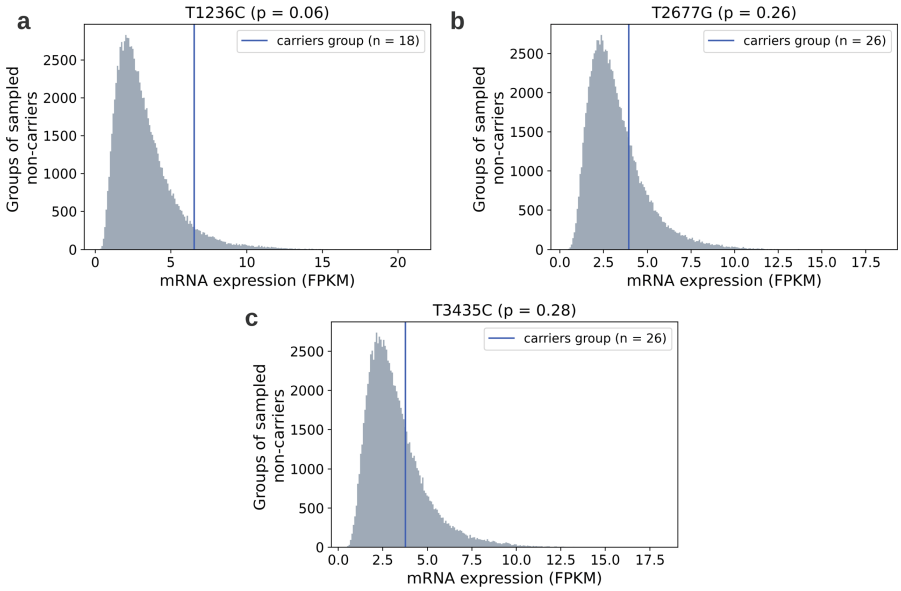

Fig. 1. MDR1 expression levels of carriers vs. non-carriers of the three variants in the TCGA database. a: T1236C; b: T2677G; c: T3435C.Comparison of the mean MDR1 expression of the carriers group (vertical line) to the mean MDR1 expression levels of 100,000 groups of randomly chosen non-carriers (distribution). Size of the carriers' group and non-carriers' groups are the same.

3.2 T1236C, T2677G and T3435C Do not Seem to Affect mRNA Splicing

SpliceAI [56] was used to assess whether the variants change MDR1 splicing. Given a sequence, the output of spliceAI is a matrix of probabilities, indicating the probability of each position in the sequence to be a donor or acceptor site (see Methods). For each variant we ran SpliceAI for the reference sequence and for the mutated sequence. Using the difference between probabilities we searched for canceled or newly generated donor and acceptors sites, as these events could lead to alternative splicing. None of the variants were found to cause significant changes.

3.3 T1236C, T2677G and T3435C Potentially Increase p-gp Levels Through Raising Global Translation Rates

The rate of translation elongation is a key component that shapes protein abundance. A change in the translation rate of a single codon can impact protein abundance if it is located in a translation bottleneck. The translation rate varies throughout the mRNA sequence and is dependent on multiple cellular conditions and factors such as accessibility of the mRNA to the ribosome, codon usage bias and availability of tRNAs. In this section we examine several measures correlated with the translation rate and evaluate whether the three variants are expected to change it, locally or globally.

T1236C, T2677G and T3435C All Decrease mRNA Folding Strength. T1236C and T3435C Potentially Increase Global Translation Rates Due to this Decrease. Minimum Free Energy (MFE) is a measure derived from predictions of the secondary structure of an mRNA sequence; a lower MFE value indicates a structure that is tightly packed and less accessible for translation. Therefore, lower MFE values are correlated with lower translation rates [68]. We computed MFE scores for the region surrounding each of the three variants (see Methods) and predicted the instigated change in MFE. The results (Fig. 2) demonstrate that all three variants cause an increase in MFE. This increase is significantly larger ($p = 0.01$, $p = 0.02$ and $p = 0.04$ for T1236C, T2677G and T3435C respectively) than the increase caused by random variants with similar characteristics in the MDR1 gene (see Methods). Moreover, positions 1236 and 3435 (with the wild type alleles "T") have relatively low MFE scores compared to the rest of the coding sequence (CDS) positions (19^{th} and 12^{th} percentile respectively) and therefore constitute possible translation bottlenecks (Fig. 7), suggesting that T1236C and T3435C could also increase global translation rates.

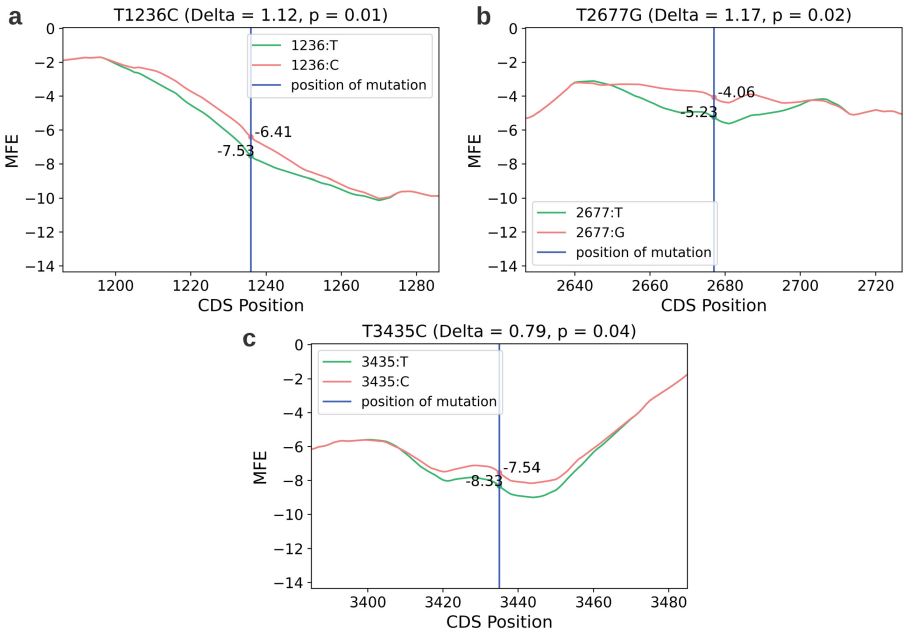

Fig. 2. Effect of the three variants on MFE in their vicinity. a: T1236C; b: T2677G; c: T3435C. x axis: the nucleotide position in the coding sequence (CDS); y axis: MFE score. The vertical line indicates the position of the variant, and the curves depict the MFE scores of the nucleotides proximal to the position of the variant, with (pink) or without it (green). (Color figure online)

T1236C, T2677G and T3435C Increase the Optimality of Codon Usage. T1236C Potentially Increases Global Translation Rates Due to This Increase. We compute Codon Adaptation Index (CAI) and Frequency Per 1000 codons (FPTC) to measure the change in codon usage bias (CUB) caused by the three variants, where CAI is computed for the synonymous variants and FPTC for the non-synonymous variant. Higher CUB suggests better adaptation of the sequence to the cellular translation machinery and resources and is correlated with faster translation [69]. Results (Fig. 3) indicate that all three variants cause a less prevalent codon to be replaced by a more prevalent codon, suggesting an increase in local translation rate. Moreover, T1236C substitutes the least prevalent codon of Glycine to the most common one. When comparing to random variants with similar characteristics in the MDR1 gene, the increase in CAI caused by T1236C is marginally significant ($p = 0.08$). Also, the CAI score of codon 412 (in which position 1236 resides) is in the 16^{th} percentile of CAI scores compared to all the codons in the CDS, further strengthening the possibility that it is a translational bottleneck and that T1236C could increase global translation rate.

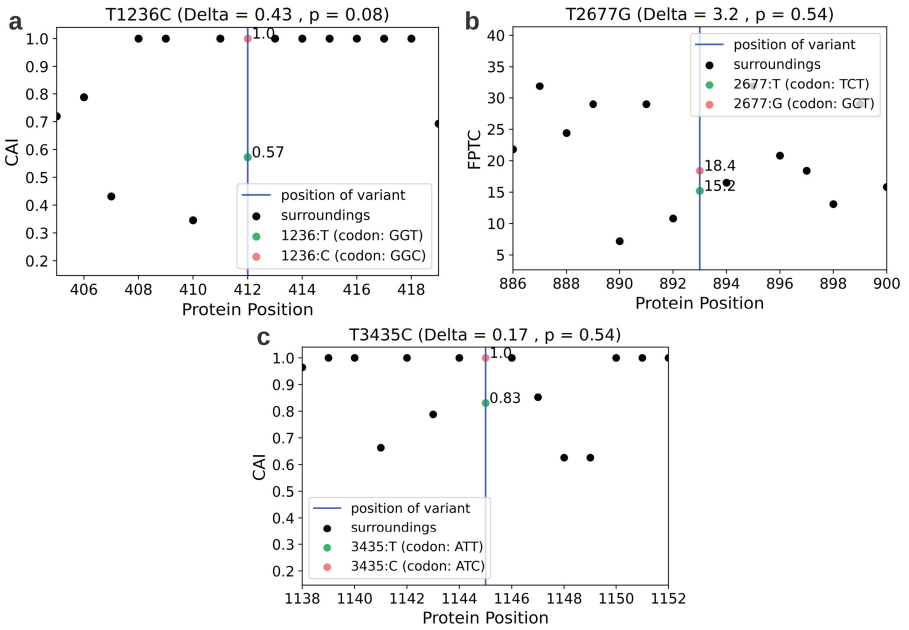

Fig. 3. Effect of the three variants on CUB. a: T1236C; b: T2677G; c: T3435C. x axis: the amino-acid position in the protein sequence; y axis: the CUB score (CAI/FPTC). The vertical line indicates the position of the variant. Black dots indicate the CUB scores in the vicinity of the variant and the dots on the vertical line indicate the CUB score of the mutated position, before (green) and after (pink). (Color figure online)

T1236C and T2677G Improve Adaptation to the tRNA Pool in Tissues Where MDR1 is Expressed. T2677G Potentially Increases Global Translation Rates Due to the Improvement in Adaptation The tRNA Adaptation Index (tAI) is a measure of translational efficiency which considers the intracellular concentration of tRNA molecules and the efficiencies of each codon–anticodon pairing [70]. Higher tAI scores are given to codons with better tRNA availability and are correlated with a higher translation rate [71]. The tRNA availability changes substantially between different tissues and organs. We examine the effect of the variants on the tAI profile in tissues where the MDR1 gene is typically expressed (see Methods) – liver, kidney, colon and brain tissues. tAI scores for the adrenal glands were not available. Figure 4 demonstrates the effect of the variants on the tAI profile in the liver tissue, but it is similar for all other examined tissues (more data will be report in the full version of this paper). The results of this analysis show that both T1236C and T2677G cause an increase in tAI. Moreover, T2677G replaces a codon with an extremely low tRNA availability to a codon with much higher tRNA availability, in all examined tissues. When comparing the tAI change caused by T2677G to changes caused by random variants with similar characteristics in the MDR1 gene, the former is found either significantly or marginally significantly larger in all relevant tissues ($p <= 0.1$). Moreover, the tAI score of the codon affected by T2677G is in the 1^{st}–5^{th} (depending on the tissue) percentile of the tAI scores of all the codons in

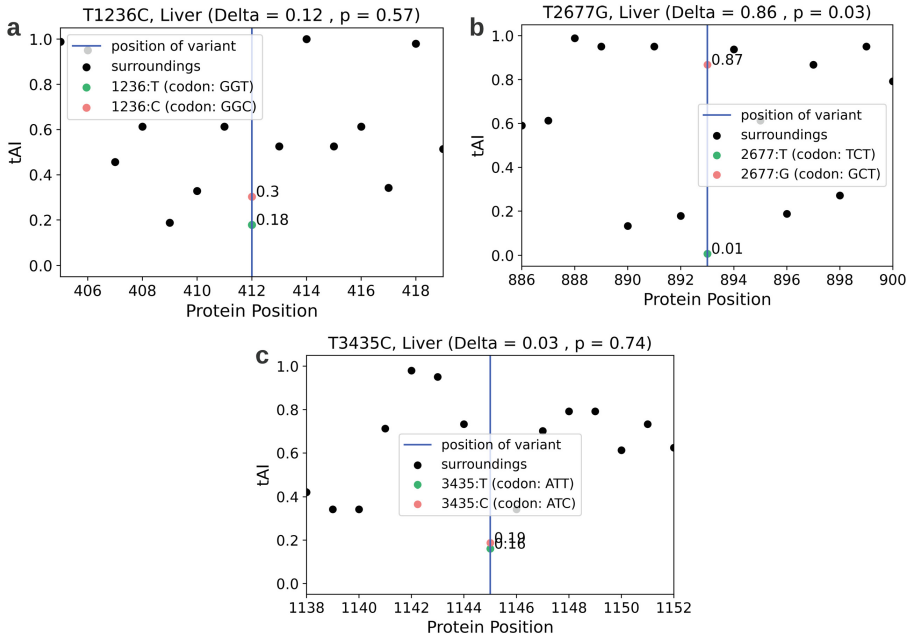

Fig. 4. Effect of the three variants on tAI. a: T1236C; b: T2677G; c: T3435C.x axis: the amino-acid position in the protein sequence; y axis: the tAI score. The vertical line indicates the position of the variant. Black dots indicate the tAI scores in the vicinity of the variant and the dots on the vertical line indicate the tAI score of the mutated position, before (green) and after (pink). (Color figure online)

the CDS, strengthening the indication that the position could be a translation bottleneck and that T2667G could increase global translation speed.

To conclude, the variants are in positions that are potential translation bottlenecks as they all obtain scores that are close to the minimal score of the entire CDS in at least one of the examined measures. All three variants lead to an increase of MFE and CUB scores, the increase in MFE being significantly larger than expected by chance. Additionally, T1236C and T2677G cause an increase in tAI scores in tissues where MDR1 is expressed, with T2677G leading to an extreme and significant tAI change. Notably, a variant can cause an increase of one factor while decreasing the others and vise-versa. In fact, when examining all synonymous variants across TCGA we find a negative correlation between the effect of variants on MFE and CAI ($p = -0.13$, $p < 10\text{^}-323$) and between the effect on MFE and tAI ($p = -0.18$, $p < 10\text{^}-323$). The increase in MFE, CAI and tAI caused by all three variants provides accumulating evidence and suggest that T1236C, T2677G and T3435C likely increase local translation rates in their vicinity, and possibly increase global translation rates and protein levels.

3.4 T3435C Potentially Modifies Co-translational Protein Folding Through Raising Local Translation Rate in a Conserved Slowly Translated Region

CTF is the mechanism in which the nascent protein begins to fold during its translation [72]. This process was shown to be governed by local translation rates [62]. To examine whether T1236C, T2677G and T3435C influence CTF we deployed a model which detects positions that have evolutionary conserved extreme MFE scores. The rationale being that MFE scores can be used as proxies for translation rates and that positions with evolutionary conserved extreme translation rates (especially slow rates) are largely conserved as such due to their importance for optimal CTF. Therefore, it is possible that variants in these positions interfere with the CTF mechanism. In order to find positions in the CDS of MDR1 with conserved extreme MFE scores, the model utilizes orthologous MDR1 genes from hundreds of organisms, calculates their MFE profiles and compares them to the MFE profiles of permuted versions of these genes (see Methods).

Model output (Fig. 5) indicates that T3435C is in a position with evolutionary conserved low MFE, surrounded by a stretch of positions with conserved low MFE (6 nucleotides upstream of T3435C and 17 downstream of it). Combining the results of this model and the results of the previous analysis which suggests that T3435C causes a local increase in translation rate, we deduce that it is possible that T3435C causes an increase in translation rate in a region of conserved low translation rates, and thus modifies CTF. Both T1236C and T2677G were not found to be in a position of evolutionary conserved extreme MFE.

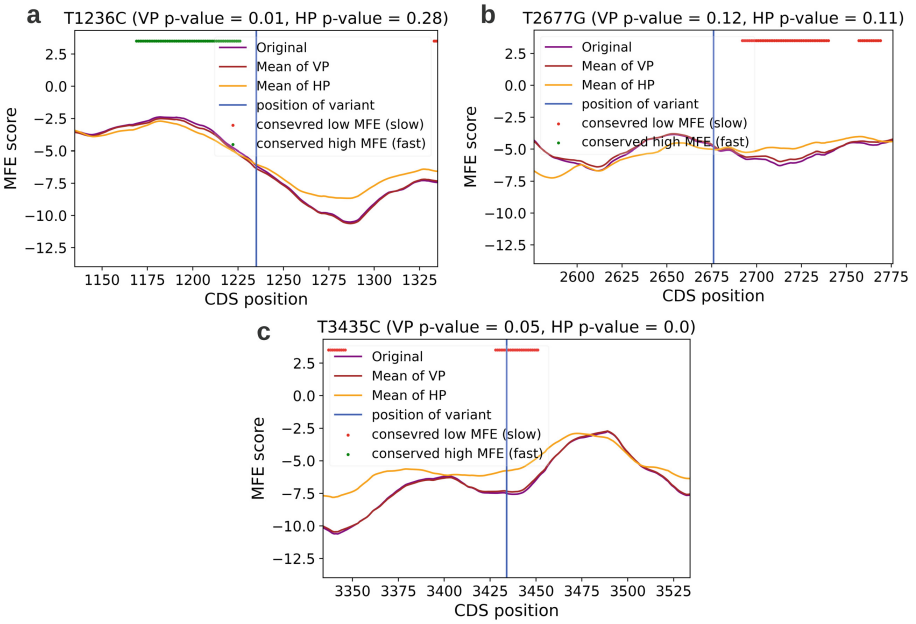

Fig. 5. Regions of evolutionary conserved low or high MFE in the vicinity of the three variants. a: T1236C; b: T2677G; c: T3435C. x axis: the nucleotide position on the CDS of the MDR1 gene. y axis: the MFE score, which is used by this model as the proxy for translation rate. The purple curve indicates the mean MFE score of the true MDR1 CDS across 383 orthologs. The orange and brown curves indicate the mean MFE score of a permuted MDR1 CDS, using two different kinds of permutations – vertical permutations (VP) and horizontal permutations (HP). Green dots are nucleotide positions for which the MFE of the true MDR1 was significantly higher than the MFE of both permutated versions, across organisms. Red dots are nucleotide positions for which the MFE of the true MDR1 was significantly lower than the MFE of both randomized versions, across organisms.

3.5 T1236C Potentially Decreases Patient Survivability

Clinical data of TCGA patients were used to assess the variants' effect on survivability. Kaplan Meier curves [64] were compared between patients that acquired one or more of the three variants in their cancerous tissue and those who did not (see Methods). Results (Fig. 6 and Fig. 8 for the haplotypes) suggest that T1236C decreases survival probability (logrank $T = 5.22$, $p = 0.02$. Bonferroni corrected $p = 0.06$) whereas T2677G and T3435C do not. Other TCGA variants that are highly correlated with the presence of T1236C (see Methods) are not correlated with decreased survivability (Fig. 9), enhancing the likelihood that T1236C is a causal variant.

When stratifying patients according to cancer types (Table 1), we find a significant negative impact ($p = 0.002$) on the survival of patients with inflammatory cancers (Fig. 10).

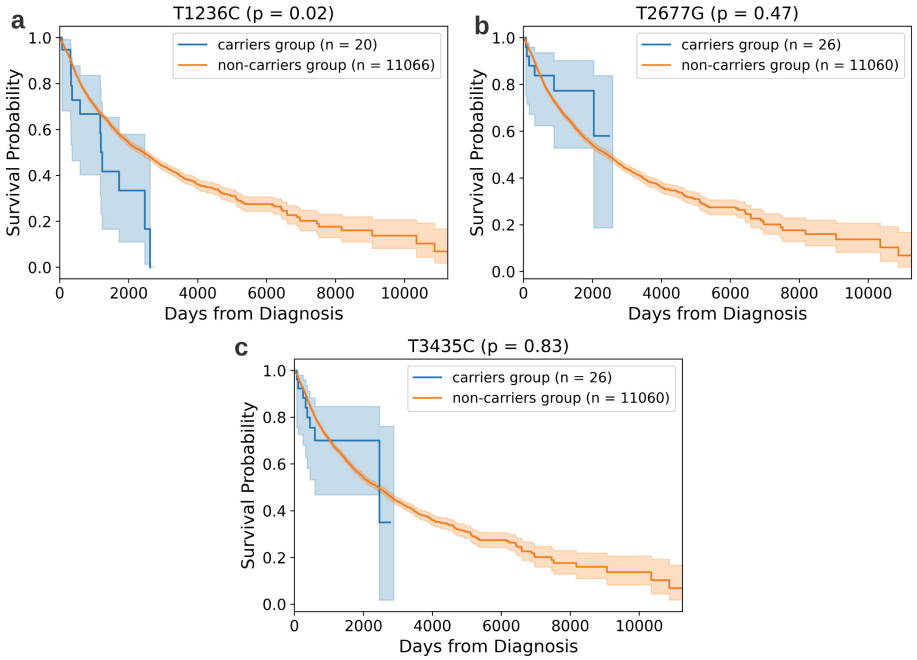

Fig. 6. Effect of the three variants on patient survivability. a: T1236C; b: T2677G; c: T3435C. Comparison of the Kaplan-Meier survival curves of the carriers' group and the non-carriers group of the three variants.

4 Discussion

Since the discovery of MDR1 and its many variants, numerous in-vitro and in-vivo studies have attempted to understand their mechanisms of action and contribution to chemotherapy resistance and cancer prognosis. Due to contradictory findings and the lack of success of MDR1 inhibitors in clinical trials [73], research of MDR1 and its variants has gradually diminished. However, the current surge in genomic data and evolving technology enables the examination of this case from a new, data-driven, perspective.

Our analyses aid to paint a clearer image, providing evidence of accumulating effects. T1236C causes an increase in CAI, tAI and MFE, all positively correlated with translation rates. Because it is in a region of low MFE and CAI, this increase could lead to a rise in global translation rates and p-gp over-expression. Moreover, T1236C was found to be correlative with worse patient prognosis, which can be explained by overexpression of the protein. T2677G also causes an increase in measures correlated in translation rates and most significantly effects tAI. It causes a codon that is extremely rare in tissues where MDR1 is expressed to be replaced with a common codon. T3435C increases MFE and CAI as well, in a region of very low MFE. Additionally, it possibly modifies p-gp structure through changing local translation rate. Altogether, when exploring the mechanism of action of T1236C, T2677G and T3435C, it is much more likely that they cause an increase in p-gp expression rather than a decrease, providing an advantage for the cancer cell. Due to these findings, we expected patients with the CGC haplotype to have

significantly worse survival, but it was not detected (Fig. 8). Perhaps the combination of the three variants leads to a more complexed effect than the accumulated impact of each one separately. Alternatively, perhaps a significant effect on survival was not observed due to the small size of the cohort and confounders such as patients' cancer type and treatment regime.

Though the findings of previous research on the matter are inconsistent, our analyses support the conclusions of several major studies. Kimchey-Zarfaty at el. Demonstrated that T3435C, combined with either T1236C or T2677G, changes MDR1 substrate specificity [24]. They hypothesized that T3435C modifies CTF (and therefore protein conformation) because it changes the local translation rate in a slowly translated region, important for correct CTF. They supported their hypothesis by identifying a cluster of non-prevalent codons in the vicinity of T3435C (as can be seen in Fig. 3c). Our model, which utilizes MFE and assesses which positions are evolutionarily conserved for slow or fast translation, indeed detects 3435 in a region that is likely conserved for slow translation, supporting their findings. Moreover, we demonstrated that T3435C increases both MFE and CAI scores, suggesting that it could increase the local translation rate in this slowly translated region and thus modify CTF. Kimchey-Zarfaty et al. also examined mRNA expression, protein expression and aberrant splicing events in wild-type and mutated p-gp and reports that no changes were caused by the variants. To the best of our knowledge, no prior study has suggested that T1236C, T2677G or T3435C cause alternative splicing; our results, obtained using SpliceAI [56], support the consensus. As for mRNA and protein expression, most in-vitro studies do not report a significant change, while some in-vivo studies do. This could be due to a complicated mechanism of action that cannot be detected out of the natural cellular environment. Nevertheless, reports of the in-vivo studies are also controversial, perhaps due to the tremendous variability in MDR1 expression between patients, small cohorts and difference in patient ethnicities, genetic backgrounds and tumor types. Combining the results of our analyses, we assess that if the variants do modify protein levels, they should cause an increase rather than a decrease. In another comprehensive study, Johnatty SE et al. [74] examine 4,616 ovarian cancer patients from the Ovarian Cancer Association Consortium (OVAC) and TCGA, that have received chemotherapy treatments. They found a marginal association of T1236C with worse overall survival and did not find association between T2677G or T3435C and survival parameters. Additionally, Chen et al. [75] performed a meta-analysis involving 3,320 patients across 15 studies and concluded that a TT genotype in position 1236 is associated with better overall survival. Both these findings are supported by our pan-cancer survival analysis on TCGA data, shown in Fig. 6.

When reviewing the analyses described in this study, we must remember the limitations of in-silico models as well; A major limitation of this study is the small cohorts of mutated patients; our carriers data is comprised of 20, 26 and 20 patients with T1236C, T2677G and T3435C respectively. In the future, once more genomic data is available, we suggest measuring both mRNA and protein levels of MDR1 of cancer patients that carry these mutations. Also, a significantly larger cohort could allow us to split patients according to their specific cancer types, ethnicities and received chemotherapy regimens while keeping the required statistical power. As MDR1 is related to the response

to chemotherapy, it is likely that the effect of the variants on patient survivability is dependent on the patient's received treatment; not all chemotherapeutic agents are substrates of p-gp, and the effect of the variant among different substrates could also vary, making this analysis highly important. Another limitation is the difficulty to create models that capture all relevant processes and interactions in highly complex biological mechanisms such as gene expression regulation. Moreover, some of the models were trained on data obtained from wet lab experiments and thus are also subjected to noise. Nonetheless, computational models have many advantages; they are often less expensive, faster and can incorporate larger quantities of data than in-vitro experiments. We believe the incorporation of computational, data-driven models is both essential and highly beneficial when analyzing the complex effects of variants on gene expression regulation, generally and specifically for the case of MDR1.

Acknowledgements. We thank Alma Davidson, Yoram Zarai, Sanjit Batra and Yun S. Song for their contributions in different analyses of this study.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Appendix

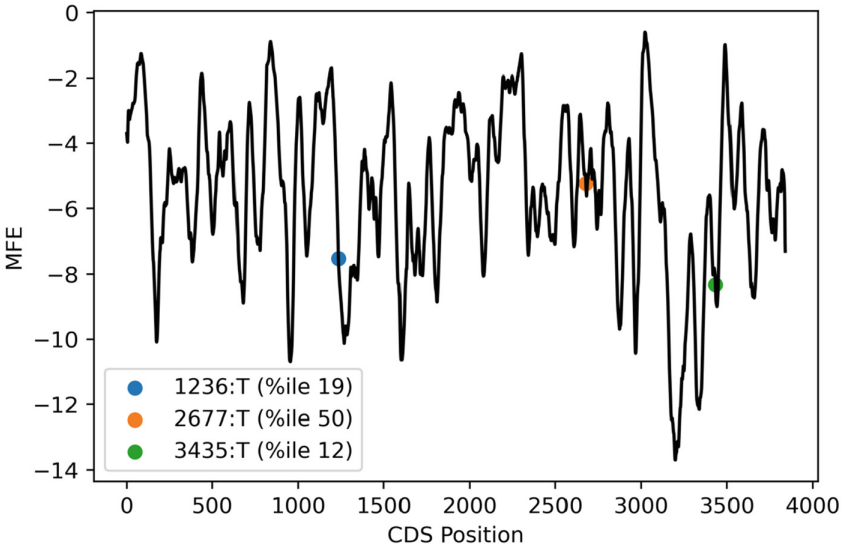

Fig. 7. MFE profile of the unmutated CDS sequence of MDR1. The positions of the variants are denoted in blue (1236), orange (2677) and green (3435). The percentile of the MFE score of the positions of the variants (when unmutated) is denoted in the legend. (Color figure online)

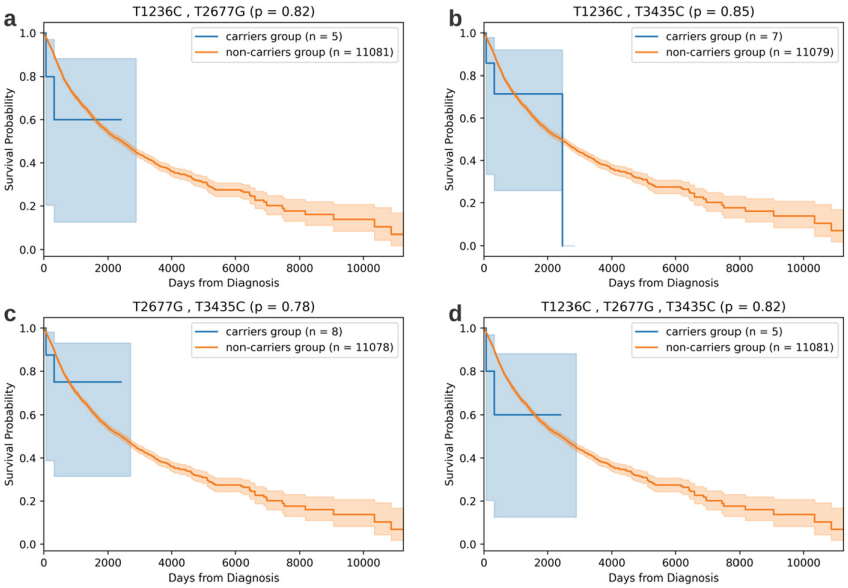

Fig. 8. Kaplan-Meier survival curves of the carriers and non-carriers of the combinations of the three variants, from the TCGA database. a: T1236C & T2677G; b: T1236C & T3435C; c: T2677G & T3435C;. d: T1236C, T2677G & T3435C.

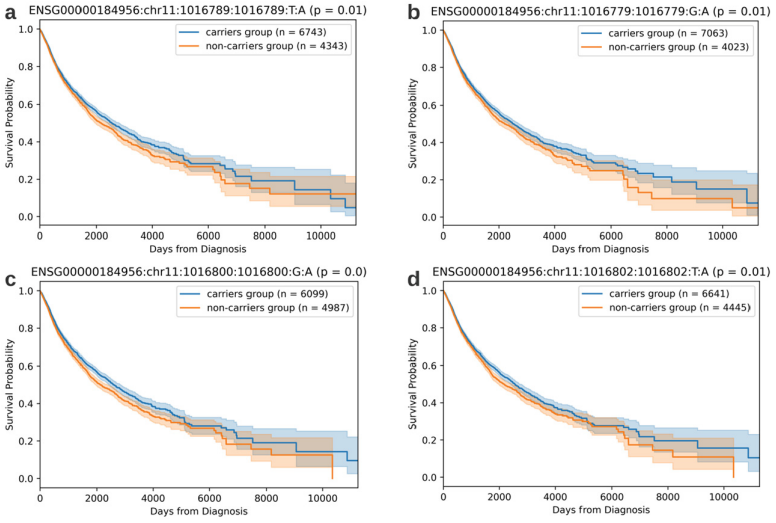

Fig. 9. Kaplan-Meier survival curves of carriers vs. non-carriers of TCGA mutations that are highly correlated with T1236C (present in the genomes of over 75% of T1236C positive patients)

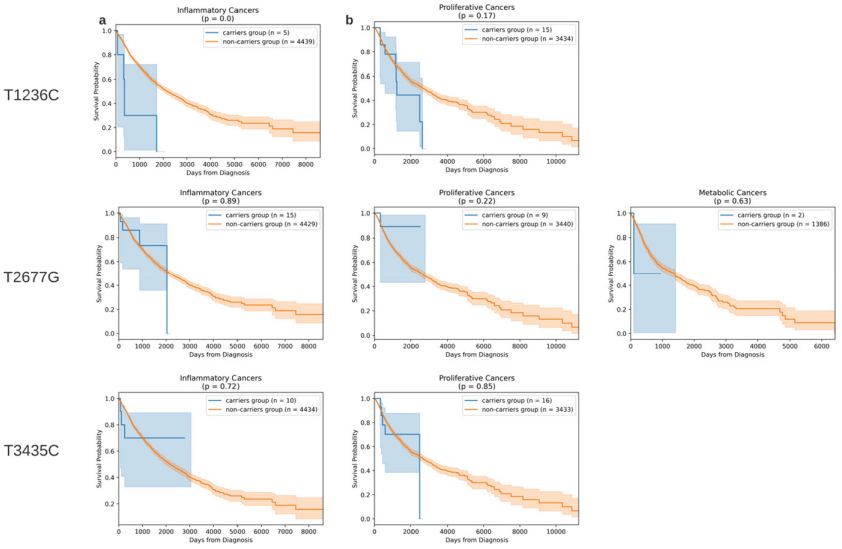

Fig. 10. Kaplan-Meier survival curves of carriers vs. non-carriers of the three variants, stratified cancer types. Left column – inflammatory cancers. Middle column- proliferative cancers. Right column- metabolic cancers. Categories are described in Table 1.

Table 1. Cancer type clusters

Cluster	Cancer Types
Metabolic Cancers	LAML, UCS, HNSC, ESCA, UVM, CHOL, LIHC
Proliferative Cancers	BLCA, SKCM, SARC, COAD, UCEC, MESO, ACC, LUAD, KIRC, KIRP, DLBC, TGCT, PRAD
Inflammatory Cancers	PAAD, LGG, CESC, GBM, READ, LUSC, BRCA, STAD, THCA, OV, KICH, PCPG, THYM

References

1. Juliano, R.L., Ling, V.: A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. BBA - Biomembranes **455**(1), 152–162 (1976). https://doi.org/10.1016/0005-2736(76)90160-7
2. Allikmets, R., Gerrard, B., Hutchinson, A., Dean, M.: Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database. Hum. Mol. Genet. **5**(10), 1649–1655 (1996). https://doi.org/10.1093/hmg/5.10.1649
3. Li, Y., Yuan, H., Yang, K., Xu, W., Tang, W., Li, X.: The structure and functions of P-Glycoprotein. Curr. Med. Chem. **17**(8), 786–800 (2010). https://doi.org/10.2174/092986710790514507

4. Sakaeda, T., Nakamura, T., Okumura, K.: MDR1 genotype-related pharmacokinetics and pharmacodynamics. Biol. Pharm. Bull. **25**(11), 1391–1400 (2002). https://doi.org/10.1248/bpb.25.1391
5. Thiebaut, F., Tsuruo, T., Hamada, H., Gottesman, M.M., Pastan, I.R.A., Willingham, M.C.: Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. Proc. Nat. Acad. Sci. **84**(21), 7735–7738 (1987). https://doi.org/10.1073/pnas.84.21.7735
6. Tanigawara, Y.: Role of P-glycoprotein in drug disposition. Ther. Drug Monit. **22**(1), 137–140 (2000). https://doi.org/10.1097/00007691-200002000-00029
7. Kartner, N., Evernden-Porelle, D., Bradley, G., Ling, V.: Detection of P-glycoprotein in multidrug-resistant cell lines by monoclonal antibodies. Nature **316**(6031), 820–823 (1985). https://doi.org/10.1038/316820a0
8. Fung, K.L., Gottesman, M.M.: A synonymous polymorphism in a common MDR1 (ABCB1) haplotype shapes protein function. Biochimica et Biophysica Acta (BBA)-Proteins Proteomics, **1794**(5), 860–871 (2009). https://doi.org/10.1016/j.bbapap.2009.02.014
9. Wang, L.-H., Song, Y.-B., Zheng, W.-L., Jiang, L., Ma, W.-L.: The association between polymorphisms in the MDR1 gene and risk of cancer: a systematic review and pooled analysis of 52 case-control studies. Cancer Cell Int. **13**, 46 (2013). https://doi.org/10.1186/1475-2867-13-46
10. Panczyk, M., Balcerczak, E., Piaskowski, S., Jamroziak, K., Pasz-Walczak, G., Mirowski, M.: ABCB1 gene polymorphisms and haplotype analysis in colorectal cancer. Int. J. Colorectal Dis. **24**(8), 895–905 (2009). https://doi.org/10.1007/s00384-009-0724-0
11. Sherry, S.T., et al.: DbSNP: the NCBI database of genetic variation. Nucleic Acids Res. **29**(1), 308–311 (2001). https://doi.org/10.1093/nar/29.1.308
12. Spooner, W., et al.: Haplosaurus computes protein haplotypes for use in precision drug design. Nat. Commun. **9**(1), 4128 (2018). https://doi.org/10.1038/s41467-018-06542-1
13. Wang, S.Y., et al.: A synonymous mutation in IGF-1 impacts the transcription and translation process of gene expression. Mol. Ther.-Nucleic Acids **26**, 1446–1465 (2021). https://doi.org/10.1016/j.omtn.2021.08.007
14. Tarrant, D., Von Der Haar, T.: Synonymous codons, ribosome speed, and eukaryotic gene expression regulation. Cell. Mol. Life Sci. **71**(21), 4195–4206 (2014). https://doi.org/10.1007/s00018-014-1684-2
15. Walsh, I.M., Bowman, M.A., Soto Santarriaga, I.F., Rodriguez, A., Clark, P.L.: Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. Proc. Nat. Acad. Sci. **117**(7), 3528–3534 (2020). https://doi.org/10.1073/pnas.1907126117
16. Gu, W., Wang, X., Zhai, C., Xie, X., Zhou, T.: Selection on synonymous sites for increased accessibility around miRNA binding sites in plants. Mol. Biol. Evol. **29**(10), 3037–3044 (2012). https://doi.org/10.1093/molbev/mss109
17. Mueller, W.F., Larsen, L.S., Garibaldi, A., Hatfield, G.W., Hertel, K.J.: The silent sway of splicing by synonymous substitutions. J. Biol. Chem. **290**(46), 27700–27711 (2015). https://doi.org/10.1074/jbc.M115.684035
18. Robey, R.W., Pluchino, K.M., Hall, M.D., Fojo, A.T., Bates, S.E., Gottesman, M.M.: Revisiting the role of ABC transporters in multidrug-resistant cancer. Nat. Rev. Cancer **18**(7), 452–464 (2018). https://doi.org/10.1038/s41568-018-0005-8
19. He, H., et al.: Association of ABCB1 polymorphisms with prognostic outcomes of anthracycline and cytarabine in Chinese patients with acute myeloid leukemia. Eur. J. Clin. Pharmacol. **71**(3), 293–302 (2015). https://doi.org/10.1007/s00228-014-1795-6
20. Hemauer, S.J., Nanovskaya, T.N., Abdel-Rahman, S.Z., Patrikeeva, S.L., Hankins, G.D., Ahmed, M.S.: Modulation of human placental P-glycoprotein expression and activity by MDR1 gene polymorphisms. Biochem. Pharmacol. **79**(6), 921–925 (2010). https://doi.org/10.1016/j.bcp.2009.10.026

21. Hoffmeyer, S., et al.: Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. Proc. Nat. Acad. Sci. **97**(7), 3473–3478 (2000). https://doi.org/10.1073/pnas.97.7.3473
22. Song, P., et al.: G2677T and C3435T genotype and haplotype are associated with hepatic ABCB1 (MDR1) expression. J. Clin. Pharmacol. **46**, 373–379 (2006). https://doi.org/10.1177/0091270005284387
23. Pang, L., et al.: ATP-binding cassette genes genotype and expression: a potential association with pancreatic cancer development and chemoresistance? Gastroenterol. Res. Pract. **2014**, 414931 (2014). https://doi.org/10.1155/2014/414931
24. Kimchi-Sarfaty, C., et al.: A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science **315**(5811), 525–528 (2007). https://doi.org/10.1126/science.1135308
25. Kroetz, D.L., et al.: Sequence diversity and haplotype structure in the human ABCB1 (MDR1, multidrug resistance transporter) gene. Pharmacogenet. Genomics **13**(8), 481–494 (2003). https://doi.org/10.1097/00008571-200308000-00006
26. Gow, J.M., Hodges, L.M., Chinn, L.W., Kroetz, D.L.: Substrate-dependent effects of human ABCB1 coding polymorphisms. J. Pharmacol Exp. Ther. **325**(2), 435–442 (2008). https://doi.org/10.1124/jpet.107.135194
27. Hung, C.C., Chen, C.C., Lin, C.J., Liou, H.H.: Functional evaluation of polymorphisms in the human ABCB1 gene and the impact on clinical responses of antiepileptic drugs. Pharmacogenet. Genomics **18**(5), 390–402 (2008). https://doi.org/10.1097/FPC.0b013e3282f85e36
28. Salama, N.N., Yang, Z., Bui, T., Ho, R.J.: MDR1 haplotypes significantly minimize intracellular uptake and transcellular P-gp substrate transport in recombinant LLC-PK1 cells. J. Pharm. Sci. **95**(10), 2293–2308 (2006). https://doi.org/10.1002/jps.20717
29. Fung, K.L., et al.: MDR1 synonymous polymorphisms alter transporter specificity and protein stability in a stable epithelial monolayer. Cancer Res. **74**(2), 598–608 (2014). https://doi.org/10.1158/0008-5472.CAN-13-2064
30. Ni, L.-N., et al.: Multidrug resistance gene (MDR1) polymorphisms correlate with imatinib response in chronic myeloid leukemia. Med. Oncol. **28**, 265–269 (2011). https://doi.org/10.1007/s12032-010-9456-9
31. Lu, Y., et al.: Host genetic variants of ABCB1 and IL15 influence treatment outcome in paediatric acute lymphoblastic leukaemia. Br. J. Cancer **110**(6), 1673–1680 (2014). https://doi.org/10.1038/bjc.2014.7
32. Zheng, Q., et al.: ABCB1 polymorphisms predict imatinib response in chronic myeloid leukemia patients: a systematic review and meta-analysis. Pharmacogenomics J. **15**(2), 127–134 (2015). https://doi.org/10.1038/tpj.2014.54
33. Chu, Y.-H., et al.: Association of ABCB1 and FLT3 polymorphisms with toxicities and survival in Asian patients receiving sunitinib for renal cell carcinoma. PLoS ONE **10**(8), e0134102 (2015). https://doi.org/10.1371/journal.pone.0134102
34. Munisamy, M., et al.: Pharmacogenetics of ATP binding cassette transporter MDR1 (1236C>T) gene polymorphism with glioma patients receiving Temozolomide-based chemoradiation therapy in Indian population. Pharm. J. **21**(2), 262–272 (2021). https://doi.org/10.1038/s41397-021-00206-y
35. Li, J.Z., Tian, Z.Q., Jiang, S.N., Feng, T.: Effect of variation of ABCB1 and GSTP1 on osteosarcoma survival after chemotherapy. Genet. Mol. Res. **13**(2), 3186–3192 (2014). https://doi.org/10.4238/2014.April.25.3
36. Olarte Carrillo, I., García Laguna, A.I., De la Cruz Rosas, A., Ramos Peñafiel, C.O., Collazo Jaloma, J., Martínez Tovar, A.: High expression levels and the C3435T SNP of the ABCB1 gene are associated with lower survival in adult patients with acute myeloblastic leukemia in

Mexico City. BMC Med. Genomics **14**(1), 1–9 (2021). https://doi.org/10.1186/s12920-021-01101-y
37. Balcerczak, E., Panczyk, M., Piaskowski, S., Pasz-Walczak, G., Sałagacka, A., Mirowski, M.: ABCB1/MDR1 gene polymorphisms as a prognostic factor in colorectal cancer. Int. J. Colorectal Dis. **25**(10), 1167–1176 (2010). https://doi.org/10.1007/s00384-010-0961-2
38. Caronia, D., et al.: Effect of ABCB1 and ABCC3 polymorphisms on osteosarcoma survival after chemotherapy: a pharmacogenetic study. PLoS ONE **6**(10), e26091 (2011). https://doi.org/10.1371/journal.pone.0026091
39. Wu, H., et al.: Roles of ABCB1 gene polymorphisms and haplotype in susceptibility to breast carcinoma risk and clinical outcomes. J. Cancer Res. Clin. Oncol. **138**(9), 1449–1462 (2012). https://doi.org/10.1007/s00432-012-1209-z
40. Knez, L., et al.: Predictive value of ABCB1 polymorphisms G2677T/A, C3435T, and their haplotype in small cell lung cancer patients treated with chemotherapy. J. Cancer Res. Clin. Oncol. **138**(9), 1551–1560 (2012). https://doi.org/10.1007/s00432-012-1231-1
41. Vivona, D., et al.: ABCB1 haplotypes are associated with P-gp activity and affect a major molecular response in chronic myeloid leukemia patients treated with a standard dose of imatinib. Oncol. Lett. **7**(4), 1313–1319 (2014). https://doi.org/10.3892/ol.2014.1857
42. Li, W., et al.: ABCB1 3435TT and ABCG2 421CC genotypes were significantly associated with longer progression-free survival in Chinese breast cancer patients. Oncotarget, **8**(67), 111041 (2017).https://doi.org/10.18632/oncotarget.22201
43. Gregers, J., et al.: Polymorphisms in the ABCB1 gene and effect on outcome and toxicity in childhood acute lymphoblastic leukemia. Pharm. J. **15**(4), 372–379 (2015). https://doi.org/10.1038/tpj.2014.81
44. Xiaohui, S., Aiguo, L., Xiaolin, G., Ying, L., Hongxing, Z., Yilei, Z.: Effect of ABCB1 polymorphism on the clinical outcome of osteosarcoma patients after receiving chemotherapy. Pak. J. Med. Sci. **30**(4), 886–890 (2014). https://doi.org/10.12669/pjms.304.4955
45. Liu, S., Yi, Z., Ling, M., Shi, J., Qiu, Y., Yang, S.: Predictive potential of ABCB1, ABCC3, and GSTP1 gene polymorphisms on osteosarcoma survival after chemotherapy. Tumor Biol. **35**(10), 9897–9904 (2014). https://doi.org/10.1007/s13277-014-1917-x
46. Zmorzynski, S., et al.: The relationship of ABCB1/MDR1 and CYP1A1 variants with the risk of disease development and shortening of overall survival in patients with multiple myeloma. J. Clin. Med. **10**(22), 5276 (2021). https://doi.org/10.3390/jcm10225276
47. Chen, Q., et al.: Prognostic value of two polymorphisms, rs1045642 and rs1128503, in ABCB1 following taxane-based chemotherapy: a meta-analysis. Asian Pac. J. Cancer Prev. **22**(1), 3–10 (2021). https://doi.org/10.31557/APJCP.2021.22.1.3
48. Graudejus, O., Wong, R., Varghese, N., Wagner, S., Morrison, B.: Bridging the gap between in vivo and in vitro research: reproducing in vitro the mechanical and electrical environment of cells in vivo. Front. Cell Neurosci. **12** (2018). https://doi.org/10.3389/conf.fncel.2018.38.00069
49. Tomczak, K., Czerwinska, P., Wiznerowicz, M.: Review the cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp. Oncol. (Pozn) **19**, A68–A77 (2015). https://doi.org/10.5114/wo.2014.47136
50. Chang, K., et al.: The cancer genome atlas pan-cancer analysis project. Nat. Genet. **45**, 1113–1120 (2013). https://doi.org/10.1038/ng.2764
51. Grossman, R., et al.: Toward a shared vision for cancer genomic data. New. Engl. J. Med. **375**, 1109–1112 (2016). https://doi.org/10.1056/NEJMp1607591
52. Cunningham, F., et al.: Ensembl 2022. Nucleic Acids Res. **50**(D1), D988–D995 (2022). https://doi.org/10.1093/nar/gkab1049
53. Hunt, S.E., et al.: Ensembl variation resources. Database, **2018**, bay119 (2018). https://doi.org/10.1093/database/bay119

54. Wang, M., et al.: PaxDb, a database of protein abundance averages across all three domains of life. Mol. Cell. Proteomics **11**, 492–500 (2012). https://doi.org/10.1074/mcp.O111.014704
55. Avsec, Ž, et al.: Effective gene expression prediction from sequence by integrating long-range interactions. Nat. Methods **18**, 1196–1203 (2021). https://doi.org/10.1038/s41592-021-01252-x
56. Jaganathan, K., et al.: Predicting splicing from primary sequence with deep learning. Cell **176**, 535–548 (2019). https://doi.org/10.1016/j.cell.2018.12.015
57. Hofacker, I., et al.: Automatic detection of conserved RNA structure elements in complete RNA virus genomes. Nucleic Acids Res. **26**, 3825–3836 (1998). https://doi.org/10.1093/nar/26.16.3825
58. Sharp, P., Li, W.-H.: The codon adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**, 1281–1295 (1987). https://doi.org/10.1093/nar/15.3.1281
59. Nakamura, Y., Gojobori, T., Ikemura, T.: Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. **28**(1), 292 (2000). https://doi.org/10.1093/nar/28.1.292
60. Hernandez-Alias, X., Benisty, H., Schaefer, M.H., Serrano, L.: Translational adaptation of human viruses to the tissues they infect. Cell Rep. **34**(11), 108872 (2021). https://doi.org/10.1016/j.celrep.2021.108872
61. Sabi, R., Tuller, T.: Modelling the efficiency of codon–tRNA interactions based on codon usage bias. DNA Res. **21**(5), 511–526 (2014). https://doi.org/10.1093/dnares/dsu017
62. Yu, C.-H., et al.: Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. Mol. Cell **59**(5), 744–754 (2015). https://doi.org/10.1016/j.molcel.2015.07.018
63. Sievers, F., et al.: Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Mol. Syst. Biol. **7**, 539 (2011). https://doi.org/10.1038/msb.2011.75
64. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. In: Kotz, S., Johnson, N.L. (eds.) Breakthroughs in Statistics. Springer Series in Statistics, pp. 319–337. Springer, New York (1992). https://doi.org/10.1007/978-1-4612-4380-9_25
65. Mantel, N.: Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother. Rep. 50(3), 163–170 (1966). http://europepmc.org/abstract/MED/5910392
66. Karson, M.: Handbook of Methods of Applied Statistics. Volume I: Techniques of Computation Descriptive Methods, and Statistical Inference. Volume II: Planning of Surveys and Experiments. I. M. Chakravarti, R. G. Laha, and J. Roy, New York, John Wiley; 1967, $9.00. J Am. Stat. Assoc. 63(323), 1047–1049 (1968). https://doi.org/10.1080/01621459.1968.11009335
67. Gao, H., et al.: Clustering cancers by shared transcriptional risk reveals novel targets for cancer therapy. Mol. Cancer **21**(1), 116 (2022). https://doi.org/10.1186/s12943-022-01592-y
68. Kudla, G., Murray, A.W., Tollervey, D., Plotkin, J.B.: Coding-sequence determinants of gene expression in Escherichia coli. Science **324**(5924), 255–258 (2009). https://doi.org/10.1126/science.1170160
69. Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., Garrels, J.I.: A sampling of the yeast proteome. Mol. Cell Biol. **19**(11), 7357–7368 (1999). https://doi.org/10.1128/MCB.19.11.7357
70. Dos Reis, M., Wernisch, L., Savva, R.: Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. Nucleic Acids Res. **31**(23), 6976–6985 (2003). https://doi.org/10.1093/nar/gkg897

71. Waldman, Y.Y., Tuller, T., Shlomi, T., Sharan, R., Ruppin, E.: Translation efficiency in humans: tissue specificity, global optimization and differences between developmental stages. Nucleic Acids Res. **38**(9), 2964–2974 (2010). https://doi.org/10.1093/nar/gkq009
72. Hardesty, B., Tsalkova, T., Kramer, G.: Co-translational folding. Curr. Opin. Struct. Biol. **9**(1), 111–114 (1999). https://doi.org/10.1016/S0959-440X(99)80014-1
73. Binkhathlan, Z., Lavasanifar, A.: P-glycoprotein inhibition as a therapeutic approach for overcoming multidrug resistance in cancer: current status and future perspectives. Curr. Cancer Drug Targets **13**(3), 326–346 (2013). https://doi.org/10.2174/15680096113139990076
74. Johnatty, S.E., et al.: ABCB1 (MDR1) polymorphisms and ovarian cancer progression and survival: a comprehensive analysis from the ovarian cancer association consortium and the cancer genome atlas. Gynecol. Oncol. **131**(1), 8–14 (2013). https://doi.org/10.1016/j.ygyno.2013.07.107
75. Chen, Q., et al.: Prognostic value of two polymorphisms, rs1045642 and rs1128503, in ABCB1 following taxane-based chemotherapy: a meta-analysis. Asian Pac. J. Cancer Prev. **22**(1), 3 (2021). https://doi.org/10.31557/APJCP.2021.22.1.3