Population variability of rhesus macaque (*Macaca mulatta*)

*NAT1* gene for arylamine N-acetyltransferase 1: Functional effects and comparison with human

Sotiria Boukouvala1, Zoe Chasapopoulou1, Despina Giannouri2, Evanthia Kontomina1, Nikolaos Marinakis1, Sophia V. Rizou1, Ioanna Stefani1, Theodora Tsirka1, Charlotte Veyssière2, Sofia Zaliou1,4, Audrey Sabbagh1, Brigitte Crouau-Roy2 & Giannoulis Fakis1

Human *NAT1* gene for N-acetyltransferase 1 modulates xenobiotic metabolism of arylamine drugs and mutagens. Beyond pharmacogenetics, *NAT1* is also relevant to breast cancer. The population history of human *NAT1* suggests evolution through purifying selection, but it is unclear whether this pattern is evident in other primate lineages where population studies are scarce. We report *NAT1* polymorphism in 25 rhesus macaques (*Macaca mulatta*) and describe the haplotypic and functional characteristics of 12 variants. Seven non-synonymous single nucleotide variations (SNVs) were identified and experimentally demonstrated to compromise enzyme function, mainly through destabilization of *NAT1* protein and consequent activity loss. One non-synonymous SNV (c.560G > A, p.Arg187Gln) has also been characterized for human *NAT1* with similar effects. Population haplotypic and functional variability of rhesus *NAT1* was considerably higher than previously reported for its human orthologue, suggesting different environmental pressures in the two lineages. Known functional elements downstream of human *NAT1* were also differentiated in rhesus macaque and other primates. Xenobiotic metabolizing enzymes play roles beyond mere protection from exogenous chemicals. Therefore, any link to disease, particularly carcinogenesis, may be via modulation of xenobiotic mutagenicity or more subtle interference with cell physiology. Comparative analyses add the evolutionary dimension to such investigations, assessing functional conservation/diversification among primates.

The *NAT1* gene (HGNC ID: 7645) for human arylamine N-acetyltransferase 1 (UniProt ID: P18440; E.C. 2.3.1.5) was originally characterized for its role in the metabolism of xenobiotics, particularly aromatic amine drugs and mutagens. Apart from its pharmacogenetic significance, *NAT1* has more recently been implicated in cell systems relevant to carcinogenesis1–4. The human *NAT1* gene, and its mouse orthologue *Nat2*, begin their expression during preimplantation development4–6 and remain transcriptionally active in most embryonic and adult tissues4,6–11. Both orthologues are controlled by a conserved Sp1-type housekeeping promoter, adjacent to an upstream non-coding exon11–13, and are potentially subject to epigenetic regulation14,15. Human *NAT1* is also controlled by a second, more tissue-specific promoter16, and regulation has been shown to implicate steroid hormones17–20, micro-RNAs21,22, protein polyubiquitinylation23,24, as well as acetylation of the enzyme and allosteric modulation by adenosine triphosphate25,26. Human NAT1 and rodent NAT2 proteins have been demonstrated to acetylate the same arylamine substrates27, including p-aminobenzoyleglutamate (pABGlu), a catabolic derivative of folate27–30.

---

1Democritus University of Thrace, Department of Molecular Biology and Genetics, Alexandroupolis, Greece. 2CNRS, Université P. Sabatier, IRD, UMR5174 EDB lab (Laboratoire Evolution & Diversité Biologique), Toulouse, France. 3UMR 261 MERIT IRD, Faculté de Pharmacie de Paris, Université Paris Descartes, Sorbonne Paris Cité, Paris, France. 4Present address: Medicom Hellas S.A., Gerakas Attikis, Greece. Correspondence and requests for materials should be addressed to G.F. (email: gfakis@mbg.duth.gr)
NAT1 carried more than one nucleotide variation compared to (MACMU)NAT1 (Reference). A total of 12 polymorphic sites were encountered in the sample studied (50 chromosomes analysed in total). Assigned allelic symbols and GenBank accession numbers (Nucleotide IDs) are provided for each haplotype. Estimated haplotype diversity (Hd) in the population was 0.817.

Table 1. (MACMU)NAT1 haplotypes and corresponding amino acid changes. Polymorphic sites are shown, numbered relative to either the nucleotide sequence of (MACMU)NAT1 intronless coding region (top part) or the deduced amino acid sequence of (MACMU)NAT1 protein (bottom part), where position c.1 is the adenosine of the translation initiation codon (ATG) and p.1 is the first amino acid (Met) of the polypeptide chain, respectively. Non-synonymous substitutions are shown in bold. N is the number of times each haplotype was encountered in the sample studied (50 chromosomes analysed in total). Assigned allelic symbols and GenBank accession numbers (Nucleotide IDs) are provided for each haplotype.

Results

Variability in the rhesus macaque (MACMU)NAT1 gene. The intronless coding region of (MACMU)NAT1 gene was sequenced for 25 rhesus macaque individuals. A total of 12 polymorphic sites were identified, of which 5 represented synonymous and 7 non-synonymous nucleotide substitutions relative to reference allele (MACMU)NAT1*1 (Nucleotide ID: KU640969.1). The determined genotypes (Supplementary Table S1) corresponded to 12 haplotypes (Table 1), which were assigned allelic symbols (MACMU)NAT1*2 – (MACMU)NAT1*12 (Nucleotide IDs: KU640968 – KU640996). Allele (MACMU)NAT1*1 showed highest identity (96.45%) to (HUMAN)NAT1*4 reference allele (Nucleotide ID: AJ307007.1) and was also most frequently encountered in the rhesus macaque sample screened, representing 34% of the 50 haplotypes analysed in total. Other frequently found haplotypes were (MACMU)NAT1*2 (18%) and (MACMU)NAT1*5 (20%), while the remaining haplotypes were less common (2–6%). With the exception of (MACMU)NAT1*3 and *8, all haplotypes carried more than one nucleotide variation compared to (MACMU)NAT1*1. Haplotypes (MACMU)NAT1*3, *5, *8, *9 and *12 contained only synonymous substitutions. Haplotypes (MACMU)NAT1*2, *4, *6, *7 and *11
Figure 1. Genomic variability downstream of NAT1 coding region in primates. A typical Alu/SINE element is found about 10 bp after the stop codon of NAT1 gene in Old World monkeys (11 species, including the rhesus macaque), but is absent in apes (6 species, including human). Moreover, a microsatellite (TAA) repeat element, spanning one consensus polyadenylation signal (polyA-1) of (HUMAN)NAT1 gene, is 12–18 bp shorter in all 16 non-human primate genomes analysed, resembling (HUMAN)NAT1*11 (9 bp shorter) and other polymorphic NAT1 alleles reported in human populations (http://nat.mbg.duth.gr/). Also shown is SNV rs1057126, which affects polyA-1 signal in (HUMAN)NAT1*10 and other reported human alleles. The figure was not drawn to scale. Detailed illustrations of this particular genomic region are provided in Supplementary Figs S1 and S2.

contained one non-synonymous, together with one or two synonymous substitutions. (MACMU)NAT1*10 allele carried three non-synonymous and one synonymous substitution (Table 1).

Of the SNVs found (Tables 1 and S1), synonymous substitution c.321C>T was most frequent (52.5%) and altered the codon for key catalytic residue p.His107, which however remained conserved. Synonymous SNV c.540C>T was also common (20.3%), while other synonymous SNVs showed lower frequencies (up to 5%). Of the non-synonymous SNVs (Tables 1 and S1), c.152G>C (p.Gly51Ala), c.244A>G (p.Asp81Val), c.267G>C (p.Leu89Phe), c.463G>C (p.Glu155Gln) and c.523T>C (p.Phe175Leu) were encountered only once in the rhesus macaque sample. Non-synonymous substitution of amino acid residue p.Gly51 (c.152G>T, p.Gly51Val; rs72466547:G>T) has been reported previously for the *6M allele of (HUMAN)NAT2 gene (the 87.4% conserved NAT1 paralogue in human) and has been predicted to potentially affect substrate binding. Substitution c.343G>T (p.Asp115Tyr) was found in two heterozygous rhesus macaque individuals carrying haplotypes NAT1*10 and *11. The most frequent (5%) non-synonymous SNV in the rhesus macaque sample was c.560G>A (p.Arg187Gln), found in three heterozygous individuals (Tables 1 and S1). This exact SNV (rs4986782:G>A) has been reported for (HUMAN)NAT1*14, with 100 Genomes Project phase 3 global minor allele frequency (MAF) of 0.6%, and has important functional consequences. The same codon is also subject to another mutation (c.559C>T, p.Arg187Ter; rs5030893:C>T with global MAF of 0.3%) in the (HUMAN)NAT1*15 allele, causing premature termination of the polypeptide chain.

Although non-synonymous SNVs have been reported in the literature for (HUMAN)NAT1 gene (http://nat.mbg.duth.gr/), the most common polymorphic alleles in human populations are characterized by functional variability in the 3'-untranslated region (3'-UTR). Allele (HUMAN)NAT1*10 is defined by rs1057126:T>A (c.*215T>A) which affects one active polyadenylation (polyA) signal of the gene. Allele (HUMAN)NAT1*11 is defined by a shortage of three TAA repeats in a microsatellite element spanning the same polyA signal. This particular region is highly variable in human populations (http://nat.mbg.duth.gr/), and has been demonstrated to modulate polyA usage of the gene and mRNA translation efficiency. We show here that the rhesus macaque and other non-human primates (both Old World monkeys and apes) also carry a shorter microsatellite stretch at this position, suggesting an expansion of the ancestral wild-type allele specific to the human evolutionary lineage (Figs 1 and S1). Moreover, the genomic distance of this particular microsatellite/polyA region from the stop codon of NAT1 gene is about 300 bp longer in the rhesus macaque and other Old World monkeys, compared with human and other apes. This difference in length is due to an Alu/SINE element which is present immediately downstream of NAT1 gene in Old World monkeys, but not in apes (Figs 1 and S2). These notable differences between human and other primates suggest substantial evolutionary and functional divergence of the NAT1 3'-UTR that merits further investigation, especially given the phenotypic significance attributed to (HUMAN)NAT1*10 and *11 alleles and their investigated association with disease.

Effects of polymorphisms on rhesus macaque (MACMU)NAT1 protein expression. A clone carrying the (MACMU)NAT1*1 reference allele in recombinant expression vector was used to generate a series of mutagenised constructs, each carrying one of the seven non-synonymous SNVs found in the rhesus macaque sample screened. One additional construct was engineered to carry both c.267G>C (p.Leu89Phe) and c.343G>T (p.Asp115Tyr), as those two SNVs were found together on the same haplotype (Supplementary Table S2).

As previously observed, recombinant expression/purification of (MACMU)NAT1_1 reference protein was always efficient, generating preparations of high yield and purity (Fig. 2). Variant (MACMU)NAT1_p.Arg187Gln also generated sufficient yields of highly pure protein, although at considerably lower levels (by 30% on average) compared with (MACMU)NAT1_1 (Fig. 2). Recombinant expression of all other variants generated poor quality preparations that made recovery of pure proteins virtually impossible, although the expected NAT1 protein band was always visible on gels (Fig. 2). The recombinant preparations of (MACMU)NAT1_1 and (MACMU) NAT1_p.Arg187Gln were in fact differentiated from those of all other variants, throughout the entire expression/purification process (Supplementary Fig. S3). Those results were reproducible upon expression of recombinant (MACMU)NAT1 protein variants multiple (up to five) times and by different (up to three) persons, over an extended period of time. Procedures always incorporated (MACMU)NAT1_1, as reference protein, and followed a well-standardized protocol in our laboratory, that has been successfully applied to numerous recombinant
NAT homologues across a wide taxonomic spectrum of organisms, including the rhesus macaque and other primates. Our observations were consistent with protein aggregation involving misfolded recombinant (MACMU)NAT1 unstable variants, both during expression (in the form of bacterial inclusion bodies), as well as in cell lysates and during purification (in the form of soluble and insoluble complexes with native \textit{Escherichia coli} proteins)\textsuperscript{75–77}. Although a variety of approaches are available for overcoming recombinant protein aggregation\textsuperscript{75}, optimizing the expression/purification conditions for individual (MACMU)NAT1 variants was beyond the scope of this study, which focused on comparison of SNV effects using a standard procedure widely employed by NAT researchers performing similar work\textsuperscript{27,35,64,78–87}. As expression of mammalian proteins in \textit{E. coli} could potentially confound the observed functional impact of different SNVs, future investigations of the characterized (MACMU)NAT1 variants could additionally employ eukaryotic cells. However, previous expression of (HUMAN)NAT1 and its variants in standard eukaryotic systems has been demonstrated to generate recombinant protein in very low yield that is barely detectable\textsuperscript{70,71}.

To verify the presence of recombinant (MACMU)NAT1 protein in preparations of expressed variants, a series of immunoblot analyses were performed, using antibodies specific for the N-terminal, middle or C-terminal part of the polypeptide chain. All preparations were demonstrated to contain recombinant (MACMU)NAT1 protein, both in the soluble fraction of each cell lysate and in the products of subsequent chromatographic purifications (Fig. 3). However, compared with (MACMU)NAT1_1 reference protein, the amount of detected variants was substantially decreased (typically by more than 50%) in recombinant preparations, and this result was consistent across different experiments and with all three antibodies. As expected, (MACMU)NAT1_p.Arg187Gln appeared to be only moderately affected, relative to (MACMU)NAT1_1 reference protein, demonstrating up to 45% reduction in the amount of recombinant protein detected in purified chromatographic fractions. By contrast,
the amount of immunodetected (MACMU)NAT1_p.Arg187Gln protein was similar to (MACMU)NAT1_reference protein in soluble fractions generated immediately after cell lysis (Fig. 3).

The combination of two amino acid substitutions in the same polypeptide chain had an additive, essentially detrimental, impact on the recombinant expression of (MACMU)NAT1_p.Leu89Phe/p.Asp115Tyr variant. The double mutant was barely detectable by immunoblot analyses, even in freshly prepared cell lysates following recombinant expression (Supplementary Fig. S4).

Overall, these results indicate that (MACMU)NAT1 variants are inherently unstable, when expressed in E.coli, suggesting that different non-synonymous substitutions may compromise protein folding and solubility. This effect is relatively moderate in the case of (MACMU)NAT1_p.Arg187Gln variant and most pronounced in the case of double mutant (MACMU)NAT1_p.Leu89Phe/p.Asp115Tyr. Bacteria possess complex systems to assess correct folding of expressed proteins and can reject misfolded proteins through proteolysis or the formation of insoluble aggregates. Although application of more sophisticated methods for detection and monitoring of protein aggregate formation may enable more accurate assessments in the future, the present work demonstrates that the likely impact of characterized non-synonymous SNVs is to compromise the inherent stability of (MACMU)NAT1 protein, as summarized in Table 2. This is in line with previous investigations of (HUMAN)NAT1, demonstrating a similar effect on protein stability for most characterized SNVs, irrespectively of whether the expression was carried out in prokaryotic (E. coli) or eukaryotic (yeast or mammalian) cells.

Effects of polymorphisms on rhesus macaque (MACMU)NAT1 enzyme activity. The recombinant preparations of (MACMU)NAT1_1 reference protein and its eight variants were assayed for enzymatic activity, using the arylamines p-aminobenzoic acid (PABA), 5-aminosalicylic acid (5AS), p-anisidine (PANS) and sulphonamethazine (SMZ) as acetyl-group acceptor substrates. SMZ is a NAT2-selective substrate that was expected to provide only marginal activity with (MACMU)NAT1 protein, unless any of the assayed variants caused a switch in substrate specificity of the isoenzyme, as has been reported previously for the p.Val231Ile polymorphism of (MACMU)NAT2. No such effect was demonstrated for any of the (MACMU)NAT1 polymorphisms tested in the present study (Fig. 4).
Because of the apparent instability of (MACMU)NAT1 variants, and the consequently unavoidable variability in the amount and purity of recovered recombinant proteins across different preparations (Table 2), a 50% threshold was used to interpret the effects of polymorphisms on enzyme activity, relative to (MACMU)NAT1_1 reference protein which was included in each set of assays. Enzymatic activity was detected for all variants, when PABA, 5AS or PANS were used as substrates, but was decreased relative to (MACMU)NAT1_1 (Fig. 4). The smallest change was observed for variant (MACMU)NAT1_p.Glu155Gln, which was >90% active relative to (MACMU)NAT1_1, and a 50% threshold was used to assess the overall impact of various SNVs on protein stability. The amount of (MACMU)NAT1_p.Arg187Gln protein showed a moderate decrease in the 100 mM imidazol fraction eluted during chromatographic purification, but such decrease was not evident when immunodetection of the same recombinant variant took place in soluble fractions generated immediately after cell lysis. A 50% enzymatic activity threshold, relative to the (MACMU)NAT1_1 reference protein, was used to assess the effect of each polymorphism as either “moderate” or “substantial”. The average value and corresponding range of relative (%) specific activities measured with PABA, 5AS and PANS are provided in parentheses. Variant (MACMU)NAT1_p.Glu155Gln provided a broad range of activities, depending on which substrate was used, and the overall effect was assessed as ‘marginal/moderate’. The predicted effect of each polymorphism on (MACMU)NAT1 protein structure/function was assessed by structural modelling and PolyPhen-2 software.

Table 2. Summary of effects of non-synonymous polymorphisms on (MACMU)NAT1 enzyme function. The effect of polymorphisms on stability of (MACMU)NAT1 variants was assessed based on observations during recombinant expression/purification processes, the appearance of recombinant proteins on SDS-PAGE gels and immunoblots, as well as on thermal stability assays performed where possible. In parentheses, the average amount and corresponding amount range of immunodetected proteins is provided relative (%) to (MACMU)NAT1_1, and a 50% threshold was used to assess the overall impact of various SNVs on protein stability. The amount of (MACMU)NAT1_p.Arg187Gln protein showed a moderate decrease in the 100 mM imidazol fraction eluted during chromatographic purification, but such decrease was not evident when immunodetection of the same recombinant variant took place in soluble fractions generated immediately after cell lysis. A 50% enzymatic activity threshold, relative to the (MACMU)NAT1_1 reference protein, was used to assess the effect of each polymorphism as either “moderate” or “substantial”. The average value and corresponding range of relative (%) specific activities measured with PABA, 5AS and PANS are provided in parentheses. Variant (MACMU)NAT1_p.Glu155Gln provided a broad range of activities, depending on which substrate was used, and the overall effect was assessed as ‘marginal/moderate’. The predicted effect of each polymorphism on (MACMU)NAT1 protein structure/function was assessed by structural modelling and PolyPhen-2 software.

| Recombinant variant | Observed effect on stabilitya | Observed effect on enzymatic activityb | Predicted effect on structurec | Predicted effect by PolyPhen-2d |
|---------------------|-------------------------------|-------------------------------------|-------------------------------|-------------------------------|
| (MACMU)NAT1_1       | Reference                      | Reference                           | Reference                     | Reference                     |
| (MACMU)NAT1_p.Gly51Ala | Substantial decrease (40%; 33–51%) | Moderate decrease (55%; 45–69%) | Benign (residue on the surface of the protein molecule) | Benign (score 0.118) |
| (MACMU)NAT1_p.Met82Val | Substantial decrease (32%; 20–49%) | Moderate decrease (68%; 56–81%) | Benign (residue on the surface of the protein molecule) | Benign (score 0.128) |
| (MACMU)NAT1_p.Leu89Phe | Substantial decrease (44%; 24–65%) | Substantial decrease (31%; 25–44%) | Damaging (residue stabilizes structural conformation important for enzyme function) | Probably damaging (score 1.000) |
| (MACMU)NAT1_p.Asp115Tyr | Substantial decrease (32%; 25–39%) | Substantial decrease (18%; 11–26%) | Damaging (residue stabilizes structural conformation important for enzyme function) | Possibly damaging (score 0.845) |
| (MACMU)NAT1_p.Leu89Phe/p.Asp115Tyr | Barely detectable (<5%) | Residual activity (<5%) | Damaging (combined effect of two compromising changes) | N/A |
| (MACMU)NAT1_p.Glu155Gln | Substantial decrease (38%; 16–68%) | Marginal/moderate decrease (78%; 44–100%) | Benign (residue on the surface of the protein molecule) | Benign (score 0.009) |
| (MACMU)NAT1_p.Phe175Leu | Substantial decrease (50%; 11–73%) | Substantial decrease (18%; 15–20%) | Damaging (residue stabilizes structural conformation important for enzyme function) | Possibly damaging (score 0.545) |
| (MACMU)NAT1_p.Arg187Gln | Marginal/moderate decrease (100%) vs. (73%; 56–82%) | Substantial decrease (27%; 20–36%) | Damaging (residue interacts with components important for enzymatic function) | Benign (score 0.444) |

Because of the apparent instability of (MACMU)NAT1 variants, and the consequently unavoidable variability in the amount and purity of recovered recombinant proteins across different preparations (Table 2), a 50% threshold was used to interpret the effects of polymorphisms on enzyme activity, relative to (MACMU)NAT1_1 reference protein which was included in each set of assays. Enzymatic activity was detected for all variants, when PABA, 5AS or PANS were used as substrates, but was decreased relative to (MACMU)NAT1_1 (Fig. 4). The smallest change was observed for variant (MACMU)NAT1_p.Glu155Gln, which was >90% active relative to (MACMU)NAT1_1, when PABA and 5AS were used. However, a 55% decrease in activity was observed with PANS, suggesting that the effects of the polymorphism could be substrate-specific: PABA and 5AS are more NAT1-selective, compared with PANS which is a substrate for both NAT1 and NAT2 isoenzymes of primates. Most severely compromised was the double mutant (MACMU)NAT1_p.Leu89Phe/p.Asp115Tyr, demonstrating some activity only with PABA. Variants (MACMU)NAT1_p.Gly51Ala and (MACMU)NAT1_p.Met82Val demonstrated moderate decrease (<50%) in enzymatic activity relative to (MACMU)NAT1_1, while the decrease was more substantial (>50%) in the case of variants (MACMU)NAT1_p.Leu89Phe, (MACMU)NAT1_p.Asp115Tyr, (MACMU)NAT1_p.Phe175Leu and (MACMU)NAT1_p.Arg187Gln, as summarized in Table 2.

Initial attempts to assess the impact of substitutions on thermal stability of (MACMU)NAT1 protein, through measurement of enzymatic activity after exposure of recombinant preparations to temperatures from 35 to 45 °C, were complicated by the fact that most variants demonstrated very low enzymatic activities even before thermal stress. In those cases where the results were assessed as reliable, thermal stability of variants appeared to decline at different rates (Fig. 5). Loss of activity was marginal for (MACMU)NAT1_1 reference protein at temperatures close to its reported denaturation midpoint temperature (36.5 ± 1.2 °C)\(^\text{29}\), which is also similar to the body temperature (37.3 °C) of the animal (Animal Ageing and Longevity Database, http://genomics.senescence.info/species)\(^\text{30}\), but became more evident above 39.5 °C. Variant (MACMU)NAT1_p.Leu89Phe appeared rather resilient to thermal stress, losing activity at similar rate as the reference protein, despite its demonstrated instability during expression/purification. Other variants were more sensitive to the effects of thermal stress (Fig. 5). To overcome the aforementioned limitations, we also tried differential scanning fluorimetry, a well-standardized method in our laboratory\(^\text{23}\), to assess thermal stability of (MACMU)NAT1 variants. However, with the exception of the reference protein, this approach was hampered by the insufficient purity of recombinant preparations. In the future, it may be useful to try alternative thermal denaturation assays, described in the literature\(^\text{24}\).
Combined, the observations of this and the previous section indicate inherent instability of (MACMU)NAT1 variants as the likely cause for reduced enzymatic activity. However, an effect on catalytic function may not be ruled out, particularly for variants with polymorphisms p.Asp115Tyr and p.Phe175Leu that demonstrated a decrease in enzymatic activity that was considerably more substantial relative to the decrease in the amount of recovered recombinant protein (Table 2). Impairment of catalytic function is also likely to be the main cause for the observed decrease in the enzymatic activity of variant (MACMU)NAT1_p.Arg187Gln, which appeared relatively stable compared with other variants in this study. Polymorphism p.Arg187Gln has also been demonstrated to cause decrease in activity of (HUMAN)NAT1 homologue, both when using human biological material and after recombinant expression in bacterial, yeast or mammalian cell systems68–71,92–94. As in the present study for (MACMU)NAT1, the effect of p.Arg187Gln on (HUMAN)NAT1 appeared to be in the form of moderate decrease in protein stability, combined with impairment of enzymatic function68,70,71. More sophisticated kinetic analyses with a range of substrates may help elucidate the exact mechanism underlying functional impairment of each characterized (MACMU)NAT1 variant, at least in those cases where stability and activity are sufficient to allow reliable enzymatic assays.

Structural effects of rhesus macaque (MACMU)NAT1 polymorphisms. Structural alignment and homology modelling of (MACMU)NAT1 protein against the published structure of (HUMAN)NAT164 was used to assess the potential functional effects of rhesus macaque polymorphisms at the molecular level (Figs 6 and S5). These predictions were further compared with the output of PolyPhen-2, a computational algorithm employing comparative sequence alignments and structural analyses to determine the impact of amino acid substitutions on proteins (Table 2).

Residue p.Gly51 is located within the α-helical domain I on the surface of the protein molecule, right before the start of α3-helix and behind p.Cys68 of the catalytic triad. No direct interactions of p.Gly51 are evident and its substitution to Ala is conservative (Figs 6a and S5). Therefore, the postulated effect is benign, which is in line with the prediction of PolyPhen-2 assigning a low impact probability score to this particular polymorphism (Table 2). Despite those computational predictions, the experimental results show p.Gly51Ala to have a moderate effect on (MACMU)NAT1 enzyme activity, presumably caused by reduction in intrinsic stability. The conformational flexibility of glycine makes the amino acid important for maintaining certain structural arrangements (e.g. tight turns), so its substitution could potentially affect protein stability95.

A similar impact can be attributed to polymorphism p.Met82Val which also affects a residue of domain I and is located on the surface of the protein molecule. Amino acid p.Met82 is the last amino acid on α4-helix, which has catalytic p.Cys68 as its first residue (Figs 6b and S5). Substitution from Met to Val is conservative and predicted as benign by PolyPhen-2 (Table 2). The experimental results, however, indicate a moderate effect on enzymatic activity, potentially caused by reduction in intrinsic stability of the protein. Valine has certain restrictions in adopting the α-helical conformation and is hydrophobic, usually found in the interior of protein molecules95. Such factors could contribute to the observed instability of variant (MACMU)NAT1_p.Met82Val.
Residue p.Leu89 is located inside the enzymatic core, on β2-strand of β-barrel domain II of the protein molecule. Substitution to Phe is predicted by PolyPhen-2 as probably damaging, with the highest score (Table 2). The same tool identified p.Leu89 as essentially conserved across NAT1 homologues in the UniProt database. The residue is located directly behind β4-strand carrying p.His107 of the catalytic triad and its adjacent p.Le106 that is important for binding of NAT1-selective aminobenzyl substrates, like p-aminosalicylic acid and PABA. Structural modelling (Fig. 6c) supports that replacement of p.Leu89 by the much bulkier Phe causes a direct steric clash with β4-strand, leading to molecule destabilization and impairment of its catalytic function, consistent with the experimental observations of the previous sections.

Substitution of p.Asp115 by the bulkier, uncharged and more hydrophobic Tyr represents a non-conservative change that was predicted by PolyPhen-2 as possibly damaging to (MACMU)NAT1 (Table 2). Amino acid p.Asp115 is located on the surface of the protein molecule and is part of domain II (Figs 6d and S5). This residue is positioned right at the turn between anti-parallel strands β4 and β5, which determine the positioning of p.His107 and p.Asp122 in the catalytic triad. The conformation of the two β-strands is stabilized via bonding between p.Asp115 and p.Arg117. As described in the previous sections, polymorphism p.Asp115Tyr appeared to compromise the protein molecule in terms of stability and possibly catalytic function. Moreover, the combination of this polymorphism with p.Leu89Phe on the same (MACMU)NAT1 variant was demonstrated to be effectively detrimental. Apart from those two polymorphisms, haplotype (MACMU)NAT1*10 also contained a third SNV causing substitution p.Glu155Gln. However, given the severe impairment of double mutant (MACMU)NAT1_p.Leu89Phe/p.Asp115Tyr, the potential additive effect of the more benign p.Glu155Gln substitution was not examined.

Residue p.Glu155 is the last amino acid on β8-strand of domain II, which is connected to its anti-parallel β9-strand via a sharp turn. The location is away from the catalytic pocket and on the surface of the enzyme molecule (Figs 6e and S5). Its substitution to Gln is conservative and has no apparent consequences. PolyPhen-2 also predicted this change as benign, assigning a very low score (Table 2). Although our experiments detected no substantial impact on enzymatic activity, an apparent decrease in stability of variant (MACMU)NAT1_p.Glu155Gln was nevertheless observed. The underlying cause of this effect is unclear, although p.Glu155 could stabilize the conformation of anti-parallel strands β8-β9, which are part of a β-sheet forming a “wall” on one side of the active site pocket.

Residue p.Phe175 is one of 17 amino acids forming the so-called “eukaryotic loop” in domain II of mammalian NATs. This component is thought to stabilize an important β-sheet of domain III, comprising four anti-parallel β-strands that form a “wall” at the bottom of the active site pocket. This stabilization is conferred via parallel alignment of the loop with the fourth strand (β15) of the β-sheet and, specifically, via “stacking” interaction of the aromatic rings of p.Phe175 on the loop with p.Phe255 on the β15-strand (Fig. 6f). The structural model predicts replacement of p.Phe175 by Leu to be damaging to this interaction and, consequently, to the protein molecule itself, both in terms of stability and potentially enzymatic function. PolyPhen-2 identified p.Phe175 as very conserved across NAT homologues and predicted its substitution as possibly damaging (Table 2). The in silico results are consistent with the experimental observations of the previous sections.

The final polymorphism in this study, p.Arg187Gln, has been structurally modelled before, as it is also found in (HUMAN)NAT1*4,8,96,97. As shown (Figs 6g and S5), residue p.Arg187 is located on β11-strand of domain II and forms hydrogen bond interactions with p.Lys188 and p.Glu182. The latter amino acid is located on the
The aforementioned “eukaryotic loop” of the enzyme and interacts via hydrogen bonding with two important residues, p.Tyr129 and p.Thr289. Residue p.Tyr129 is key for NAT1 acceptor substrate selectivity and also interacts with p.Thr289, which is the residue located just before the carboxyl terminus of the 290 amino acid polypeptide chain. This network of interactions maintains the shape of the active site pocket and of the “eukaryotic loop”, while also stabilizing the carboxyl terminus in a position deep within the catalytic core of the enzyme that enables contact with the donor substrate. Although the rather conservative substitution of p.Arg187 to Gln could potentially maintain some of the required hydrogen bond interactions, instability in the polypeptide chain folding would be unavoidable, in line with the moderate loss of protein observed during recombinant expression/purification. Moreover, the expected change in the conformation of the active site pocket would explain the substantial decrease in enzymatic activity observed for variant p.Arg187Gln both in the rhesus macaque and human. Contrary to the solid experimental evidence, PolyPhen-2 predicted this polymorphism as benign, although the score was close to 0.5 (Table 2). Despite their value in guiding the assessment of SNV effects on protein function, computational predictions must always be interpreted with caution, as they may not be fully aligned with experimental evidence.

Discussion

The evolution of xenobiotic metabolizing enzymes is shaped by the ever-changing chemical environments organisms need to adapt to in order to survive. Genetic polymorphisms can be favourable, in that respect, as they confer a range of metabolic activities towards a plethora of potentially harmful xenobiotic substances. Moreover, population frequencies of high- vs. low-activity enzyme variants may shift to one or the other direction, depending on the prevailing xenobiotic challenges in a specific environment. In humans, this process is driven not only by natural
selective pressures, but also by the lifestyle choices of our species. The NAT enzymes metabolize toxic aromatic amines found in cooked food and other byproducts of everyday human activity since prehistoric times. Variability in NAT genes has thus been studied in the context of understanding how subsistence patterns may have affected the genetic makeup of human populations. Such knowledge is useful for elucidating how even apparently modest differentiations in gene-environment interactions may influence disease susceptibility or drug response.

Comparison of humans with other primates can further discriminate between variability that is ancestral or specific to our own species. Although previous phylogenetic analyses have demonstrated a dynamic pattern consistent with adaptive evolution of NAT genes in primates, those studies have been limited by their entirely in silico methodology and the fact that they have examined NAT variability only between species. A new study, investigating population variability at NAT1, NAT2 and NATP loci across different hominin species, demonstrated opposite diversity patterns for the chimpanzee (Pan troglodytes) and bonobo (P. paniscus) compared with human. Genetic diversity was higher for NAT1 in the two Pan species and lower in human, with the opposite pattern observed for the NAT2 gene. Here, we investigated NAT1 variability in a population sample of the Old World monkey rhesus macaque, also characterizing the possible functional consequences of detected variants. Our results are in accordance with the aforementioned study, indicating high haplotypic diversity (Hd = 0.817) of (MACMU)NAT1 coding region for the specific rhesus macaque population studied, and with 16% of the haplotypes carrying at least one non-synonymous SNV with some functional impact. It appears that, in the rhesus macaque, polymorphic NAT1 alleles may modulate acetylation of arylamine xenobiotics more effectively compared with human, where population frequencies of non-synonymous SNVs in NAT1 coding region are extremely low (typically <1%, with few exceptions). The low haplotypic diversity (Hd = 0.073) reported for (HUMAN)NAT1 coding region is consistent with the action of purifying selection predicted for the gene. As mentioned above, (HUMAN)NAT1 is considered to have a more fundamental role in cells, unlike its paralogous (HUMAN)NAT2, which is highly polymorphic and exhibits the functional characteristics of a typical xenobiotic metabolizing enzyme.

Although (MACMU)NAT1 is phylogenetically orthologous to (HUMAN)NAT1, its higher degree of variability could be the outcome of environmental pressures differentiating the evolutionary lineages of the two species. However, despite this high level of diversity, all heterozygous (MACMU)NAT1 genotypes were found to contain at least one allele with no predicted phenotypic consequences. Specifically, of the 25 genotypes determined, seventeen contained only functional alleles (i.e. combinations of *1, *3, *5, *8 and *12 encoding for the reference amino acid sequence), while the remaining eight genotypes combined one functional with one compromised (*2, *4, *6, *7, *10, *11) allele. Genotypes consisting of two compromised alleles were not found (Supplementary Table S1). It is possible that rhesus macaque NAT1 variability may be maintained, as far as the corresponding enzymatic function is not entirely abolished in the individual. This could preserve the same endogenous function attributed to (HUMAN)NAT1, although the diversity pattern appears to differ in the evolutionary lineages of the two species. Although such observations are useful, in view of current hypotheses relating (HUMAN)NAT1 with functions beyond xenobiotic metabolism, other factors (potentially related to the small size or the particular makeup of the rhesus macaque population from which our sample was drawn) could also account for the high variability observed.

As described above, (HUMAN)NAT1 is also differentiated from other primate orthologues at the 3′-UTR of the gene. The reference alleleic sequence downstream of the coding region bears a microsatellite repeat element that encompasses an active polyA signal of the gene. Variability in this element is common (~40% global frequency) in human populations, and is encountered both in the form of SNV, e.g. rs1057126:T>A in (HUMAN)NAT1, and in the form of simple sequence length polymorphism, as in (HUMAN)NAT1*11 and other alleles. Although those variations appear to confer some functional effects, their association with disease susceptibility has been controversial. It would be interesting to examine if the functional significance attributed to the 3′-UTR of (HUMAN)NAT1 gene is corroborated by studies with NAT1 orthologues of other primates, given the substantial sequence and structural variability we report here.

Despite those differences, certain NAT1 variants are found in both the rhesus macaque and human. As discussed, NAT1 polymorphism c.560G>A (p.Arg187Gln) was the most prevalent non-synonymous SNV in our rhesus macaque sample and is relatively common in human populations. The functional impact of this polymorphism was similar for the two species, compromising the activity of NAT1 isoenzyme. Other rhesus macaque SNVs of this study have also been reported for (HUMAN)NAT1 gene by large scale genome-wide population studies, like the 100 Genomes Project, including synonymous SNVs at positions c.15 (rs74456038:A>G) and c.321 (rs755240678:C>T), as well as the non-synonymous SNV at position c.523 (rs762246777:T>C; p.Phe175Leu). Moreover, similarly to the characterized (MACMU)NAT1 variants, impaired intrinsic stability appears to be the major defect of low-activity (HUMAN)NAT1 variants, such as NAT1*17 (c.190C>T, p.Arg64Trp; rs56379106:C>T) and *22 (c.752A>T; p.Asp251Val; rs56172717:A>T). Although pinpointing genetic aberrations that are strongly associated with disease is relatively straightforward, it is more difficult to unravel subtle genomic differences that may influence adaptability to exogenous challenges and shape the landscape of factors contributing to complex disorders. Xenobiotic metabolizing enzymes play diverse roles in cells that may not be limited merely to protection of organisms from exogenous substances. Therefore, any link to disease, particularly carcinogenesis, may be either via modulation of the mutagenic potential of xenobiots, or via a less obvious interference with cell physiology, as has long been speculated for human NAT1 and is supported by recent studies mentioned above. Comparative genetic analyses between human and other primates add the evolutionary dimension to such investigations, particularly when combined with experimental insights into functions that may have been either conserved or diversified in the phylogenetic lineages separating our species from its closest relatives. To that end, it would be useful to conduct more comprehensive interdisciplinary investigations into the molecular, structural and enzymological aspects of variants reported for primate NAT homologues.
Methods

Amplification and sequencing of (MACMU)NAT1 gene. For this in vitro study, the 870 bp intronless coding region of (MACMU)NAT1 gene was amplified from genomic DNA samples of a rhesus macaque population described previously 105, 106. PCR was performed with flanking primers 1A (5′-AGCCATAATTAGCCTACTC-3′) and 1AR (5′-GTACAGAAGATACATGATAGG-3′), and the reactions contained 1 μL GoTaq DNA polymerase (Promega) in 1x reaction buffer with 1.5 mM MgCl₂, 0.1 mM each dNTP, 0.1 μM each primer and 25 ng genomic DNA as template. Thermal cycling conditions were as follows: 95 °C (10 min), followed by 38 cycles of 94 °C (1 min), 53 °C (50 s) and 72 °C (2 min), with final extension at 72 °C (10 min). Specific PCR amplification was verified by agarose gel electrophoresis and the products (1350 bp) were purified and sequenced from both directions (Müllingen, Labège, France).

Generation of polymorphic (MACMU)NAT1 constructs. To date, the NAT1 homologues of three macaque species (M. mulatta, M. sylvanus and M. fascicularis) have been cloned and characterized 61, 107, and those encode identical polypeptide chains. The (MACMU)NAT1 reference sequence (Nucleotide ID: KU640969.1) of the rhesus macaque was already available in pET28b(+) expression vector (Novagen) for the purposes of this study. Incorporation of each non-synonymous SNV found in the various (MACMU)NAT1 haplotypes was performed by site-directed mutagenesis using the QuikChange II kit (Agilent), as previously described 7. The mutagenised constructs were verified by sequencing of the insert (GATC Biotech, Konstanz, Germany).

Recombinant protein expression and purification. Expression of polyhistidine-tagged (MACMU) NAT1 protein variants was carried out in 200 ml cultures of E. coli BL21(DE3)pLysS cells as previously described 61, 73. The soluble fraction of bacterial lysates was recovered and immediately subjected to purification by affinity chromatography through Nickel-charged columns (Qiagen). Recombinant proteins were eluted with increasing concentration of imidazol and analysed by quantification at 280 nm on a Nanodrop 2000 Spectrophotometer (Thermo Scientific). Protein separation was carried out by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and gels were photographed with BioRad ChemiDoc XRS + Imaging System. All handling of proteins took place on ice or at 4 °C, and subsequent experiments were carried out within one day from protein production (while maintaining on ice in a cold chamber), to avoid freeze-thawing.

Immunoblot analysis. Immunoblot detection was performed according to standard practices, using the BioRad Mini-Protein Tetra Cell System and Criterion Blotter. Gels were blotted onto Immuno-Blot PVDF membrane for antibody binding and subsequent chemiluminescence detection with Immun-Star Western kit (both products from BioRad). The antibodies used were: (a) commercial horseradish peroxidase (HRP) conjugated monoclonal antibody A7058 (Sigma-Aldrich) against the polyhistidine tag of recombinant proteins, diluted 1:12,000; (b) commercial non-selective NAT rabbit polyclonal antibody ARP44197 (kindly provided by Aviva Systems Biology, San Diego, CA, USA), diluted 1:2000; (c) NAT1-selective rabbit polyclonal antiserum #183, raised against the bovine serum albumin (BSA) conjugated C-terminal peptide of human NAT1 (kindly provided by Professor Edith Sim, University of Oxford, UK), diluted 1:4000. The binding of antibodies ARP44197 and #183 was detected with HRP-conjugated monoclonal anti-rabbit IgG secondary antibody A1949 (Sigma-Aldrich), diluted 1:12,000; (b) commercial non-selective NAT rabbit polyclonal antibody ARP44197 (kindly provided by Aviva Systems Biology, San Diego, CA, USA), diluted 1:2000; c) NAT1-selective rabbit polyclonal antiserum #183, raised against the bovine serum albumin (BSA) conjugated C-terminal peptide of human NAT1 (kindly provided by Professor Edith Sim, University of Oxford, UK), diluted 1:4000. The binding of antibodies ARP44197 and #183 was detected with HRP-conjugated monoclonal anti-rabbit IgG secondary antibody A1949 (Sigma-Aldrich), diluted 1:10,000. Quantification of affinity chromatography purified proteins (100 mM imidazol fractions) was carried out by spectrophotometry at 280 nm (Nanodrop). Total protein quantification in cell lysates was carried out using the Bradford reagent (Sigma-Aldrich), as instructed by the manufacturer. The ChemiDoc XRS + Imaging System was used to detect and quantify the chemiluminescent signal on immunoblot membranes.

Enzyme activity assays. The enzyme activity assays 108 were performed with 1 μg of affinity chromatography purified NAT1 protein (100 mM imidazol fraction), according to well-standardized procedures in our laboratory 61, 73, 74. The acyl-group acceptor substrates used (0.5 mM) were the NAT1-selective PABA (PubChem CID: 978) and 5AS (PubChem CID: 4075), the NAT2-selective SMZ (PubChem CID: 5327), all three pharmaceutical arylamines, and the less selective PANS (PubChem CID: 7732), a toxic arylamine. Acetyl-CoA (0.4 mM) was used as acyl-group donor substrate and the enzymatic release of CoA was monitored over time with Ellman’s reagent. Reactions were performed in duplicate.

Thermal stability assays. For comparing thermal stability of recombinant (MACMU)NAT1 variants of the rhesus macaque, protein expression/purification was carried out and the enzyme activity was recorded using PABA as substrate. The procedures were as described in the previous sections. Each recombinant protein preparation was then incubated for 10 min on a PCR DNA Engine Tetrad Gradient Cycler (MJ Research) across a temperature range spanning the denaturation midpoint temperature (Tm = 36.5 ± 1.2 °C) previously determined for (MACMU)NAT1 recombinant protein by differential scanning fluorimetry 109. Following thermal stress, the activity of each recombinant variant was measured with PABA during a 5 min enzymatic assay, as described above.

Computational analyses. BioEdit version 7.1.9 109 was used to inspect sequences and perform multiple alignments on ClustalW. NAT1 haplotypes were inferred from the unphased genotypes of sequenced individuals with PHASE v2.1.1 software 110, using the default parameter values in the Markov chain Monte Carlo simulations. The algorithm was applied ten times with different random seed numbers to check for consistency of the results across the independent runs. We chose the results from the run displaying the best average goodness-of-fit of the estimated haplotypes to the underlying coalescent model. Primate sequences for NAT1 were retrieved from the Genome Database (https://www.ncbi.nlm.nih.gov/genome/), using BLASTn, and the search for repetitive elements downstream of NAT1 gene was carried out with RepeatMasker open version 4.0.8 (http://www.
The presentation of SNVs in the manuscript is according to the guidelines of the Human Genome Variation Society (http://www.hgvs.org/mutnomen/) and each haplotype was named following the recommendations of the NAT Gene Nomenclature Committee (http://nat.mbg.duth.gr/). The reference allele (Nucleotide ID: KU640969.1) was assigned symbol (MACMU)NAT1*1 and the remaining haplotypes were named in the order of their annotation during the study. A GenBank accession number (Nucleotide ID) was assigned per haplotype.

Structural alignment of protein sequences was carried out with T-COFFEE Expresso (http://tcoffee.org/cat/apps/coffee/do/expresso) and graphically visualised with ESPript3.0 (http://esprit.ibcp.fr). Modelling of (MACMU)NAT1 was performed against the known structure of human NAT1 protein (PDB ID: 2PQT) using the Swiss-Model online tool (http://swissmodel.expasy.org/). Graphical visualisation of models was performed on PyMOL (Schrödinger, LLC). The effect of each polymorphism on (MACMU)NAT1 protein was also predicted using PolyPhen-2 (HumDiv model, http://genetics.bwh.harvard.edu/pph2/), an online bioinformatics tool assigning impact probability scores in the scale from 0.00 (benign) to 1.00 (probably damaging). The potential consequences of amino acid substitutions were further assessed according to Russell and colleagues (http://www.russelllab.org/aas/).

References
1. Minchin, R. F. et al. Arylamine N-acetyltransferase I. Int J Biochem Cell Biol 39, 1999–2005 (2007).
2. Butcher, N. J. & Minchin, R. F. Arylamine N-acetyltransferase 1: a novel drug target in cancer development. Pharmacol Rev 64, 147–165, https://doi.org/10.1124/pr.110.094275 (2012).
3. Sim, E., Walters, K. & Boukouvala, S. Arylamine N-acetyltransferases: from structure to function. Drug Metab Rev 40, 479–510, https://doi.org/10.1080/03602530380168603 (2008).
4. Payton, M., Smelt, V., Upton, A. & Sim, E. A method for genotyping murine arylamine N-acetyltransferase type 2 (NAT2): a gene expressed in preimplantation embryonic stem cells encoding an enzyme acetylatng the folate catabolite p-aminozoxylglutamate. Biochem Pharmacol 58, 779–783 (1999).
5. Smelt, V. A. et al. Expression of arylamine N-acetyltransferases in pre-term placentas and in human pre-implantation embryos. Hum Mol Genet 9, 1101–1107 (2000).
6. Boukouvala, S., Price, N. & Sim, E. Identification and functional characterization of novel polymorphisms associated with the genes for arylamine N-acetyltransferases in mice. Pharmaco geneticas 12, 385–394 (2002).
7. Pacifici, G. M., Bencini, C. & Rate, A. Acetyli transferase in humans: development and tissue distribution. Pharmacology 32, 283–291 (1986).
8. Chung, J. G., Levy, G. N. & Weber, W. W. Distribution of 2-aminofluorene and p-aminobenzoic acid N-acetyltransferase activity in tissues of C57BL/6j and B6.A-Nats slow acetylator congenic mice. Drug Metabol Dispos 21, 1057–1063 (1993).
9. Stanley, L. A. et al. Immunochernical detection of arylamine N-acetyltransferase during mouse embryonic development and in adult mouse brain. Teratology 58, 174–182, https://doi.org/10.1002/tci.1096-9926(199811)58:5<174:aaid-teta3-3.0.co;2-q (1998).
10. Boukouvala, S. & Sim, E. Structural analysis of the genes for human arylamine N-acetyltransferases and characterisation of alternative transcripts. Basic Clin Pharmacol Toxicol 96, 343–351, https://doi.org/10.1111/j.1742-7843.2005.pto_02.x (2005).
11. Husain, A. et al. Functional analysis of the human N-acetyltransferase 1 major promoter: quantitation of tissue expression and identification of critical sequence elements. Drug Metab Dispos 35, 1649–1656, https://doi.org/10.1124/dmd.107.016485 (2007).
12. Boukouvala, S., Price, N., Plant, K. E. & Sim, E. Structure and transcriptional regulation of the Nat2 gene encoding for the drug-metabolizing enzyme arylamine N-acetyltransferase type 2 in mice. Biochem J 375, 593–602, https://doi.org/10.1042/bj20030812 (2003).
13. Butcher, N. J., Arulpragasam, A., Goh, H. L., Davey, T. & Minchin, R. F. Genomic organization of human arylamine N-acetyltransferase Type I reveals alternative promoters that generate different 5’UTR splice variants with altered translational activities. Biochem J 387, 119–127, https://doi.org/10.1042/bj20030803 (2005).
14. Wakefield, L., Boukouvala, S. & Sim, E. Characterisation of CpG methylation in the upstream control region of mouse Nat2: evidence for a gene-environment interaction in a polymorphic gene implicated in folate metabolism. Gene 452, 16–21, https://doi.org/10.1016/j.gene.2009.12.002 (2010).
15. Paterson, S., Sim, K. L., Tjiang, J. M., Minchin, R. F. & Butcher, N. J. Histone deacetylase inhibitors increase human arylamine N-acetyltransferase-1 expression in human tumor cells. Drug Metab Dispos 39, 77–82, https://doi.org/10.1124/dmd.110.036202 (2011).
16. Barker, D. F. et al. Functional properties of an alternative, tissue-specific promoter for human arylamine N-acetyltransferase 1. Pharmacogenet Genomics 16, 515–525, https://doi.org/10.1097/01.ppc.0000415295.29342.26 (2006).
17. Butcher, N. J., Tetlow, N. L., Cheung, C., Broadhurst, G. M. & Minchin, R. F. Induction of human arylamine N-acetyltransferase type I by androgens in human prostate cancer cells. Cancer Res 67, 85–92, https://doi.org/10.1158/0008-5472.can-06-2635 (2007).
18. Wakefield, L. et al. Arylamine N-acetyltransferase 1 expression in breast cancer cell lines: a potential marker in estrogen receptor-positive tumors. Genes Chromosomas Cancer 47, 118–126, https://doi.org/10.1002/gcc.20512 (2008).
19. Butcher, N. J. & Minchin, R. F. Arylamine N-acetyltransferase gene regulation by androgens requires a conserved heat shock element for heat shock factor-1. Carcinogenesis 31, 820–826, https://doi.org/10.1093/carcin/bgp042 (2010).
20. Bonamassa, B., Ma, Y. & Liu, D. Glucocorticoid receptor-mediated transcriptional regulation of N-acetyltransferase I gene through dimeric promoter. AAPS J 14, 581–590, https://doi.org/10.1208/s12248-012-9370-5 (2012).
21. Endo, Y. et al. miR-1290 and its potential targets are associated with characteristics of estrogen receptor alpha-positive breast cancer. Endocr Relat Cancer 20, 91–102, https://doi.org/10.1530/erc-12-0207 (2013).
22. Endo, Y. et al. Immunohistochemical determination of the miR-1290 target arylamine N-acetyltransferase 1 (NAT1) as a prognostic biomarker in breast cancer. BMC Cancer 14, 990, https://doi.org/10.1186/1471-2407-14-990 (2014).
23. Liu, F. et al. Arylamine N-acetyltransferase aggregation and constitutive ubiquititation. J Mol Biol 361, 482–492, https://doi.org/10.1016/j.jmb.2006.06.029 (2006).
24. Liu, F., Koepp, D. M. & Walters, K. J. Artificial targeting of misfolded cytosolic proteins to endoplasmic reticulum as a mechanism for clearance. Sci Rep 5, 12088, https://doi.org/10.1038/srep12088 (2015).
25. Minchin, R. F. & Butcher, N. J. The role of lysine(100) in the binding of acetylcoenzyme A to human arylamine N-acetyltransferase 1: implications for other transferases. Biochem Pharmacol 94, 195–202, https://doi.org/10.1016/j.bcp.2015.01.015 (2015).
26. Minchin, R. F., Rosenkranz, K. J., Burow, R. & Butcher, N. J. AllostERIC regulation of arylamine N-acetyltransferase I by adenosine triphosphate. Biochem Pharmacol 158, 153–160, https://doi.org/10.1016/j.bcp.2018.10.013 (2018).
27. Kawamura, A. et al. Eukaryotic arylamine N-acetyltransferase. Investigation of substrate specificity by high-throughput screening. Biochem Pharmacol 69, 347–359, https://doi.org/10.1016/j.bcp.2004.09.014 (2005).
28. Minchin, R. F. Acetylation of p-aminozoxylglutamate, a folate acid catabolite, by recombinant human arylamine N-acetyltransferase and U937 cells. Biochem J 307(Pt 1), 1–3 (1995).
63. Boukouvala, S., Sabbagh, A. & Fakis, G. In "Arylamine N-Acetyltransferases in Health and Disease" (eds. B. M. Mortensen, C. Vangenot, T. Tsirka, S. M. Carlisle and D. W. Hein), 125-140. Springer, Berlin Heidelberg (2019).

64. Wu, H. Comparative analysis of xenobiotic metabolising N-acetyltransferases from ten non-human primates as a model for humans. Biochem Pharmacol 83, 773-782, https://doi.org/10.1016/j.bcp.2012.01.037 (2012).

65. Laurieri, N. et al. Structural basis of substrate-binding specificity of human arylamine N-acetyltransferases. J Mol Biol 348, 921-929, https://doi.org/10.1016/j.jmb.2005.02.016 (2005).

66. Witham, K. L. et al. From arylamine N-acetyltransferase to folate-dependent acetyl CoA hydrolase: impact of folic acid on the activity of (HU)NAT1 and its homologue (MOUSE)NAT2. PLoS One 9, e96370, https://doi.org/10.1371/journal.pone.0096370 (2014).

67. Zeng, X. et al. Knockout of human arylamine N-acetyltransferase 1 (NAT1) in MDA-MB-231 breast cancer cells leads to increased reserve capacity, maximum mitochondrial capacity, and glycolytic reserve capacity. Mol Cancer 57, 1458-1466 (2018).

68. Stepp, M. W. et al. Knockout of human arylamine N-acetyltransferase 1 (NAT1) in MDA-MB-231 breast cancer cells leads to increased reserve capacity, maximum mitochondrial capacity, and glycolytic reserve capacity. Mol Cancer 57, 1458-1466 (2018).

69. Carlisle, S. M. et al. Untargeted polar metabolomics of transformed MDA-MB-231 breast cancer cells expressing varying levels of human arylamine N-acetyltransferase 1. Metabolomics 12, 1056-1066 (2016).

70. Carlisle, S. M. et al. Role for human arylamine N-acetyltransferase 1 in the methionine salvage pathway. Biochimica et Biophysica Acta 1813, 1913-1921, https://doi.org/10.1016/j.bbabio.2011.06.002 (2011).

71. Wakefield, L., Cornish, V. J., Long, H., Griffiths, W. J. & Sim, E. Deletion of a xenobiotic metabolizing gene in mice affects folate metabolism. Biochim Biophys Acta 1789, 169-177, https://doi.org/10.1016/j.bbamet.2008.10.001 (2009).

72. Mortensen, H. M. et al. Characterization of genetic variation and natural selection at the arylamine N-acetyltransferase genes in global human populations. Pharmacogenomics 12, 1545-1558, https://doi.org/10.2217/pgs.11.88 (2011).

73. Vangenot, C. et al. Humans and chimpanzees display opposite patterns of diversity in arylamine N-acetyltransferase genes. G3 (Bethesda) 9, 1100-1108, https://doi.org/10.1093/g3/jxy065 (2019).

74. Tsirka, T. et al. Comparative analysis of xenobiotic metabolising N-acetyltransferases from ten non-human primates as in vitro models of human homologues. Sci Rep 8, 108-20904 (2018).

75. Sabbagh, A. et al. Rapid birth-and-death evolution of the xenobiotic metabolizing NAT gene family in vertebrates with evidence of adaptive selection. BMC Evol Biol 13, 62, https://doi.org/10.1186/1471-2148-13-62 (2013).

76. Boukouvala, S., Sabbagh, A. & Fakis, G. In "Arylamine N-Acetyltransferases in Health and Disease" (eds. B. M. Mortensen, C. Vangenot, T. Tsirka, S. M. Carlisle and D. W. Hein), 125-140. Springer, Berlin Heidelberg (2019).

77. Witham, K. L. et al. Genetic and small molecule inhibition of arylamine N-acetyltransferase-1 deletion. PLoS One 8, e77923, https://doi.org/10.1371/journal.pone.0077923 (2013).

78. Stepp, M. W., Doll, M. A., Carlisle, S. M., States, J. C. & Hein, D. W. Folate-Dependent Hydrolysis of Acetyl-Coenzyme A by Recombinant Human and Rodent Arylamine N-Acetyltransferases. Biochem Biophys Res Commun 364, 556-560, https://doi.org/10.1016/j.bbrc.2007.10.026 (2007).

79. Cao, W. et al. Low levels of exogenous N-acetyltransferase can be achieved in transgenic mice. Pharmacogenomics J 5, 253-261, https://doi.org/10.1038/sj.tpj.6500319 (2005).
103. Podgorna, E. et al. Variation in NAT2 acetylation phenotypes is associated with differences in food-producing subsistence modes and ecoregions in Africa. *BMC Evol Biol* 15, 263, https://doi.org/10.1186/s12862-015-0543-6 (2015).
104. Sabbagh, A., Darlu, P., Vangenot, C. & Poloni, E. S. In *Arylamine N-Acetyltransferases in Health and Disease* 165–193 (2018).
105. Bonhomme, M., Castrero, S., Blancher, A. & Crouau-Roy, B. Assessing natural introgression in 2 biomedical model species, the rhesus macaque (Macaca mulatta) and the long-tailed macaque (Macaca fascicularis). *J Hered* 100, 158–169, https://doi.org/10.1093/jhered/esu093 (2009).
106. Blancher, A. et al. Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (Macaca fascicularis fascicularis). *J Hered* 99, 254–264, https://doi.org/10.1093/jhered/esu003 (2008).
107. Uno, Y., Murayama, N. & Yamazaki, H. Molecular and Functional Characterization of N-Acetyltransferases NAT1 and NAT2 in Cynomolgus Macaque. *Chem Res Toxicol* 31, 1269–1276, https://doi.org/10.1021/acs.chemrestox.8b00236 (2018).
108. Brooke, E. W. et al. An approach to identifying novel substrates of bacterial arylamine N-acetyltransferases. *Bioorg Med Chem* 11, 1227–1234 (2003).
109. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95–98 (1999).
110. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73, 1162–1169, https://doi.org/10.1086/379378 (2003).
111. Hein, D. W., Boukouvala, S., Grant, D. M., Minchin, R. F. & Sim, E. Changes in consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenet Genomics* 18, 367–368, https://doi.org/10.1097/FPC.0b013e3282f60db0 (2008).

Acknowledgements
We thank Professor Edith Sim (University of Oxford, UK) for kindly providing NAT1-specific polyclonal antiserum #183.

Author Contributions
S.B., G.F., B.C.R. and A.S. conceived of and designed the study, supervised implementation of the experimental procedures and analysed the data; Z.C., D.G., E.K., N.M., S.V.R., I.S., T.T., C.V. and S.Z. (presented in alphabetical order) carried out the experiments and performed initial data analyses with equal contributions. S.B. and G.F. wrote the manuscript. The materials and datasets generated during the study are available from G.F. on reasonable request. All authors reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-47485-x.

Competing Interests: T.T. was recipient of PhD scholarships from the Onassis Foundation and the State Scholarships Foundation of Greece. The EDB lab was supported by “Investissement d’Avenir” Grants (CEBA, ANR 10LabX-0025; TULIP, ANR-10-labX-41). S.Z. is currently an employee of Medicon Hellas S.A., Greece. None of the funding bodies had any involvement in the design or conduct of this study or preparation of the manuscript.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019