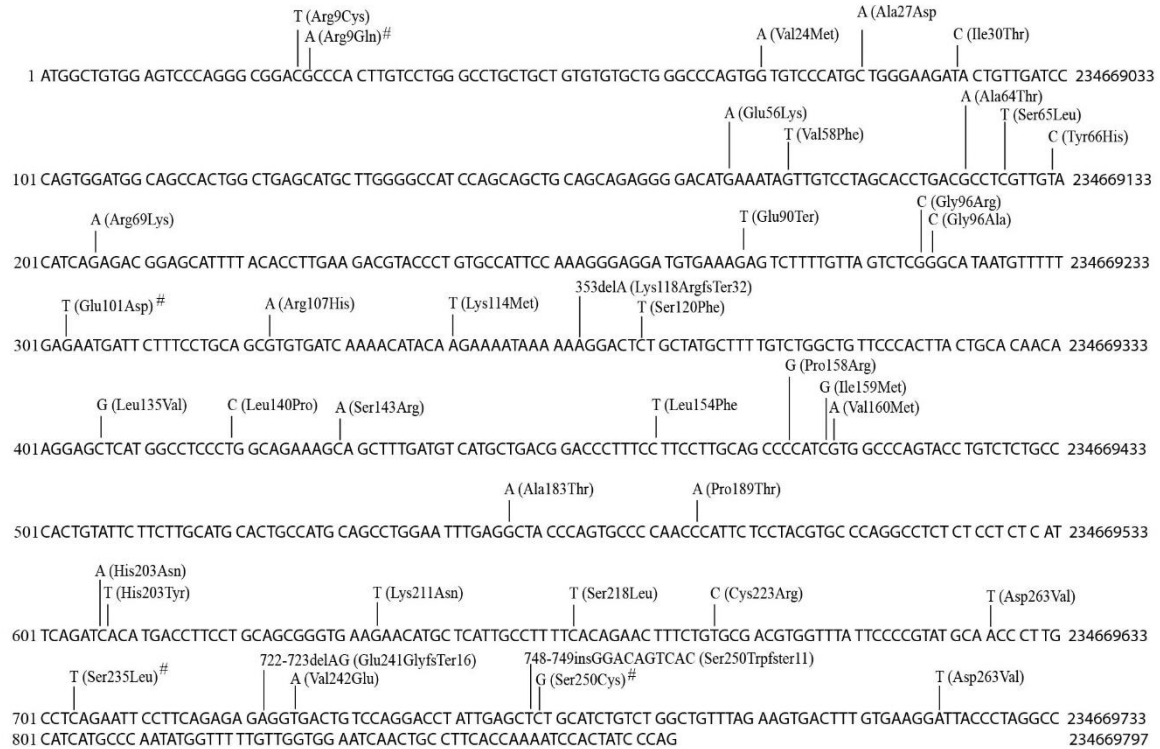
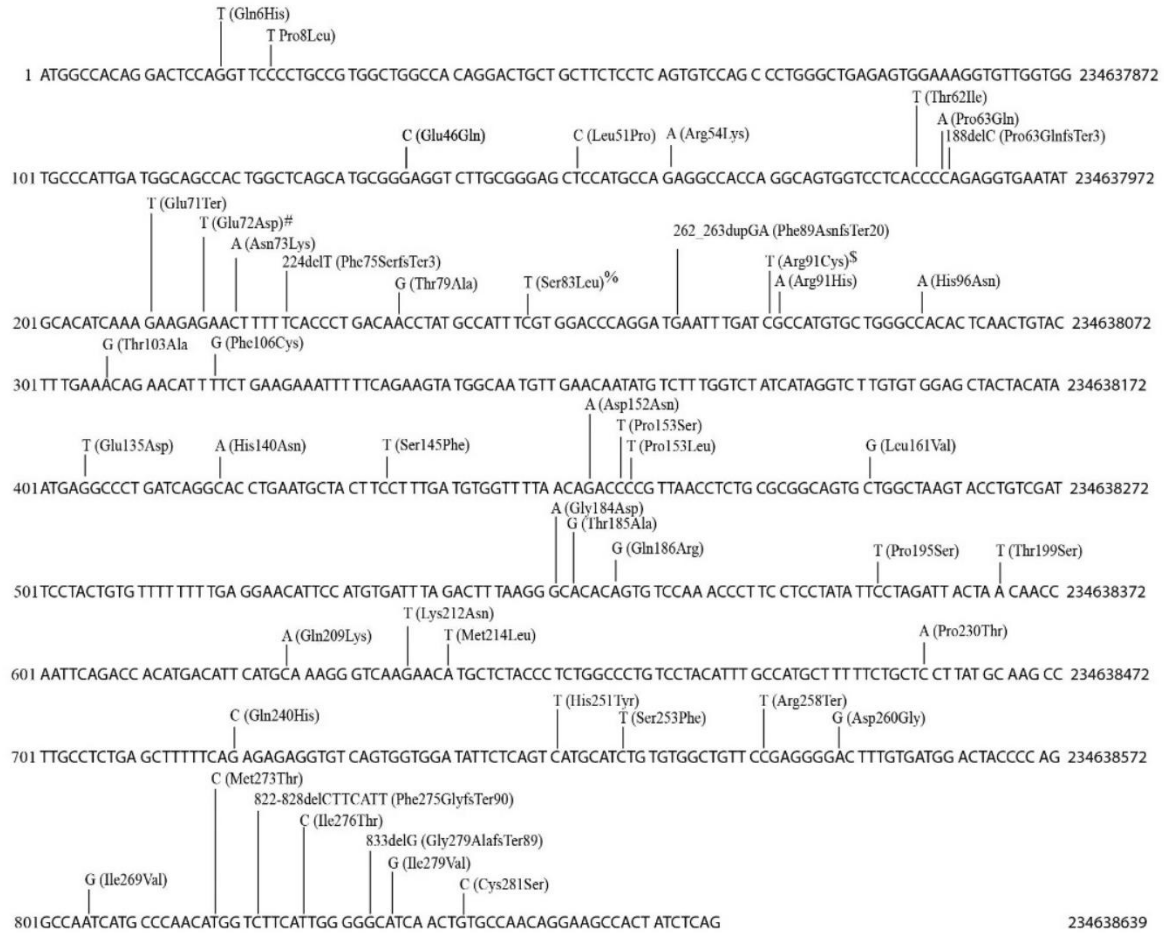


# The Somatic Mutation Landscape of UDP-Glycosyltransferase (*UGT*) Genes in Human Cancers

Dong Gui Hu\*, Shashikanth Marri, Julie-Ann Hulin, Ross A McKinnon, Peter I. Mackenzie, and Robyn Meech



**Figure S1.** Mutations within the *UGT1A1* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A1* exon 1 (864 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_00463.3, left) positions. Mutations (missense, small indels) and the resulting changes at the protein level (NP\_000454.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice).



**Figure S2.** Mutations within the *UGT1A3* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A3* exon 1 (867 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_019093.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_061966.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), % (four times).



**Figure S3.** Mutations within the *UGT1A4* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A4* exon 1 (867 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_007120.3, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_009051.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), @ (eight times).

T (Gly4Ter) G (Gln11Arg)  
 1 ATGGCCACAG GACTCCAGGT TCCCTGCCG CAGCTGGCCA CAGGACTGCTGCTCTCTAGTGTCCAGCCCTGGGCTGAGAGTGGGAAGGTGCTGGTGG 234621737  
 105delC (Thr36Leu/Ister30) A (Arg49Gln) C (Val58Ala) G (Met67Val)  
 101 TGCCCACTGATGGCAGCCTGGCTCAGCATGCGGGAGGCCCTTGGGACCTCCATGCGA GAGGCCACCAGGTGGTGGTCTCACCTGG AGGTGAATAT 234621837  
 A (Trp84Arg) A (Trp84Ter) C (Gln86Pro) A (Glu88Lys)<sup>9%</sup> A (Phe89Ile) A (Phe89Leu) T (Gln98His)  
 201 GTACATCAA GAAGAGA AACT TTTTACCCT GACAACGAT GCCATTTCAT GGACCCAGGA CGAATTGAT CGCCTT TTGCTGGGTCACACTCAATCGTTC 234621937  
 A (His105Asn) A (Leu107Met) G (Cys128Gly)  
 301 TT TGAAACAG AACATCT TCT GATGAAAT TT TCTAGAAGAA TGGCAATTAT GAACAATATG TCTTTGATCA TACATAGGTC TTGTGTGGAGTACTGCATA 234622037  
 A (Leu137Met) T (Ser145Phe) A (Val148Met) G (Thr151Arg) G (Pro153Arg) A (Ala158Thr) T (Ala162Ser) T (Ile167Phe)  
 401 ATGAGGCCCT GATCAGGCAC CTGCATGCTA CTTCCTTTGA TGTG GTTCTA ACAGACCCCT TTCACCTCTG CGCGGCGGTG CTGGCTAAGTACCTGTCTGAT 234622137  
 A (Val170Met) T (Leu173Phe) T (Pro177Ser) C (Ser191Pro) G (Ser191Cys)  
 501 TCCTGCTGIG TTTTCTTGA GGAACATCC ATGTGATTTA GACTTTAAGG GCACACAGTG TCCAAACCCCT T CCTCTATA TTCCTAGAT T ACTAACGACC 234622237  
 C (Asp203His) A (Leu208Met) A (Arg210Lys) G (Met214Val) T (Val227Phe) T (Ala229Ser) T (Ala232Val)  
 601 AATTCAGACC ACATGACAT CTGCAAAGG GTCAAGAACA TGCTCTACCC TCTGGCCCTG TCCTACCTTT GCCATGCTGT TTCTGCTCCTTATGCAAGCC 234622337  
 G (Ser236Cys) 725\_726dupAG (Val243Arg/Ister21) T (Ser253Phe) A (Val254Met) A (Arg258Gln) G (Pro266Arg)  
 701 TTGCCTCTGA GCTTTTTCAG AGAGAGGTGT CAGTGGTGGA TCTGTGACCCATGCATCTG TGTGGCTGTTCCGAGGGGACTTTGTGATGGATTACCCAG 234622437  
 C (Met273Thr) T (Ile276Phe) A (Gly284Arg)  
 801 GCCGATCAIG CCCAACATGG TCTTCAITGG GGGCATCAAC TGTGCCAACG GGAAGCCACT ATCTCAG 234622504

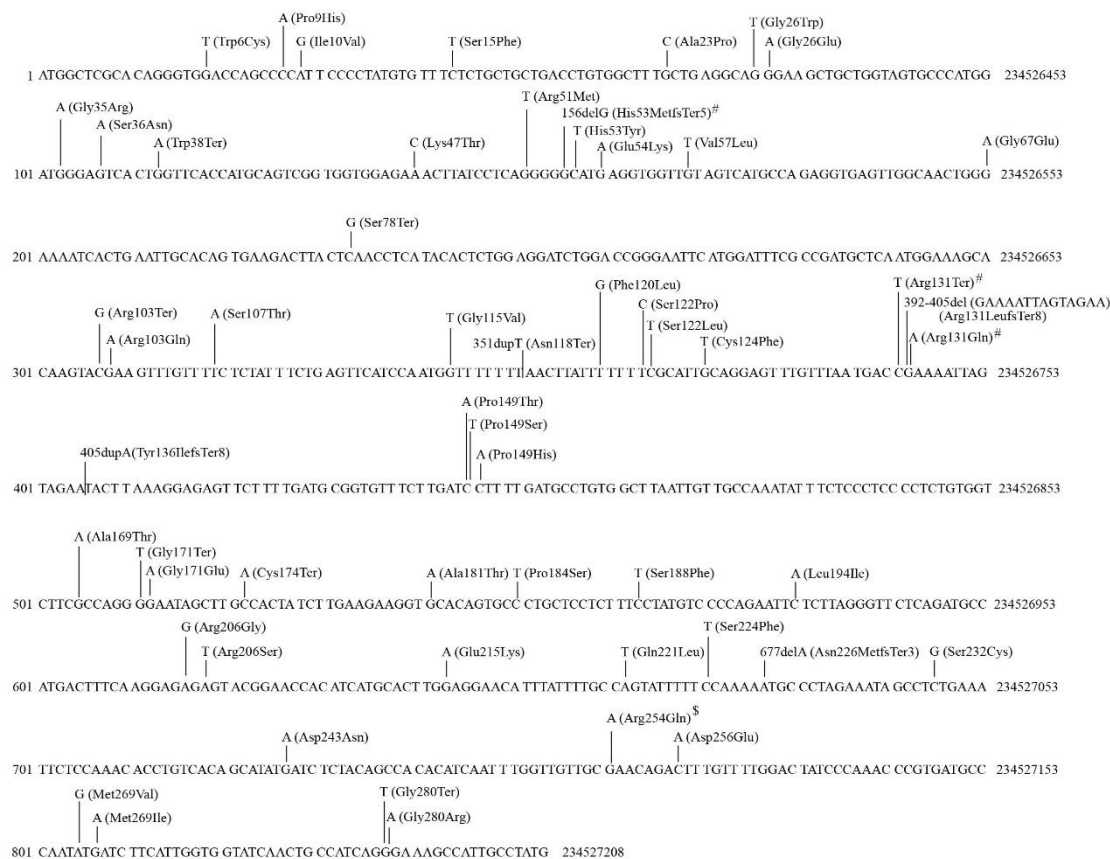
**Figure S4.** Mutations within the *UGT1A5* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A5* exon 1 (867 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_019078.2, left) positions. Mutations (missense, nonsense and small indels( and the resulting changes at the protein level (NP\_061951.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by % (four times).

T (Met1Ile) A (Arg6His) T (Arg10Ile) T (Trp21Leu) A (Leu30Met) T (Val32Phe)  
 1 ATGGCCTGCC TCCTTCGCTC AITTCAGAGA AITTCCTCAG GGGTTTCTT CTTAGCACTT TGGGGCATGG TTGTAGGTG A CAAGCTGCTG GTGGTCCCTC 234601750  
 T (Gln34Leu) A (Gly36Arg) T (Arg52Irp) A (Glu55Lys) A (Glu62Lys) T (Leu66Phe)  
 101 AGGACGGGAG CCACIGGCTT AGTAIGAAG ATATAGTTGA GGTTCCTCAGT GACCGGGGTCATG AGAATGT AGTGGTGGTG CCTGAAGTTA AITTCCTT 234601850  
 T (Glu69Ter)<sup>#</sup> G (Thr74Ala) G (Tyr78Ter) G (Pro79Ala) A (Arg90His) T (His98Lys) G (His98Arg) A (Ala100Asp)  
 201 GAAAGAATCCAAATACTACACAAGAAAAATCTATCCAGTGCCGTATGACCAAGAAGAGCT GAAGAACCGT TACCAATCAT TTGGAACAATCACT T TGCT 234601950  
 A (Arg102Gln) G (Gln109Glu) A (Leu121Met) C (Ile124Thr)  
 301 GAGCGATCAT TCCTAACTGC TCCTCAGACA GAGTACAGGAATAACATGAT TGTATTGGCCTGTACTTCA TCAACTGCCAGAGCCTCTG CAGGACAGGG 234602050  
 T (Asn137Tyr) G (Phe138Leu) C (Glu141Asp) T (Asp145Val) T (Pro151Ser)  
 401 ACACCTGAA CTTCCTTAAG GAGAGCAAGT TTGATGCTCT TTTCACAGAC CCAGCCTTACCCTGTGGGGT GATCCTGGCT GAGTATTGG GCCTACCAIC 234602150  
 A (Phe171Tyr) G (Arg172Gly) T (Pro175Ser)<sup>#</sup> T (Ser177Phe)  
 501 TGTGTACCTC TTCAGGGGTT TTCCGTGTTC CCTGGAGCAT ACATTCAGCA GAAGCCCA CCGTGTGTCC TACATTCCCA GGTGTACACAAAGTTTCA 234602250  
 616delT (Ser206Profs1er38) T (Arg208Ter) C (Val209Leu) C (Val214Leu) C (Phe226Leu) G (Phe226Cys) A (Glu230Lys)  
 601 GACCACATGA CTTT TTCCCA ACGAGTGGCC AACTTCCTTG TTAATT TGTT GGAGCCCTAT CTATT TTATT GTCTGT TTTC AAAGTATGAA GAATCTCGCAT 234602350  
 T (Ser234Leu) T (Gln248Ter)  
 701 CAGCTGTCCT CAAGAGAGAT GTGGATAA TCACCTTAIA TCAGAAGGTC TCTGTT TGGC TGTTAAGATA TGACTT TGTG CTTGAATATC CTAGGCCGGT 234602450  
 G (Ile277Met) T (Cys279Phe) T (Asp284Lys)  
 801 CATGCCCAAC ATGGTCTTCA TTGGAGGTAT CAACTGTAAG AAGAGGAAAGACTTGTCTCAG 234602511

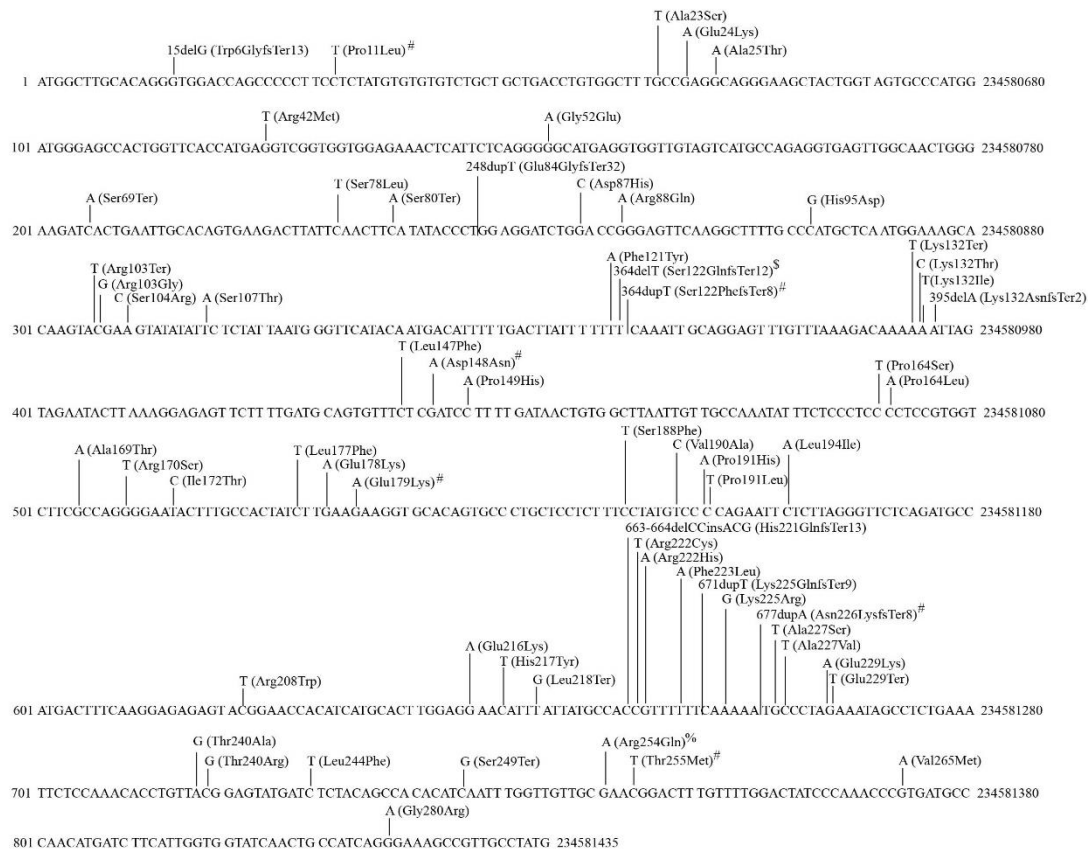
**Figure S5.** Mutations within the *UGT1A6* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A6* exon 1 (861 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001072.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001063.2) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice).

A (Arg3His) T (Gly8Cys) T (Pro11Leu) 51\_52delGCinsT(Ileu18Ter) C (Cys20Arg) T (Gly26Trp)  
 1 ATGGCTCGTG CAGGGTGGAC TGGCTCCTT CCCCTAATGTGTCTACTGCTGACCTGTGGCTTTGCCA AGGCAGGGAA GCTGCTGGTA GTGCCATGG 234590683  
 156dupG (His53AlafsTer2)  
 101 ATGGGAGCCACTGGTTCACCATGCAGTCGGTGGTGAGAACTCATCTCAGGGGGCATGAGGTGGTCGTAGTCATGCCAGAGGTGAGTTGGCAACTGGG 234590783  
 A (Arg97His) T (Arg97Leu) T (Thr99Met)  
 201 AAGATCACTG AATGCACAGTGAAGACTTACTCAACCTCATACTCTGG AGGATCAGGA CCGGGAGTTTCATGGTTTTCGCCGATGCTCG CTGGACGGCA 234590883  
 G (Ser78Ter) T (Arg88Trp) T (Val92Phe) 364delT(Ser122GlnfsTer12)\$ 365\_392del/insTCAAATGCAGGAGTTTGTTTAAGGACAA (Ser122GlnfsTer12) C (Lys132Asn) C (Leu133Ser)  
 301 CCAITGCGAA GTGCATT TTC TCTATTAACA AGTTCATCCA ATGGTATT TT TGACTTATTT T TTTCAAAT GCAGGAGT TT GTT TAATGAC CGAAAATTAG 234590983  
 A (Arg103Gln) T (Ser112Leu) A (Phe117Tyr) T (Ala158Val) T (Tyr160Phe) T (Leu163Phe) T (Pro164Ser)  
 401 TAGAATACCT AAAGGAGAGT IGTTTIGATG CAGTGTTCCT CGATCCTT TT GATGCCCTGTG GCCTAAATGT TGC CAAATAT TTTCTCCCTCC CCTCTGTGGT 234591083  
 A (Glu135Lys) C (Lys138Gln) G (Ala144Gly) T (Pro149Leu)# A (Ala152Thr) G (Leu187Val) T (Ser188Phe)# A (Leu193Ile)  
 501 CTTGCCAGG GGAATATTT GCCACTATCT TGAAGAAGGT GCAC AGTGCCCTGCTCCTT TTCCTATGTCGCCAGACTTCTCTAGGGTT CTCAGACGCC 234591183  
 G (Phe173Val) G (Leu177Val) A (Gly180Ser) T (Gly180Cys) T (Cys220Phe) 671dupT (Lys225GlnfsTer9)#  
 601 ATGACTTTCA AGGAGAGAGT ATGGAACAC ATCATGCAC TGGAGGAACA TTTATTTGCCCTAATTT T TCAAAAATGCTTTAGAAATA GCCTCTGAAA 234591283  
 T (Pro238Ser) T (Thr240Met)\$ T (Arg254Ter) A (Arg254Gln)# T (Pro262Leu) A (Val265Glu)  
 701 TTCTCCAAACCCCTGTACGGCATAATGATC TCTACAGCCA CACAATTT TGGTTGTGC GAACGTGACTT TGTGTTGGAG TATCCCAAAC CCGTGATGCC 234591383  
 C (Val283Ala)  
 801 CAATATGATC TTCATTGGTG GTATCAACTG TCATCAGGGA AAGCCAGTGCCTATG 234591438

**Figure S6.** Mutations within the *UGT1A7* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A7* exon 1 (855 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_019077.3, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_061950.2) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times).

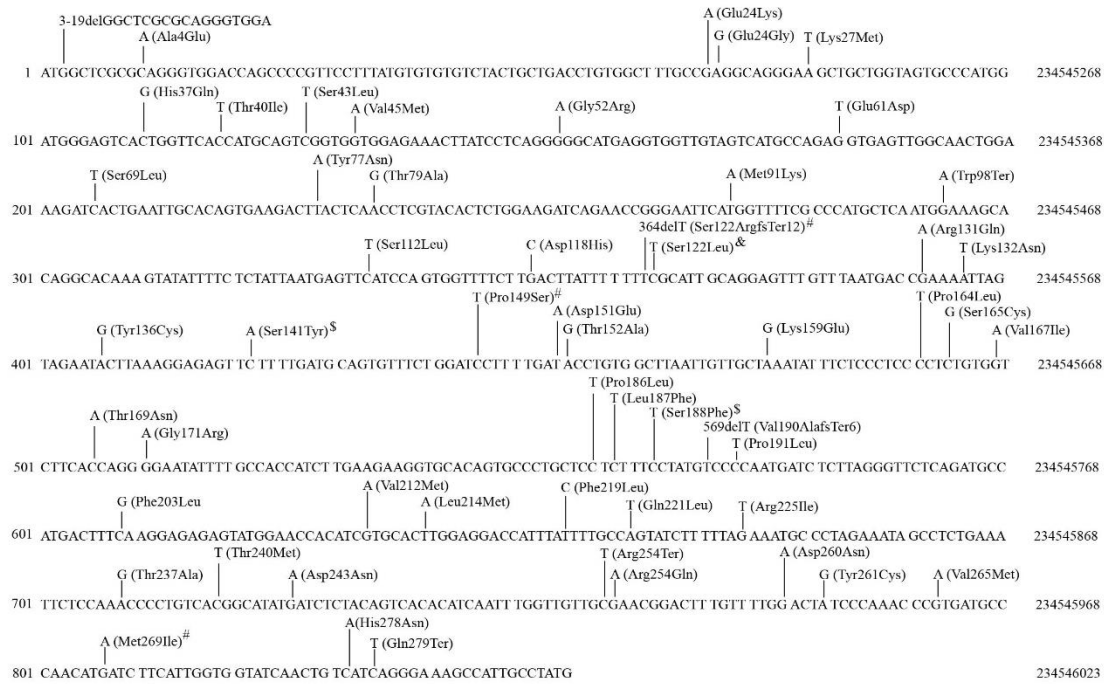


**Figure S7.** Mutations within the *UGT1A8* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A8* exon 1 (855 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_019076.5, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_061949.3) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times).

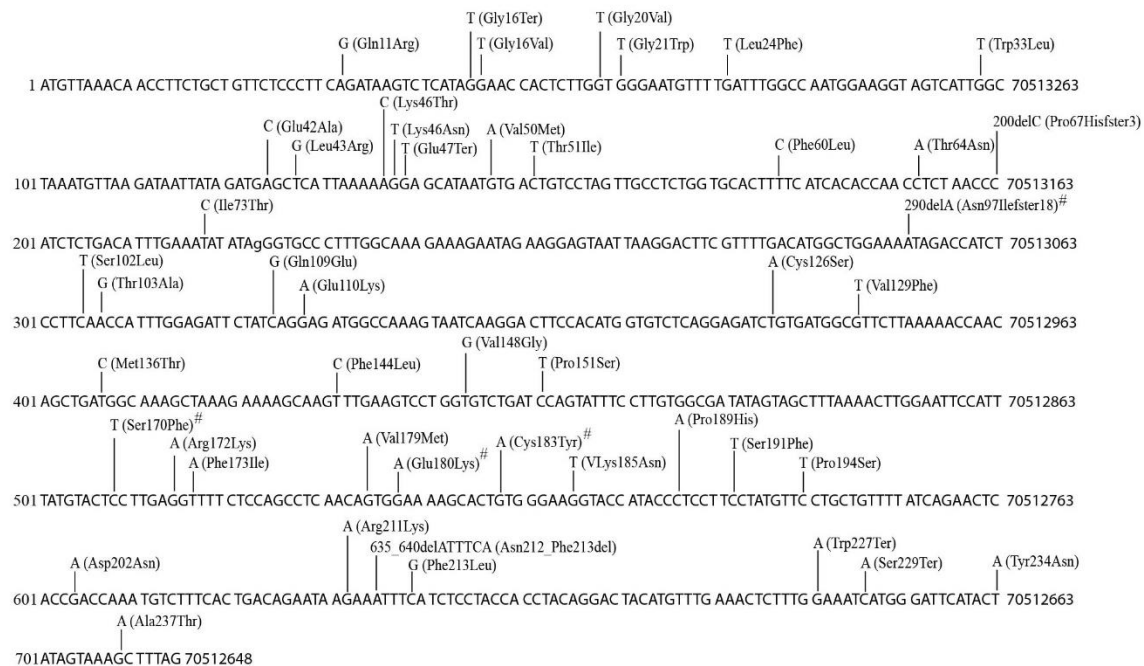


**Figure S8.** Mutations within the *UGT1A9* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A9* exon 1 (855 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_021027.3, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_066307.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), % (four times).

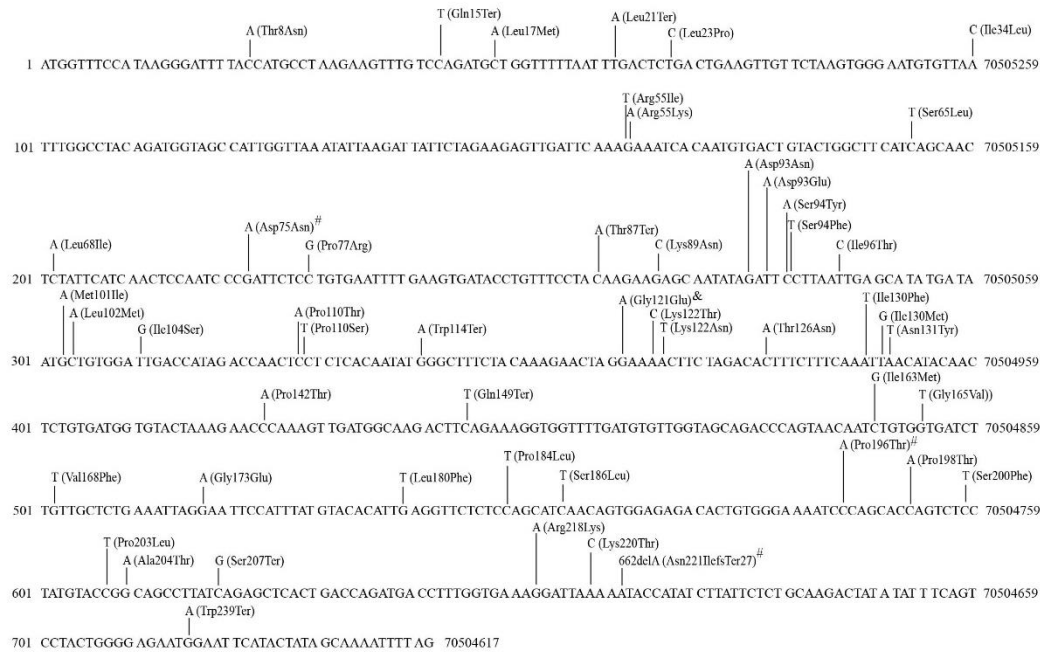




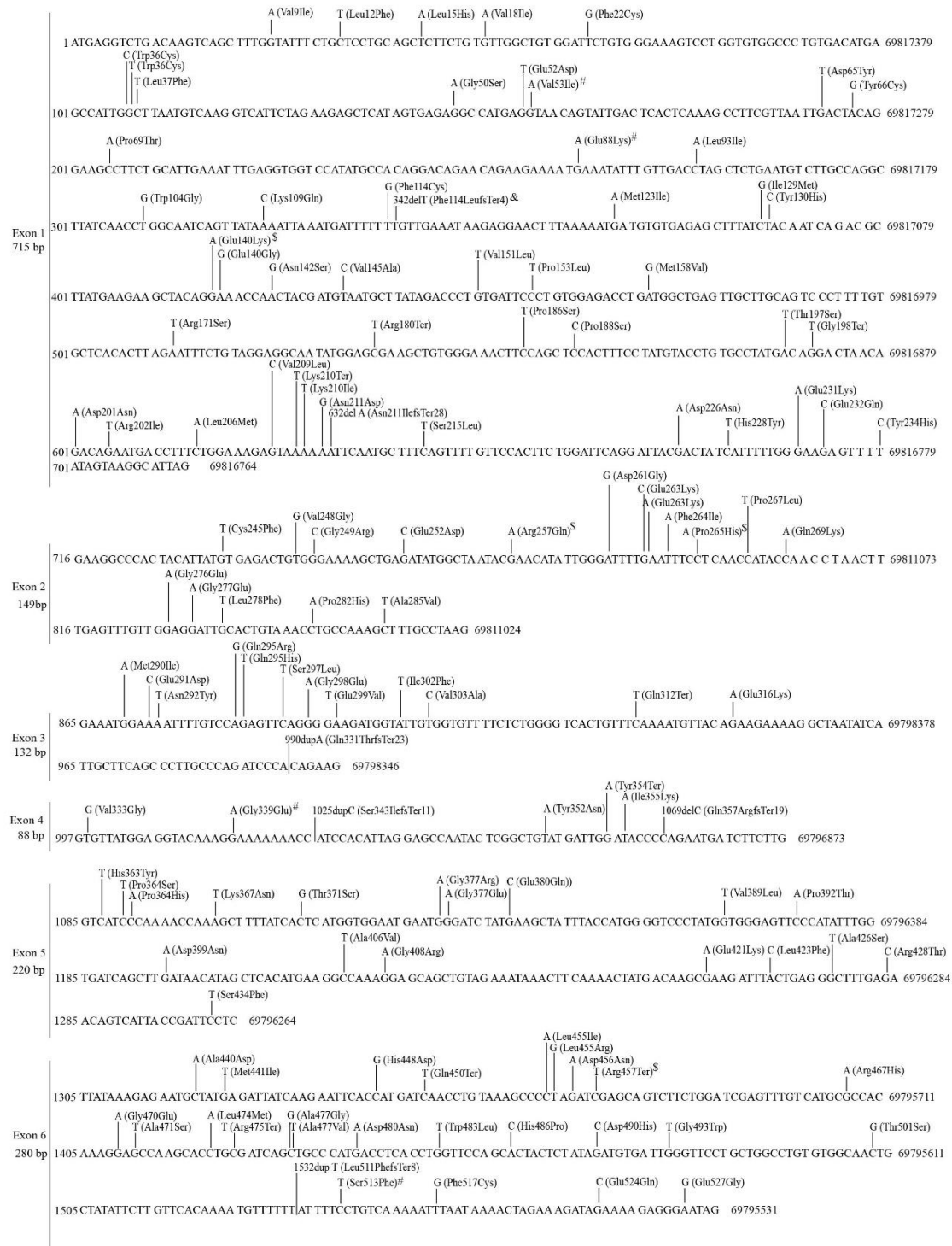
**Figure S9.** Mutations within the *UGT1A10* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT1A10* exon 1 (855 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_019075.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_061948.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), % (four times), & (five times)



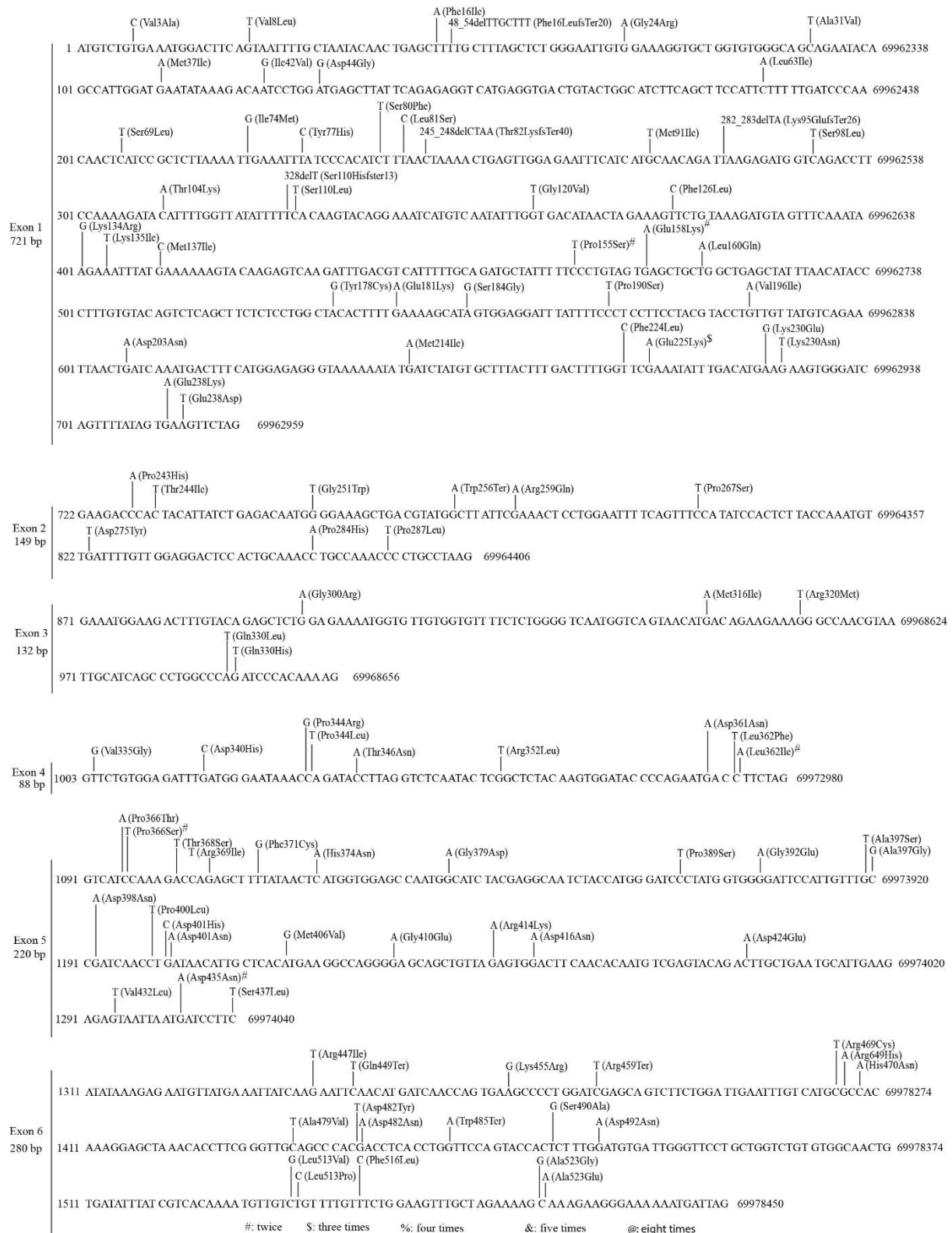
**Figure S10.** Mutations within the *UGT2A1* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT2A1* exon 1 (715 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_006798.5, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_006789.3) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice).



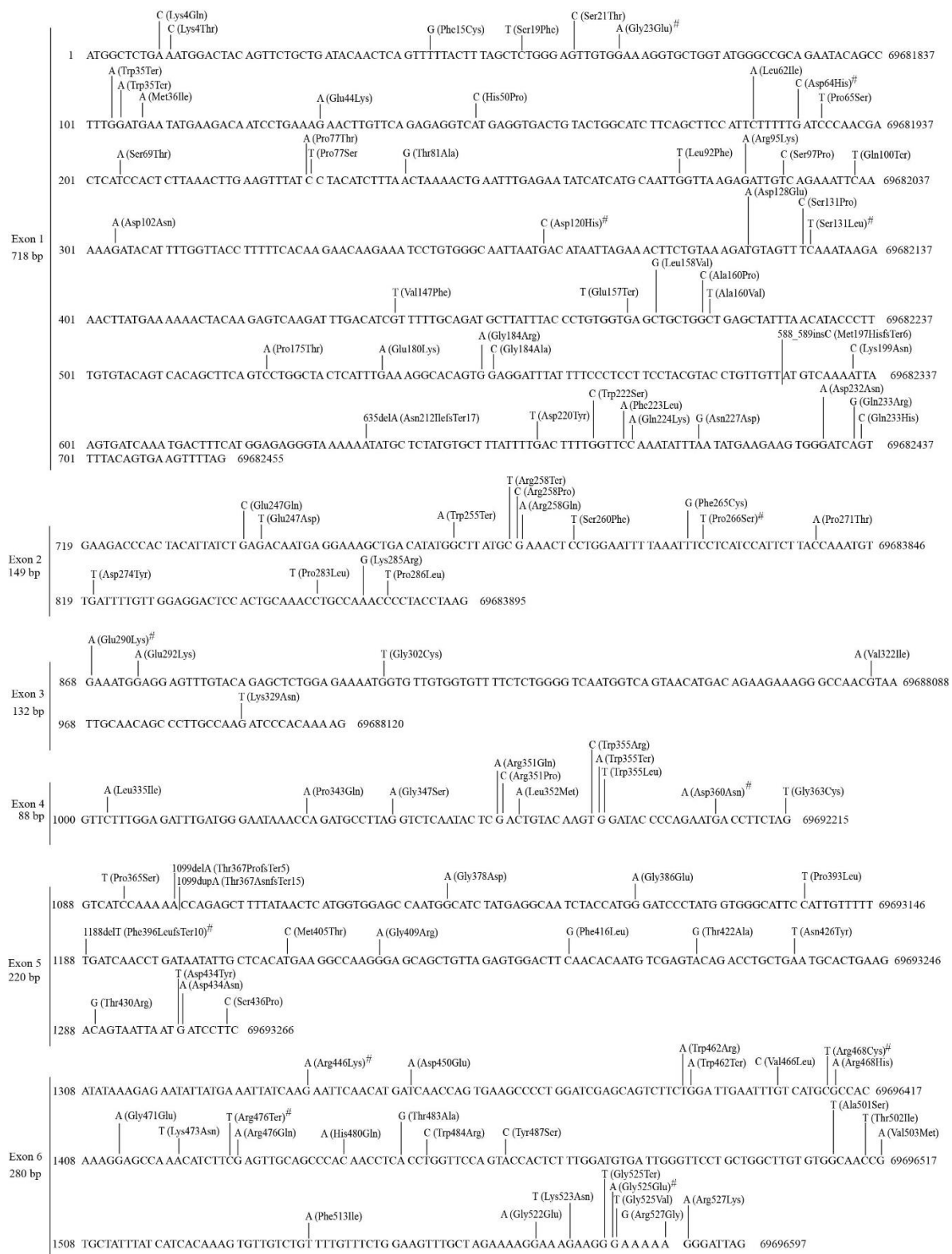
**Figure S11.** Mutations within the *UGT2A2* exon 1 in TCGA tumors. Data shown is the coding region of the *UGT2A2* exon 1 (742 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001105677.2, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001099147.2) are indicated above the reference mRNA sequence.



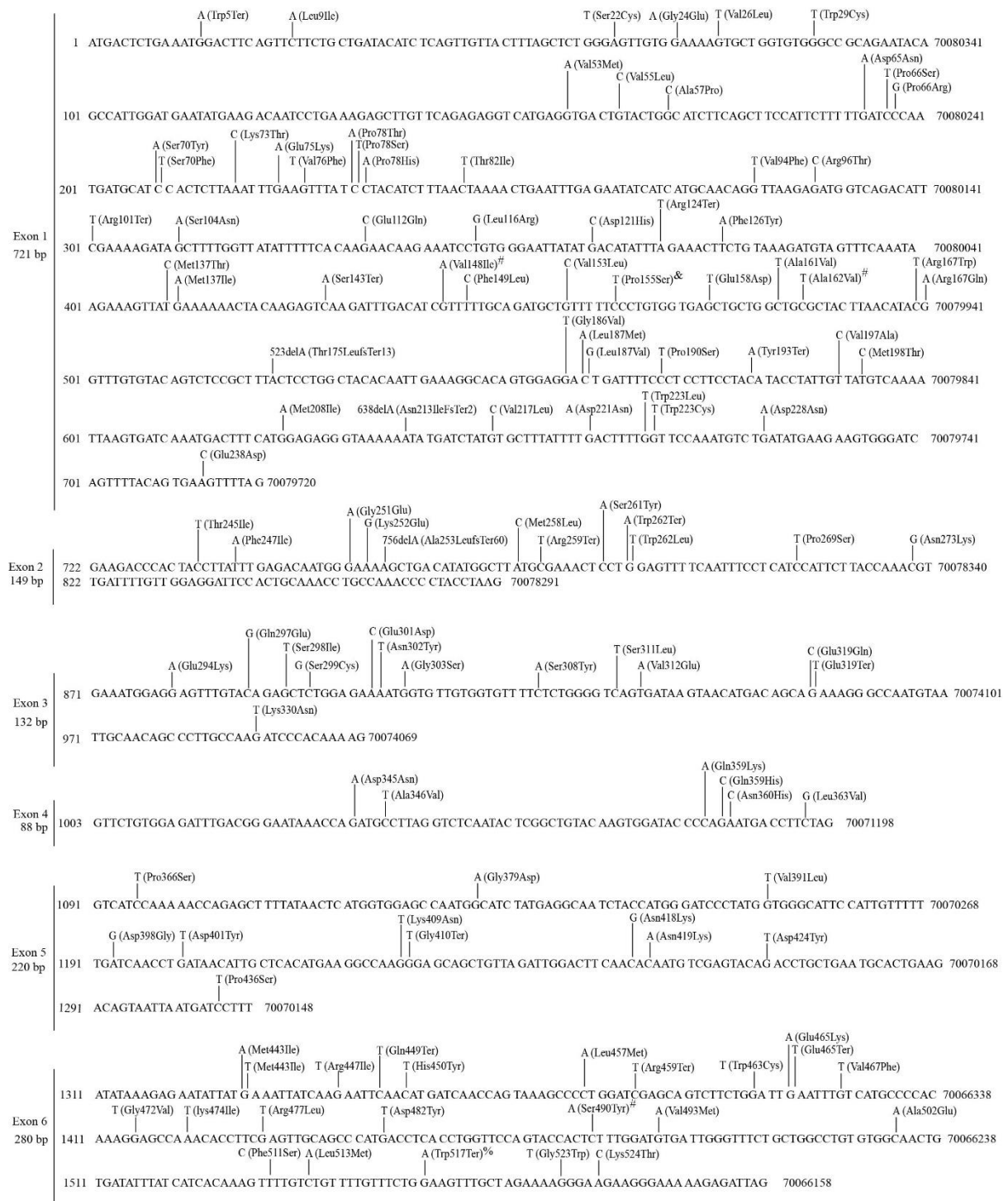
**Figure S12.** Mutations at the *UGT2A3* gene in TCGA tumors. Data shown is the coding region of the *UGT2A3* gene (6 exons, 1584 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_024743.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_079019.3) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), & (five times).



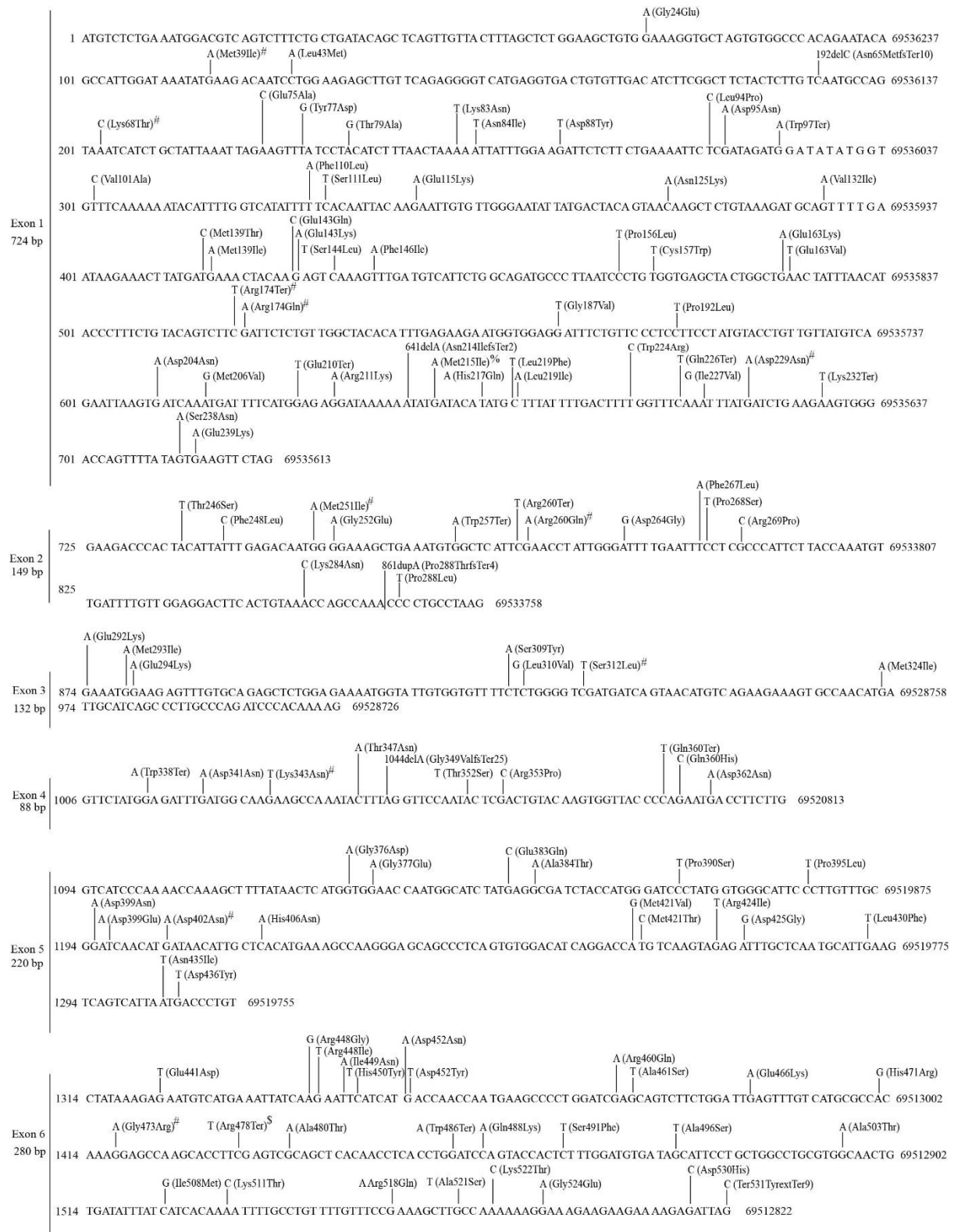
**Figure S13.** Mutations at the *UGT2B7* gene in TCGA tumors. Data shown is the coding region of the *UGT2B7* gene (6 exons, 1590 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001074.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001065.2) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times).



**Figure S14.** Mutations at the *UGT2B10* gene in TCGA tumors. Data shown is the coding region of the *UGT2B10* gene (6 exons, 1587 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001075.6, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001066.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice).

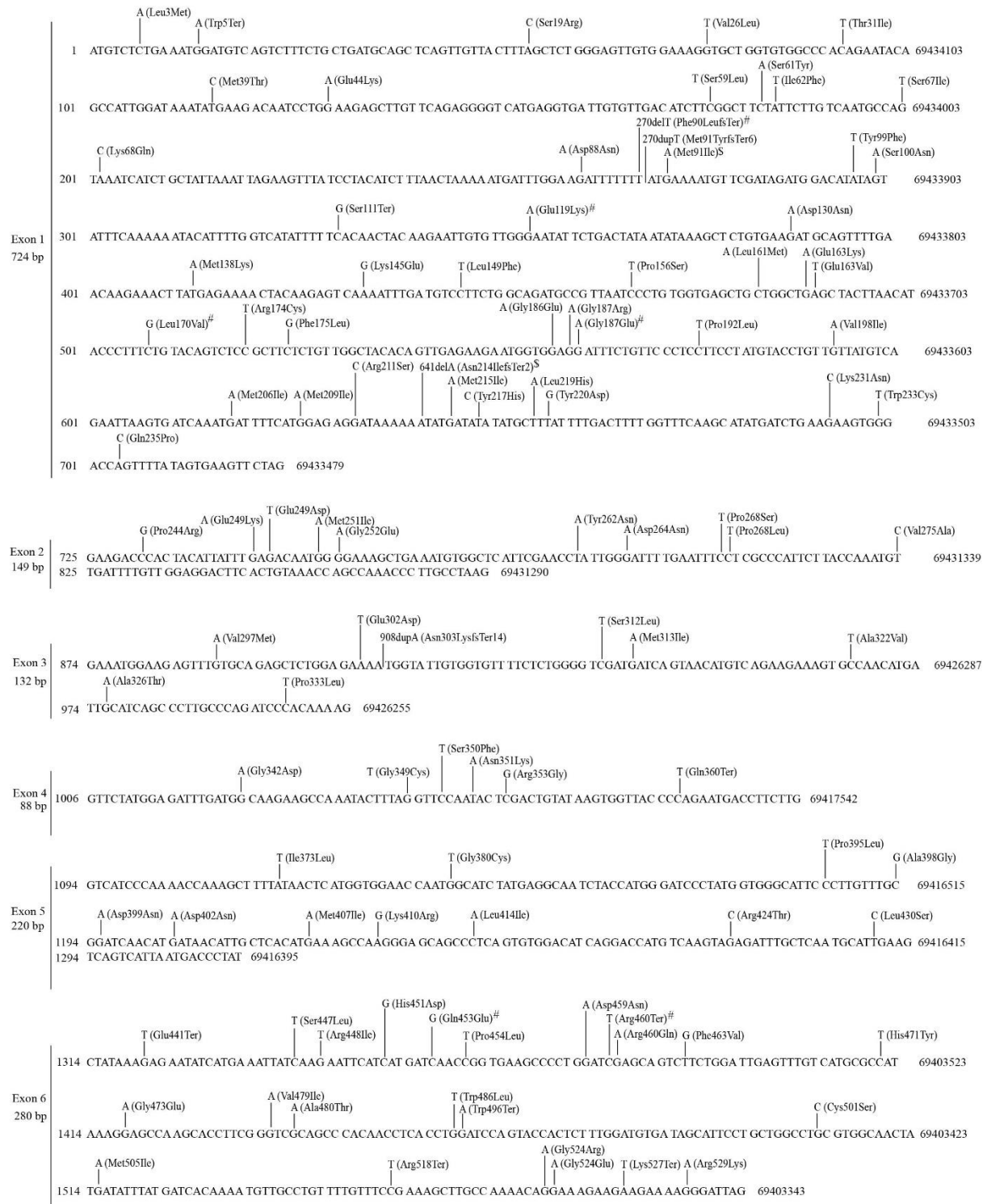


**Figure S15.** Mutations at the *UGT2B11* gene in TCGA tumors. Data shown is the coding region of the *UGT2B11* gene (6 exons, 1590 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001073.3, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001064.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), & (five times).

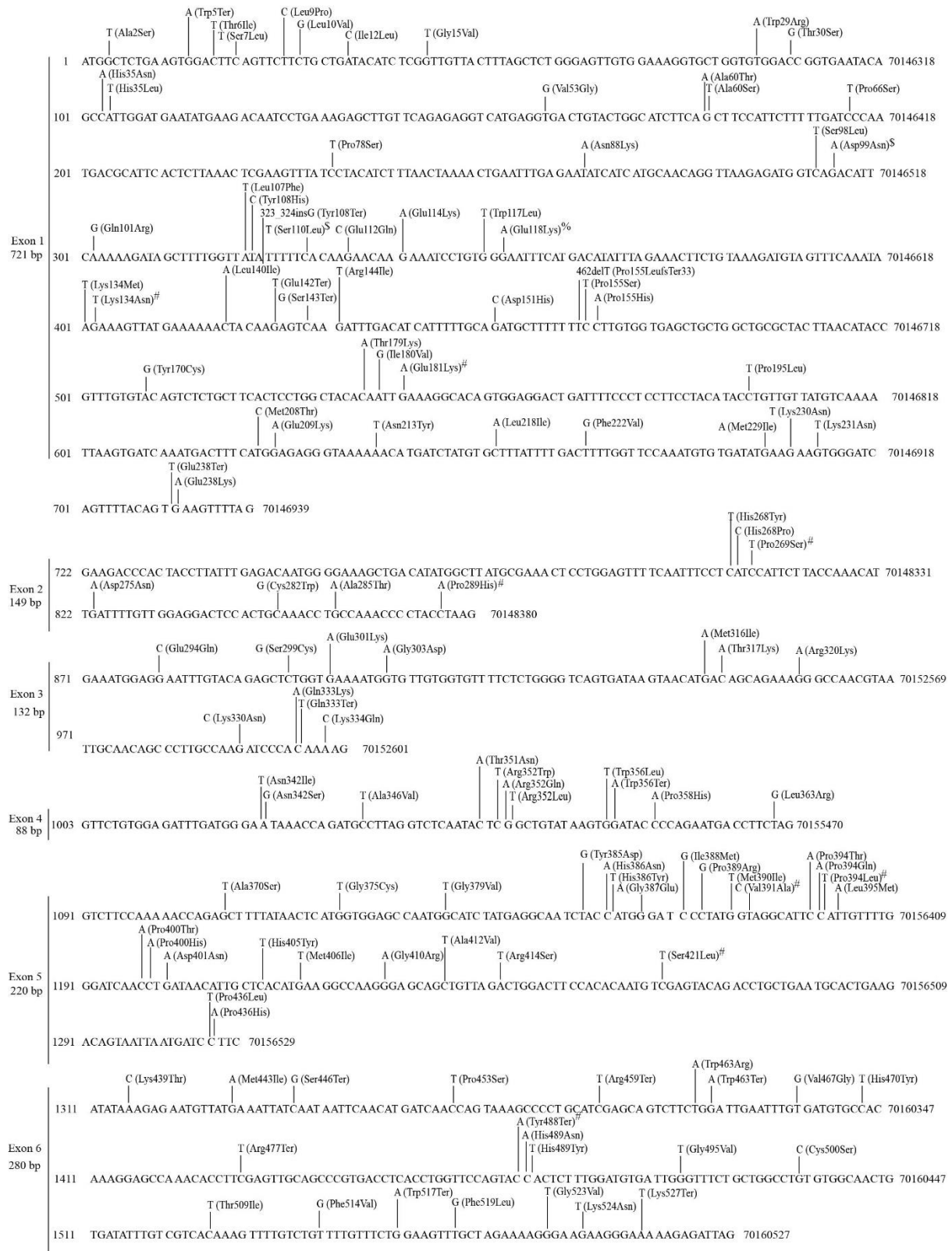


**Figure S16.** Mutations at the *UGT2B15* gene in TCGA tumors. Data shown is the coding region of the *UGT2B15* gene (6 exons, 1593 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001076.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001067.2) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), % (four times).





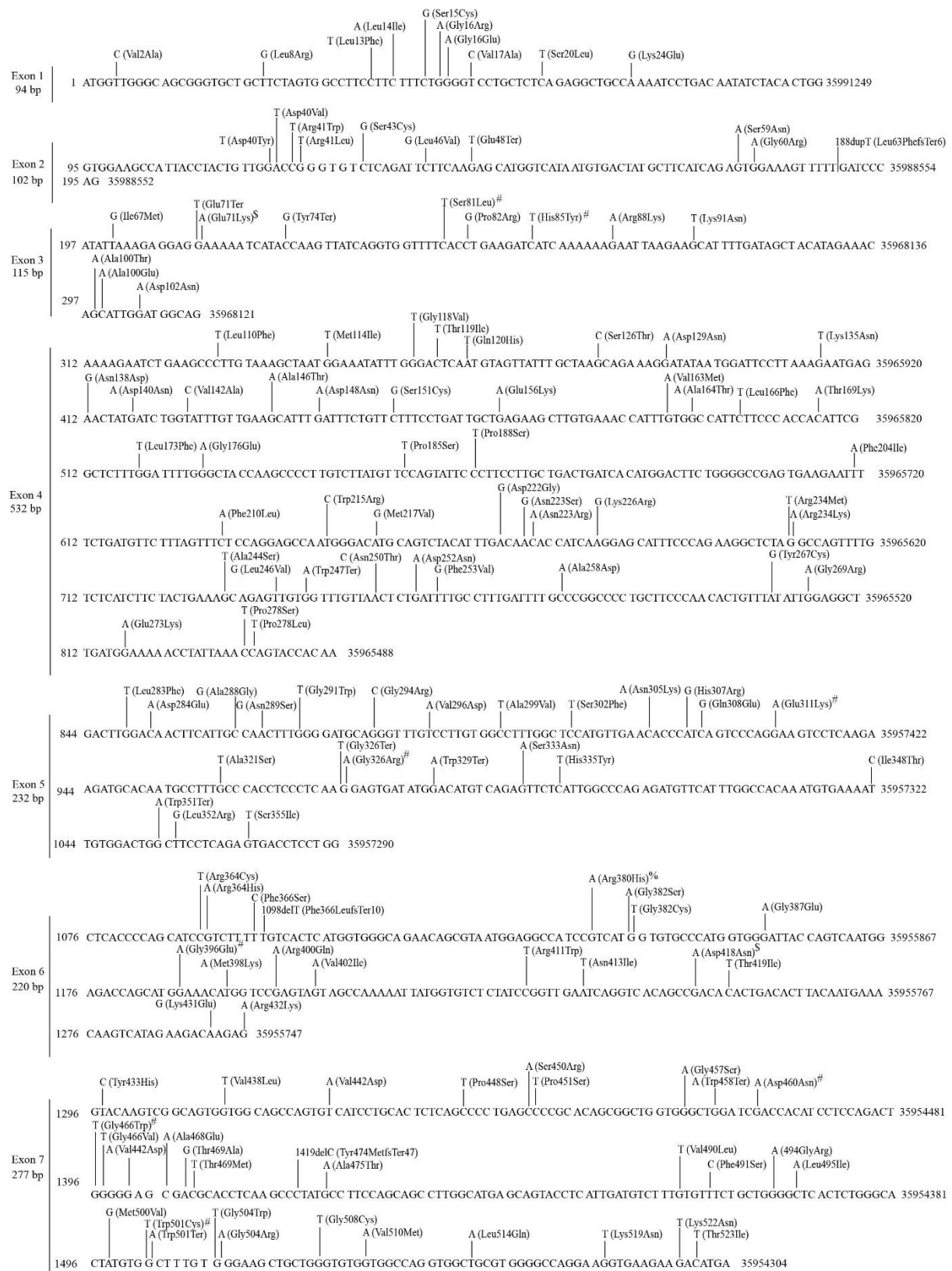
**Figure S17.** Mutations at the *UGT2B17* gene in TCGA tumors. Data shown is the coding region of the *UGT2B17* gene (6 exons, 1593 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001077.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001068.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times).



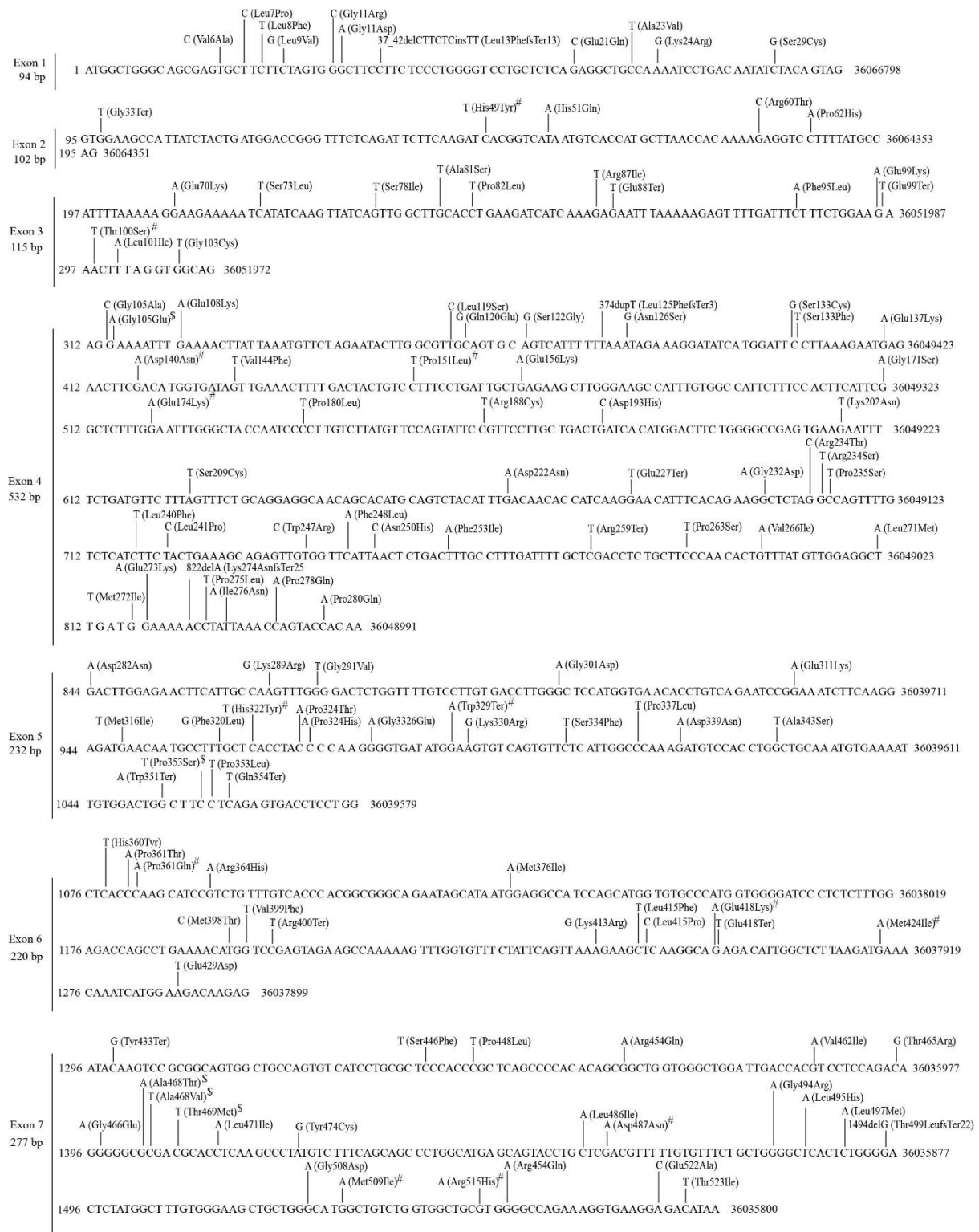
**Figure S18.** Mutations at the *UGT2B28* gene in TCGA tumors. Data shown is the coding region of the *UGT2B28* gene (6 exons, 1590 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_053039.2, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_444267.1) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), % (four times).

A		
1A1	atlaselfqrevtvqdllssasvwlfrsdfvkdypripmpnmvfvggincilhqnplsgef	290
1A3	aslaselfqrevsvvdilshasvwlfrgdfvmdypripmpnmvfiggincanrkplsgef	291
1A4	aslaselfqrevsvvdlsyasvwlfrgdfvmdypripmpnmvfiggincangkplsgef	291
1A5	aslaselfqrevsvvdlvshasvwlfrgdfvmdypripmpnmvfiggincangkplsgef	291
1A6	eelasavlkrdvdiitlyqkvswwlrydfvleyprpvmnpnmvfigginckkrkdlsgef	289
1A7	leiaselqtpvtaydlyshtsiwllrtdfvleyppkpvmpnmifigginchqgkplpmef	287
1A8	leiaselqtpvtaydlyshtsiwllrtdfvldypkpvmnpnmifigginchqgkplpmef	287
1A9	leiaselqtpvtaydlyshtsiwllrtdfvldypkpvmnpnmifigginchqgkplpmef	287
1A10	leiaselqtpvtaydlyshtsiwllrtdfvldypkpvmnpnmifigginchqgkplpmef	287
2B4	dqfysevlgrpttlsetmakadiwllrnywdfqfphpllpnvefvvgglhckpakplpkem	292
2B7	dqfysevlgrpttlsetmgkadwllrnswnfqfpypllpnvdfvvgglhckpakplpkem	292
2B10	dqfysevlgrpttlsetmrkadiwllrnswnfkfphpllpnvdfvvgglhckpakplpkem	291
2B11	dqfysevlgrpttlfetmgkadiwllrnswnsfqfphpllpnvdfvvgglhckpakplpkem	292
2B15	dqfysevlgrpttlfetmgkaemwllrtywdfefprpflpnvdfvvgglhckpakplpkem	293
2B17	dqfysevlgrpttlfetmgkaemwllrtywdfefprpflpnvdfvvgglhckpakplpkem	293
2B28	dqfysevlgrpttlfetmgkadiwllrnswnsfqfphpllpnidfvvgglhckpakplpkem	292
↑ 259 (UGT2B28)		
B		
2B4	ykwipqndllg 359	
2B7	ykwipqndllg 359	
2B10	ykwipqndllg 358	
2B11	ykwipqndllg 359	
2B15	ykwlpqndllg 360	
2B17	ykwlpqndllg 360	
2B28	ykwipqndllg 359	
1As	ykwipqndllg 354	
C		
2B4	vmrhkgakhlr 472	
2B7	vmrhkgakhlr 472	
2B10	vmrhkgakhlr 472	
2B11	vmrhkgakhlr 472	
2B15	vmrhkgakhlr 473	
2B17	vmrhkgakhlr 473	
2B28	vmchkgakhlr 472	
1As	vmchkgaphlr 471	

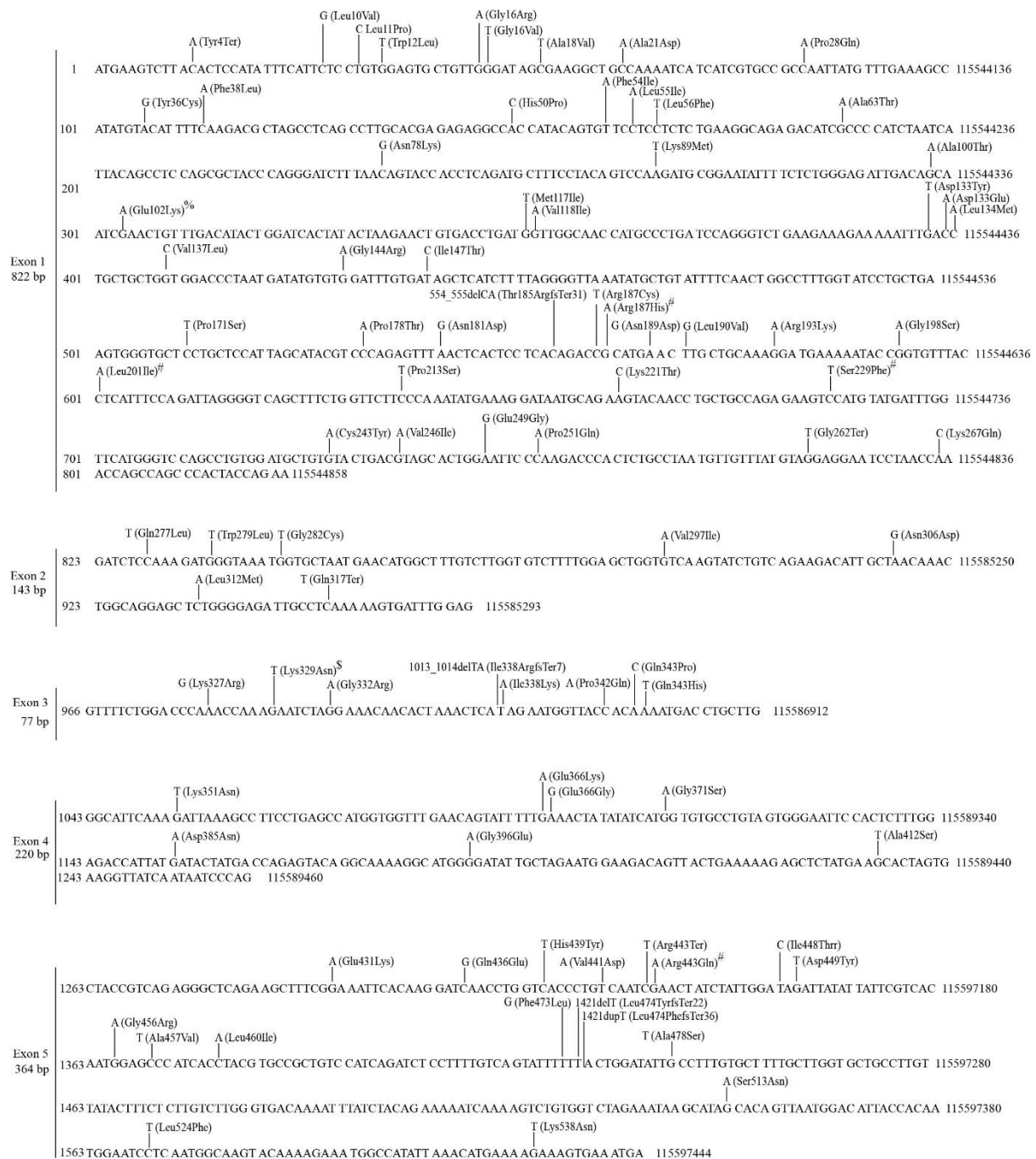
**Figure S19.** Mutations within the codons of conserved amino acids across the UGT1A and UGT2B family enzymes. The amino acid sequences of the UGT1A and UGT2B enzymes are aligned using the Clustal Omega program. Data shown are the sequence alignments surrounding the conserved amino acids [in Red, positions given at the bottom (A) or at the left (B, C)] whose codons were mutated in TCGA tumors. As indicated above the conserved amino acids, these mutations lead to 1) amino acid substitution (missense), 2) no change in amino acid sequence (silent mutation) or 3) premature stop codons (nonsense mutation, specified by \*).



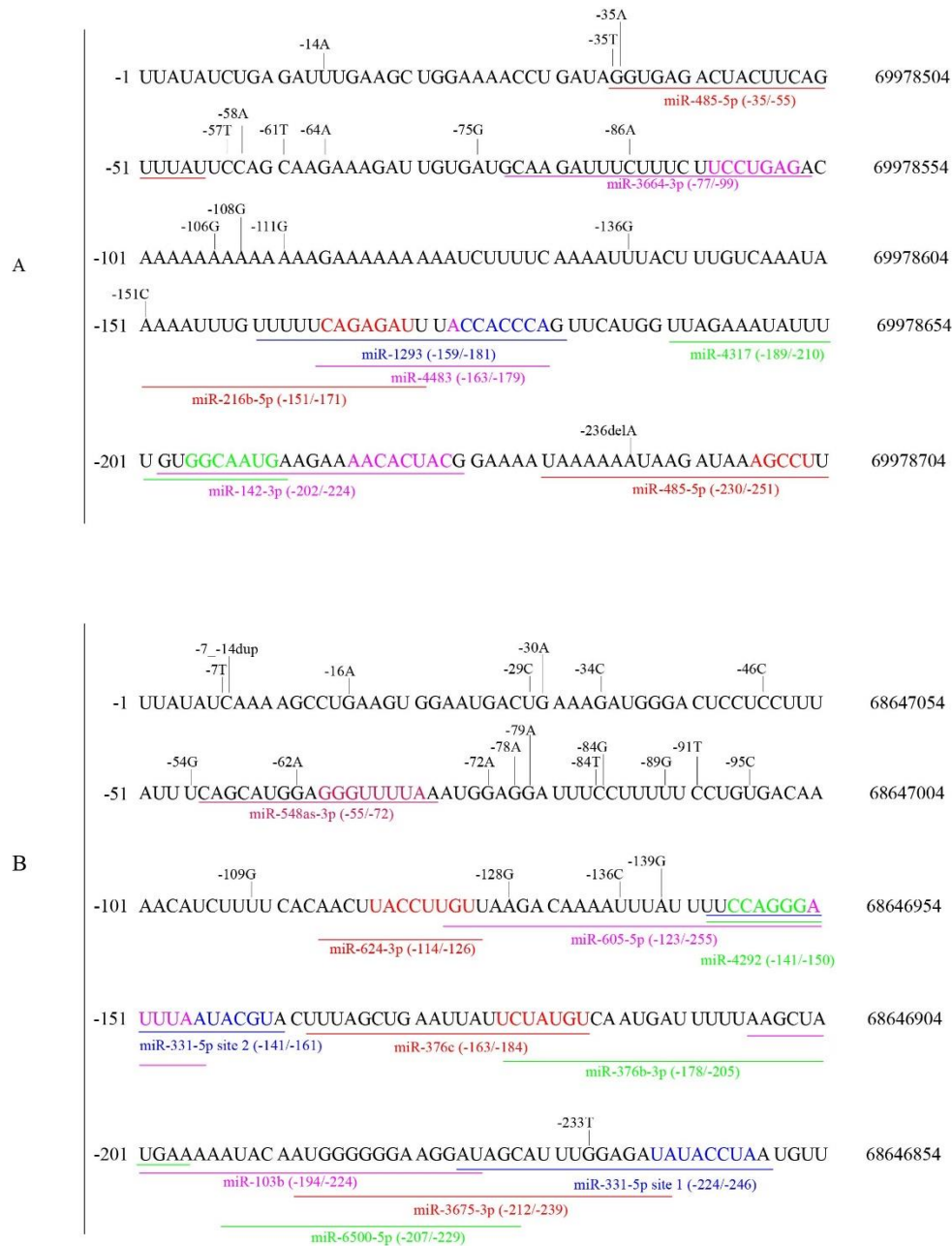
**Figure S20.** Mutations at the *UGT3A1* gene in TCGA tumors. Data shown is the coding region of the *UGT3A1* gene (7 exons, 1572 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_152404.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_689617.3) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), % (four times), & (five times).



**Figure S21.** Mutations at the *UGT3A2* gene in TCGA tumors. Data shown is the coding region of the *UGT3A2* gene (7 exons, 1590 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_174914.4, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_777574.2) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), % (four times).



**Figure S22.** Mutations at the *UGT8* gene in TCGA tumors. Data shown is the coding region of the *UGT8* gene (5 exons, 1590 bp) with genomic (GRCh37/hg19, right) and cDNA (NM\_001128174.3, left) positions. Mutations (missense, nonsense and small indels) and the resulting changes at the protein level (NP\_001121646.2) are indicated above the reference mRNA sequence. Recurrent mutations are indicated by # (twice), \$ (three times), % (four times).



**Figure S23.** Somatic mutations within/adjacent to known miRNA target sites at UGT 3'UTRs in TCGA tumors. Data shown are the first 250-nt sequence of the 3' UTR region for UGT2B7 (A) or UGT2B15 (B) with positions indicated at the cDNA (left) and genomic (GRCh37/hg19) (right) levels. Somatic mutations identified in TCGA tumors are given above the sequences. The sequences underlined and colored are known functional miRNA target sites (positions indicated in the BRACKETs) with the miRNA seed target sites highlighted in matched colors.