LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-207512

A comparison of somatic mutational spectra in healthy study populations from Russia, Sweden and USA

P. Noori, S. Hou, I. M. Jones, C. B. Thomas, B. Lambert

October 27, 2004

Carcinogenesis

Published in Carcinogenesis (2005) 26, 1138-1151

Deleted: Title page

A comparison of somatic mutational spectra in healthy study populations from Russia, Sweden and USA

Peri Noori[1], Saimei Hou[1], Irene M. Jones[2], Cynthia B. Thomas[2], Bo Lambert[1,3]

[1]Department of Biosciences, The Karolinska Institute, Novum, SE-14157 Huddinge, Sweden, [2]Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, L-441, P.O.Box 808. Livermore, CA 94550, USA

Short title: HPRT mutational spectra

[3] Corresponding author:
Bo Lambert
Department of Biosciences
The Karolinska Institute
SE-141 57 Huddinge
Sweden
Tel: +46-8-6089254
Fax: +46-8-6081501
E-mail: bo.lambert@cnt.ki.se

Key words: mutational spectra, *HPRT*, human, T-lymphocytes, *in vivo*

2 ABSTRACT

Comparison of mutation spectra at the hypoxanthine-phosphoribosyl transferase (*HPRT*) gene of peripheral blood T lymphocytes may provide insight into the aetiology of somatic mutation contributing to carcinogenesis and other diseases. To increase knowledge of mutation spectra in healthy people, we have analysed *HPRT* mutant T-cells of 50 healthy Russians originally recruited as controls for a study of Chernobyl clean-up workers (Jones et al. Radiation Res. 158, 2002, 424). Reverse transcriptase polymerase chain reactions and DNA sequencing identified 161 independent mutations among 176 thioguanine resistant mutants. Forty (40) mutations affected splicing mechanisms and 27 deletions or insertions of 1 to 60 nucleotides were identified. Ninety four (94) single base substitutions were identified, including 62 different mutations at 55 different nucleotide positions, of which 19 had not previously been reported in human T-cells. Comparison of this base substitution spectrum with mutation spectra in a USA (Burkhart-Schultz et al. Carcinogenesis 17, 1996, 1871) and two Swedish populations (Podlutsky et al, Carcinogenesis 19, 1998, 557, Podlutsky et al. Mutation Res. 431, 1999, 325) revealed similarity in the type, frequency and distribution of mutations in the four spectra, consistent with aetiologies inherent in human metabolism. There were 15-19 identical mutations in the three pair-wise comparisons of Russian with USA and Swedish spectra. Intriguingly, there were 21 mutations unique to the Russian spectrum, and comparison by the Monte Carlo method of Adams and Skopek (J. Mol. Biol. 194, 1987, 391) indicated that the Russian spectrum was different from both Swedish spectra (P=0.007, 0.002) but not different from the USA spectrum (P=0.07), when Bonferroni correction for multiple comparisons was made ($p < 0.008$ required for significance). Age and smoking did not account for these differences. Other factors causing mutational differences need to be explored.

3 INTRODUCTION

Although somatic mutagenesis is closely linked to carcinogenesis and other diseases, very little is known about the actual causes of mutation in normal human cells *in vivo*. Most mutations arise as a result of an error during replication or repair of a damaged DNA template. Many different endogenous metabolic processes as well as exogenous agents have the ability to produce DNA damage. Human somatic in vivo mutagenesis is probably driven by endogenous causes as well as environmental exposures, but to what extent one of these dominates over the other, is not well known.

Epidemiological studies have provided associations between human cancer morbidity and environmental exposures and life style, but many of these associations have not yet been mechanistically explained. Multiple mutations are implicated in carcinogenesis, and somatic mutagenesis is one possible link between environmental exposure and cancer disease. Results from a recent study of the influence of dietary factors on the frequency of somatic in vivo mutation provided a mechanistic support for a cancer-protective effect of vegetables and fruit by modulation of somatic mutagenesis. [1].

A mutation has a certain degree of specificity in the sense that it may bear the signature of the type of damage that induced it, be it a spontaneous mistake during normal DNA replication or repair, some endogenous metabolite, or an environmental chemical or radiation exposure. The spectrum of gene specific mutations in a tissue, i.e. the frequency distribution of different types of mutation along a defined nucleotide sequence in DNA, could provide information about the aetiology of mutations. Results from studies of the p53 tumour suppressor gene-specific mutational spectrum in various human tumours from different regions of the world have provided evidence of the environmental factors implicated in skin, liver and lung carcinogenesis [2]. Similarly, the spectrum of somatic mutation in somatic cells from individuals of a healthy population could serve as a useful *in vivo* marker of past and present exposure to genotoxic agents, and help to explain why some specific environmental factors are associated with increased cancer risk.

Several studies of *HPRT* (hypoxanthine-guanine phosphoribosyl transferase) gene mutations in human cultured cells and T-lymphocytes in vivo have provided

evidence for age, exposure and genetics to influence mutation frequency. An increased frequency of mutation (MF) with increasing age in normal, healthy people is generally observed (reviewed in [3,4]). Moreover, certain occupational exposures [5,6] and life style factors such as smoking (reviewed in [3,4]) have been associated with increased MF, while the intake of specific dietary items seems to protect from mutations and cancer [1]. Some inherited polymorphisms of genes involved in metabolism and DNA repair have also been shown to influence the *HPRT* MF (for example [5-9]).

In order to study the possible influence of environmental and life style factors on somatic mutagenesis, we have identified 161 *HPRT* mutations in T-lymphocytes in a population of 50 healthy Russians, and compared the base substitutions in this Russian spectrum with previously established spectra of base substitution mutations in populations from USA [10] and Sweden [11,12].

4 MATERIALS & METHODS

4.1 Subjects.

The Russian study population has been described previously [13,14]. All individuals are males. The recruitment of subjects, questionnaire and obtaining of samples were reviewed by Institutional Review Boards at all institutions involved and all subjects gave informed consent prior to participation in the study. The study population originally included two groups from Russia; one who had potential to receive low levels of radiation exposure while serving as “Clean-up workers” during the clean-up of the Chernobyl Nuclear Power Plant accident, and one comprising a set of healthy individuals selected among friends and relatives of the clean-up workers, but who had not themselves been exposed to radiation or involved in the clean-up work. Only the control group was included in the present study. Information on age and smoking history was available for most subjects as reported in a self-administered questionnaire described elsewhere [13,14].

The USA and Swedish populations, which are included for comparisons, have also been described previously. The USA population comprised healthy smokers and non-smokers from the Raleigh-Durham area of North Carolina [10]. The Sweden 1

population comprised a group of healthy garage workers, laboratory personnel and fine mechanics, all non-smokers. The garage-workers had been occupationally exposed to diesel exhausts, and as a result had increased levels of aromatic DNA adducts, but no increase of *HPRT* mutant frequency compared to the laboratory personnel and fine mechanics who served as controls [6]. The Sweden 2 population were collected as healthy controls for lung cancer patients, and comprised smokers as well as non-smokers. The mean age of this population was older than any of the other populations [15]. All individuals included in the two Swedish spectra were either working or living in the county of Stockholm. The composition of the populations with regard to age gender and smoking habits are shown in Table 1.

4.2 Experimental procedures

The collection and shipping of blood samples, the cell culture methods and the determination of *HPRT* MF, as well as the expansion and storage of mutant clones for molecular analysis have been described in detail previously [13,14,16,17]. In brief, blood samples were collected and shipped in vacutainers, and were received in California within 2-4 days of being drawn in Russia. Mononuclear cells were isolated and cultured immediately in growth medium (RPMI 1640 containing the mitogen phytohemagglutinin (PHA), 5% fetal bovine serum and antibiotics), then counted and plated in round-bottomed wells in medium supplemented with T-cell growth factors with or without 6-thioguanine (1µg/ml) for determination of *HPRT* MF.

For each donor about 15 thioguanine resistant clones were expanded by dilution from one well to 48-96 wells, depending upon the size and vigour of the clone in the initial well. All wells of a clone were pooled and harvested after 7 to 28 days. An aliquot of the cells was lysed for DNA analysis and the rest frozen in a controlled freezing chamber and stored in liquid nitrogen.

The mutants were analysed for *HPRT* deletions with PCR-based methods aiming at detection of retention, loss or change of eight gene fragments containing the exons and flanking intron sequences of genomic *HPRT* DNA. In the population of healthy controls who have been further analysed in the present work, approximately 19% of mutants were found to contain a genomic *HPRT* deletion [17]. Mutant clones, which showed ”no detectable change” in the deletion analysis and contained sufficient amount of DNA for

further study, were selected for analysis of point mutations by reverse transcriptase-PCR and DNA sequencing methods. Frozen aliquots of these clones in DMSO medium were sent in two separate shipments in dry ice containers to Stockholm in 1999, approximately four years after they had been collected and stored. Upon arrival, samples in one of the shipments were partly melted, and few of these clones were suitable for analysis. The second shipment was in good condition. The cell pellets were thawed, washed in 5 ml 1 x PBS, diluted with 4 ml PBS and redistributed into 4 tubes. One tube was used directly for RNA isolation, while the 3 other tubes were put into the 80 ^{0}C freezer.

RNA was isolated with a Purescript kit (GENTRA Systems). cDNA synthesis was for 1.5 hr at 37 ^{0}C in 4 µl M-MLV RT x 5 buffer (Promega) containing 500 µM of each dNTP (Promega), 1.6 µM of reverse primer Y3, 1 U/µl RNAsin (Pharmacia) and 2.5 U/µl M-MLV reverse transcriptase (Promega). Reverse transcriptase-PCR was carried out as described [12], except that biotinylated primers were not used. The nested PCR product was cleaned up using MicroSpin Column (S-400 HR) (Amersham Pharmacia Biotech).

Cycle sequencing was performed with Big Dye v.2 (Applied Biosystems): 4 µl Big Dye Terminator mix, 2 µl of 5 x sequencing buffer, 1 µl of 20 µM primer and 11 µl nested PCR product (approximately 400 ng of double stranded cDNA) was added to 20 µl of water. Cycle sequencing was run at 96 ^{0}C 3 min, 96 ^{0}C 10 sec, 50 ^{0}C 5 sec and 60 ^{0}C 4 min, for 25 cycles. The PCR product was cleaned by ethanol precipitation. Sequencing and primers were as in Podlutsky at al. [12]. The reaction was run on a 377A Automated Sequencer (Applied Biosystems), and the sequences were analysed using Sequence Navigator and Edit View (Applied Biosystems).

4.3. Data analysis

The probability that two or more base substitutions in a set of random mutations would occur at the same site was calculated using the Poisson distribution and the Bonferroni correction to account for multiple comparisons [18]. The calculations are based on the assumptions that (i) all observed mutations are independent (ii) there are 300 mutable sites in the *hprt* coding sequence [11,12,19]. For a set of 94 simple base substitutions, as in the present work, the probability of observing 5 or more mutations at any single site is < 0.006,

while 4 mutations in one position yields a P-value of 0.09. For the compiled spectrum of 382 mutations, the probability of observing 8 or more mutations in one position is < 0.025.

Mutational spectra in the three study populations were compared with the Monte Carlo method of Adams and Skopek [18] and the program described by Cariello et al. [20]. Two spectra were compared at a time, and positions showing no mutations were not used in the calculations. All P-values were based on 30000 iterations. A P-value of <0.05 means that the spectra are different in a pair-wise comparison, but since six comparisons were made, a Bonferroni correction for multiple comparisons with a corrected significance level of 0.0083 (0.05/6) was applied.

5 RESULTS

5.1 Background data for the Russian study population

The healthy Russian population studied in the present work was originally selected as a control population in a multi-endpoint study of genetic biomarkers in peripheral blood samples of workers who participated in the clean-up work after the nuclear power accident in Chernobyl in 1986. Most of the blood samples were collected in 1994-1995. The results of the multi-endpoint study have been published in several reports [13,14]. *HPRT* MF and *HPRT* gene deletions were two of the biomarkers used, and they have been reported on in separate publications as well [17,21,22].

The aim of the present study was to compile a mutational spectrum of single base substitution mutations for comparisons with previously established mutational spectra in healthy populations from USA [10] and Sweden [11,12]. The *HPRT* mutant clones studied in this work included only those which had shown ”no detectable change” in the previous deletion analysis [17]. Moreover, since smoking is one factor that may influence the mutational spectrum, the intention was to include equal numbers of smokers and non-smokers. Due to exhaustion of material, and problems during shipping, storage and thawing of the samples, these criteria were not fully achieved. As shown in Table 1, mutant clones from 25 current smokers and 17 non-smokers were included. Three former smokers and five donors for whom smoking data were missing were also included to bring the total number of donors to 50. The average age and *HPRT* MF in this subset of Russian controls were similar to the original, complete control population. Within the

present study population, smokers and non-smokers showed similar means for age and *HPRT* MF (Table 1).

5.2 Numbers and types of mutations

A mutation was identified in 176 clones from the 50 individuals studied. In some donors, more than one mutant clone with the same mutation was identified. Identical mutations in different clones from one donor may be replicates of one original mutation, or they may represent separate, unique mutational events. Since no attempts were made to further characterize these mutants, each distinct mutation was counted only once per individual in the spectrum analysis, but the mutant clones with identical mutations are all listed in Tables 2 - 5. Discounting these ”clonal replicates”, there were a total of 161 mutations (Table 2). The mean number of mutations per donor was 3.26. In 46 donors, between 1 and 5 mutations were collected. In 4 donors, the number of mutations/donor ranged from 7 to 17 (Fig 1).

As shown in Table 2, the distribution of different types of mutations were similar in smokers and non-smokers. Single base pair substitution in the coding region comprised 58% of the mutations, while mutations affecting splicing accounted for 25%, and 16% were small deletion/insertion. The latter type of mutation is likely to be underrepresented in comparison with the other types of mutation in Table 2, since deletions/insertions and other rearrangements had already been screened for in genomic DNA from mutants of these subjects and excluded from the present analysis. Thus, the deletions/insertions detected in cDNA are small changes that have escaped the band-shift analysis of genomic DNA in the previous study [17]. The relative frequency of missense and nonsense mutations (Table 2) was the same in non-smokers, smokers and the group of formers smokers and individuals with unknown smoking status ($P = 0.9$).

5.3 Mutations affecting splicing functions

In 44 mutants (including 4 possible clonal replicates), the *HPRT* cDNA had either lost one or several exons, or contained intron sequences indicating mutations affecting the splicing functions (Table 3). In most of these mutants, a single exon was missing. Two mutants had a duplication of exon 2+3, and exon 6 respectively. In several mutants,

cryptic splice sites were used in intron 1 and 5 and in exon 8 and 9. All of these types of mutation have been described previously [23,24], and no attempts were made to further characterize the underlying change in genomic DNA. One mutant appeared to have two independent mutations; in addition to the loss of exon 6 there was also a 5-nucleotide deletion in exon 2 (Table 3, 1052ns/90). This mutant is also listed among the deletion/insertions in Table 4.

5.4 Deletion/insertion mutations

In 28 mutants (including one possible clonal replicate), the *HPRT* cDNA was found to contain small deletions and insertions ranging from 1 to 60 nucleotides, which are all listed in Table 4. Two mutants contained two changes, one was classified as compound, and the other one as complex. The first one (1052ns/90) showed loss of the entire exon 6 in addition to a 5-nucleotide deletion in exon 2; two apparently independent changes, as already mentioned above. The complex mutation (Table 4, 1379u/50) comprised one base substitution and one dinucleotide deletion separated by 5 nucleotides in exon 7. Each of these changes is likely to give rise to a TG-resistant phenotype. The base substitution, 488T>C, predicts a change of residue 162Leu>Ser, and has previously been reported in the Human HPRT Mutation Database [25], and in T-cells in vivo [10]. The dinucleotide deletion gives rise to a stop codon in the 9th codon downstream from the deletion point. However, since these two changes are located so close to each other, it is most likely that they are part of the same complex mutational event.

Of the remaining 26 mutants, 13 were ±1 nucleotide deletions, and 13 were deletions of 4-60 nucleotides (Table 4). There were 9 deletions of 1 nucleotide as compared to 4 insertions of one nucleotide, and both types of change were more common among smokers than non-smokers. There were no deletion/insertion mutations in exon 1 and 5, which are the two shortest exons, but as many as 10 (36% of all) were in exon 2, which is twice as many as expected from a random distribution according to exon length. The other exons showed between 1 and 5 mutations each, with a distribution close to expectation. Thus, it seems that exon 2, and especially the 5´ half of the exon, is a region particularly prone to undergo deletion/insertion mutagenesis, as also observed and discussed in our previous work [26,27].

Interestingly, there were two identical mutations in exon 8, with a breakpoint within the hypothetical palindromic sequence spanning nucleotide positions 533-557, that previously was pointed out as a possible hotspot mutational hotspot region in lung cancer patients [28].

5.5 Base pair substitutions

In total, 94 single base pair substitutions were identified; 50 transitions and 44 transversions comprising 62 different substitutions at 55 different nucleotide positions. There were 85 missense and 9 nonsense mutations, all of which are listed in Table 5 with position, type of change, sequence context and predicted amino acid change. Different types of substitution were observed at seven positions; 3G >A or C, 74C > G or T, 131A > G or T, 136A > G or C, 197G > A or T, 368C > G or A, 606G > C or T.

Surprisingly many of these mutations were novel in the sense that they have not previously been reported to occur. Out of the 62 different base substitutions in Table 5, as many as 8 (13%) are new additions to the 279 single base pair substitutions that are included in the human HPRT Mutation Data Base [25] (annotated as "new" in Table 5), and as many as 21 (34%) are not included among the 169 different kinds of base substitutions that we have reported previously in human T-cells in vivo [10-12] (annotated as "new T" in Table 5). However, two of these 21 mutations (Table 5, 104T>A and 533T>G) are included among 48 different kinds of base substitutions that were detected in a study of T-cell mutations in Russian twins [29].

One of the 8 new mutations, 430C>T, creates a stop codon, 143Gln>Term. The other seven mutations are all missense: two (130G>C, 614T>A) occur in positions where different kinds of base substitutions had been reported previously, and 5 (136A>C, 136A>G, 410T>G, 479T>A, 487T>G) occur in positions were no base substitution mutations have been reported before. When these results are added to the data base of presently known human HPRT mutations, the number of positions in the 657 bp coding region of the human HPRT gene that can give rise to missense or nonsense mutation by single nucleotide substitution amounts to 292, which corresponds to 44% of the total number of nucleotides. Although this is one further step towards saturation, the HPRT

mutational spectrum is not yet completed, since there are still 14 possible nonsense mutations that have not been reported so far.

It is also interesting that, among the new mutations, there were some that changed one of the first or last two nucleotides of an exon, and still produced cDNA with no signs of exon skipping. These mutations were the second nucleotide of exon 3 (136A>G, and 136A>C), the second nucleotide of exon 7 (487T>G) and the first nucleotide of exon 8 (533T>G). The absence of a splicing effect of other mutations in the first or last position of an exon, such as the last nucleotide of exon 2 (del134G in), the first nucleotide of exon 6 (403G>C), and the first (del610C) and second (611A>G) nucleotide of exon 9 have been discussed before [24].

The types of base substitutions are summarized in Table 6. Transitions were predominating over transversions among non-smokers while the reverse was true among smokers (current, ex or others). Although the difference between smokers and non-smokers was not significant (P = 0.2 for subtype distribution and P = 0.1 for all transitions vs. all transversions), this finding confirms our previous observations of a higher frequency of GC>TA and AT>CG transversions, and lower frequency of GC>AT transitions among smokers than non-smokers [10,12].

5.6 Comparison of base substitutions mutations in Russian, USA and Swedish populations

In previous work we have reported on *HPRT* mutational spectra in two healthy Swedish populations [11,12] and one USA population [10]. Summary data for these study populations are shown in Table 1. The spectra of base substitution mutations from these studies can be compared with the present Russian mutational spectrum (Table 7). The USA spectrum comprises 94 mutations representing 66 different base substitutions in 57 positions. The Swedish spectra comprise 87 and 107 mutations, each representing 62 different base substitutions in 54 and 53 positions, respectively. Thus, with regard to size and complexity, these three previous spectra are similar to the present Russian spectrum, with 94 independent mutations representing 62 different base substitutions at 55 positions (Table 7)

The comparative analysis of mutational spectra considers numbers and kinds of mutations as well as the positions at which the mutations occur. Figure 2A and 2B shows

the distribution of mutated nucleotide positions in the *HPRT* cDNA in the four different spectra, and Fig. 3 shows the mutations in detail.

There seems to be an overall agreement between the spectra, in that there is a non-random distribution of the mutations, with (1) relatively few mutations between bases 5 – 108 (with the exception of position 74C), 223 – 367, and 618-657, (2) several apparent hotspot positions to which mutations from all of the spectra contribute. The overall impression is that of a considerable overlap in the distribution of mutations between these four spectra from different parts of the world. However, there are also obvious differences in the type and distribution of mutation between the spectra.

In total, there are 382 mutations at 126 different positions in the four spectra together, but only 11 positions are mutated in all spectra, and only five site specific mutations occur in all four spectra (Figures 2 and 3). The site specific mutations are 3G>A, 143G>A, 197G>A, 508C>T and 617G>A. The 11 common positions that are mutated in all four spectra are 119G>T or A, 208G>A or T, 464C>G or T, 539G>A or T, 568G>A or T or C, and 611A>T or G, plus the 5 mentioned above.

In the compiled spectrum of 382 single base pair substitutions, any position with 8 or more mutations is significantly different from a random distribution (see Materials and Methods). There are eight such sites in the compiled spectrum, 3G, 74C, 143G, 146T, 197G, 508C, 611A and 617G (Figure 2). All but two of these hotspot positions are mutated in all of the three country spectra. The exceptions are 74 C (with 9 mutations, but none in the Sweden 1 spectrum) and 146T (with 10 mutations, but none in the Russian spectrum). In the separate spectra, five mutations are needed for a position to qualify as a hotspot (see Materials and Methods). In the Russian spectrum, positions 197G and 617G are significant hotspots with 8 and 5 mutations each. In the USA spectrum, 197G, 617G and 508C are hotspots [10]. Hotspots in the Sweden 1 spectrum are 146T and 197G, and in the Sweden 2 spectrum 143G, 197G, and 617G. Thus, there is a considerable overlap between the four spectra with regard to mutational hotspots.

In contrast, there are several positions that seem to distinguish one spectrum from the other. In the Russian spectrum, there are four 118G>A transitions and four 611A>T transversions. The latter mutation was also detected in a study of mutants recovered from Russians by Curry et al. [29]. None of these mutations occur in any of the other spectra. In the USA spectrum, there are two transitions (C>T) and two transversions (C>G) at

551C, a position that is not mutated in any of the other spectra. In the Sweden 1 spectrum, there are five 146T>C transitions, a significant hotspot that does not occur in any of the other spectra. In the Sweden 2 spectrum, there are 11 transitions 143G>A, again a significant hotspot, which shows at most two mutations of the same kind at this position in the other spectra. Thus, these mutations could possibly be related to more regional environmental exposures. Analyses of spectra overall presented below appear to rule out age and smoking status as being responsible for the differences between the Sweden 1 and Sweden 2 spectra.

Between 27 and 44% of the mutations that occur in each spectrum are unique, in the sense that they occur only in that spectrum and not in any of the others. The Russian spectrum contains more of unique mutations and unique mutated positions than any of the other spectra (Table 7A). The degree of similarity between two spectra was evaluated by counting overlapping mutations, i.e. mutations that are identical in a pair-wise comparison of two spectra. As shown in Table 7B, there is more overlap of mutations between the USA-spectrum and the three other spectra, and less overlap of the two Swedish spectra and the Russian spectrum.

To study whether the Russian, USA and Swedish mutational spectra are similar or not, data were subjected to statistical analysis, using the Monte Carlo method devised by Adams and Skopek [18]. This statistical method accounts for the location and nature of the mutation, and is suitable for the comparison of mutational spectra with limited amount of data [20,30]. The results are shown in Table 8. The Russian spectrum was found to be significantly different from the two Swedish spectra, with P-values of 0.007 and 0.002. In contrast, no statistically significant differences were obtained between the Russian and USA spectra, or the Swedish and USA spectra. However, the two Swedish spectra were statistically significant from each other, with a P-value of 0.001 (Table 8). Additional analyses of subgroups in the two Swedish study populations were performed to investigate the contribution of age and smoking to mutation spectra. The spectrum of mutations in smokers within the Sweden 2 population was not different from that in non-smokers of similar age in the same population (P=0.62, N=107), and the spectrum of mutation in the older non-smoking subjects in the Sweden 2 population was not different from the spectrum of mutation in the young non-smoking subjects of the Sweden 1

population (P=0.17, N=137). Thus, age and smoking did not account for the difference between spectra of these two Swedish populations.

6 DISCUSSION

The present results add 161 independent mutations to the database of human somatic *HPRT* mutation *in vivo*. Overall, the distribution, types and frequencies of these mutations in a healthy, Russian population have many features in common with earlier observations in other populations of various types of *HPRT* mutation in T-cells, with regard to splicing mutations, small deletions and base substitutions. The types of splicing mutations include exclusion of exons as well as inclusion of intron sequences, and exon duplications, all of which have been described previously (review in [24]). The breakpoints for the small deletions are frequently associated with short repeats or monotonous base sequences, indicating that they are formed as a result of a slippage mechanism, and there is a cluster of breakpoints in the first half of exon 2, which is in accord with earlier observations [26] [27]. The single base substitutions show a non-random distribution, with several hotspots at positions that coincide with earlier observations in healthy control populations [10-12].

The comparison of the mutational spectra between the three populations from Russia, USA and Sweden revealed considerable similarities, both with regard to the overall distribution of mutations, and the hotspot positions. The particular strength with this comparison, in contrast to earlier analyses of mutational spectra compiled *in silico* from many smaller data sets produced in various laboratories (e.g. [4]), is that all the data were obtained in two laboratories with similar criteria for selection of study population and analytical procedures and methods.

The extensive overlap between the three mutational spectra from different parts of the world strongly supports the view that a great part, perhaps a predominating part, of the mutations are caused by endogenous factors inherent to human physiology and metabolism, rather than by some more or less specific life style factors or environmental exposures. For example, positions 197 and 617 were hotspots in 4 or 3 of the populations, respectively, whose spectra were compared in this study. At both positions, G to A

transitions predominated four-fold over G to T transversions. The existence of these hotspots and the specific mutations at them suggest a consistency of frequency of damage formation, misrepair and/or misreplication across the populations, phenomena to which the shared GTGTGT sequence context may contribute. Hence, our results indicate that 197 and 617 are hotspots for multiple mutagenic events, and in multiple populations, as also observed by others ({Cariello, 1993 #190} {Cariello, 1994 #192}{Ma, 2000 #245}). This result is consistent with the sequence (context) having features that invite more damage and/or poorer repair and/or replication errors.

On the other hand, the presence of some frequently mutated positions or significant hotspots in one but not the other spectra, e.g. positions 118G, 143G, 146T, 551C and 611A, may reflect the influence of modulating factors or perhaps represent direct fingerprints of some specific environmental exposures to mutagenic agents. At the DNA level such agents would be expected to induce sequence specific damage formation for which there is potential for misrepair or misreplication. The large number of sites in the *HPRT* coding sequence that inactivate the HPRT protein {Duan, 2004 #244} {Podlutsky, 1998 #196} {Cariello, 1993 #190} enhances the potential to detect differences in spectra.

In view of the diversity of these spectra, and the wide distribution of mutations along the *HPRT* coding sequence, with as many as 127 mutated positions, and 37 of these showing more than one kind of base substitution, it is of interest both that some pair-wise comparisons did not reveal differences of spectra despite differences among study groups for age and smoking, and others revealed differences of spectra despite similarities in these variables. There was no statistical difference between either the USA spectrum and the Russian spectrum, or the USA spectrum and the two Swedish spectra. In contrast, the two Swedish spectra and the Russian spectrum were significantly different, even when applying the Bonferroni correction factor for multiple comparisons. Part of the explanation for this apparent "discrepancies" is that the statistical method used for analysis of mutational spectra (18, 20) compares only two spectra at a time and only at base positions where mutations have occurred in at least one of the two groups. For instance, the two Swedish spectra may not share so many base positions, but each of the them could have a considerable number of positions that are mutated in the US spectrum as well. In this context, it is of interest that nither position 197 nor 617 is frequently

substituted among inherited mutations in Lesch-Nyhan patients ({Jinnah, 2000 #246}). Seeking explanations for both similarities and differences in mutation spectra is necessary for understanding the strengths and limitations of assessing mutation spectra in addition to analysing covariates for mutant frequency. This quest will be complex, requiring new statistical tool for identifying the elements in mutation spectra that make them distinctive, continued research into the variables that affect the frequency and nature of mutations at specific sequences, and ongoing iteration between epidemiological detection of key exposures and the detailed mechanisms of somatic mutagenesis.

Possible explanations for the differences between spectra could be age and/or smoking, factors which differed more between the Swedish and Russian populations, than between the Russian and USA populations (see Table 1). However, neither of these factors seemed to explain the difference between the two Swedish populations. The spectrum of mutations in smokers within the Sweden 2 population was not different from that in non-smokers of similar age in the same population, and the spectrum of mutation in the older non-smoking subjects in the Sweden 2 population was not different from the spectrum of mutation in the young non-smoking subjects of the Sweden 1 population. Thus, age and smoking did not seem to influence the spectrum of base substitution mutations in these two Swedish populations, which is agreement with previous observations in other studies that were not able to detect significant effects of smoking or age on the *HPRT* mutational spectrum in T-cells [4,10]. The apparent lack of influence of age and smoking on the mutational spectrum could be attributable to similarity in the net mutagenic effect of two different but highly complex exposures associated with aging and smoking. A number of agents can lead to the same mutation, as discussed below.

The difference between the Swedish and Russian mutational spectra are not primarily associated with the prominent and significant hotspot positions, but with the occurrence of new and unique mutations at sites with low mutation frequency. As mentioned above, in the Russian spectrum with 94 mutations, there were many new mutations, and many that were not detected in either the USA or the Swedish spectra. Also Curry et al [29] observed many new *HPRT* mutations in T-cells from Russian twins. These authors [29] also found the spectrum of Russian Twin mutations to be significantly different from an age-matched Western mutant dataset, that included data from Burkhart-Schultz et al [10] and Podlutsky et al [11] used in the present work. However, the spectra

studied by Curry et al [4,29], included not only base substitutions, but also other categories of mutation, which make the result difficult to interpret. When the 55 single base substitution mutations in Russian twins reported by Curry et al [29] were compared with the present Russian set of 94 mutations, the two spectra turned out to be significantly different with a P-value of < 0.0001 (Table 9B). Moreover, The Russian Twin spectrum of Curry et al [29] was significantly different also from the USA-spectrum (P< 0.0001, N=149) and the Sweden 1 and Sweden 2 spectra (P=0.002, N=142 and P<0.0001, N=162 respectively). The 6 sets of Russian twins studied by Curry et al [29] may have had distinct exposures that contribute to these differences.

In conclusion, these comparisons of mutational spectra of single base substitutions in the *HPRT* gene in T-cells of populations from Russia, USA and Sweden have demonstrated an overall similarity in the type, frequency and distribution of mutations. The results suggest that most mutations are induced by mechanisms that are inherent to human metabolism and little influenced by differences in life style or environmental exposures. In this regard the *HPRT* mutation spectra results are consistent with analysis of mutations in genes associated with development of cancer [31,32]. However, similarities of mutation spectra do not rule out differences in aetiology. The same mutation may be induced by multiple distinct events, some associated with endogenous agents, others with exogenous agents. For example, G > T transversions may be associated with 8-hydroxyguanine, an oxidatively damaged base arising from normal metabolism or from oxidative agents in cigarette smoke, or with exogenous exposures to a variety of agents that produce bulky adducta such as polycyclig aromatic hydrocarbons in cigarette smoke, or aflatoxin B1 (reviewed in [33]). Such an overlap in spectrum may explain the increase in *HPRT* mutant frequency associated with smoking despite little difference in mutation spectrum.

The differences in mutation spectra, which do exist in spite of the overall similarities, are intriguing (Fig. 3). A minor fraction of the mutations are certainly caused by factors that are different between Russia and Sweden, as well as between populations living in the same region within Sweden. Smoking does not seem to be involved in causing this difference, since the Russian spectrum did not differ from the spectrum in the USA population, with similar smoking habits and tobacco products as in Sweden. Age, another well documented factor that is associated with an increase in the frequency

of *HPRT* mutations in T-cells, is also not likely to be responsible for the difference, because there was no statistical difference between the mutational spectra of the non-smokers in the two Swedish populations with more than 25 years difference in mean age. Thus, other factors causing mutational differences between populations need to be explored.

7 ACKNOWLEDGEMENTS

The contribution of Dr. Andrej Podlutsky in the initial phase of this work is gratefully acknowledged. The Swedish Cancer Society provided financial support through grants to B.L. and S.H. This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48 with support from the National Institutes of Health for CA59431 (IMJ). Work performed in Russia was in part supported by the Russian Ministry of Science and the Russian Ministry of Health, in collaboration with Pavel Pleshanov, Irini Vorobstova and Ludmila Tureva, with assistance of Vladimir Knyajev, Ministry of Science and Technology Policy of Russia.

8 REFERENCES

1. Nyberg, F., Hou, S.-M., Pershagen, G. and Lambert, B. (2003) Influence of diet on the mutant frequency at the hypoxanthine-guanine phosphoribosyltransferase (*hprt*) locus in T-lymphocytes. *Carcinogenesis*, **24**, 689-696.
2. Hussain, S.P. and Harris, C.C. (1999) p53 mutation spectrum and load: the generation of hypotheses linking the exposure of endogenous or exogenous carcinogens to human cancer. *Mutat Res.*, **428**, 23-32.
3. Cole, J. and Skopek, T.R. (1994) International Commission for Protection Against Environmental Mutagens and Carcinogens. Working paper no. 3. Somatic mutant frequency, mutation rates and mutational spectra in the human population in vivo. *Mutat. Res.*, **304**, 33-105.
4. Curry, J., Larissa Karnaoukhova, G.C.G. and Glickman, B.W. (1999) Influence of sex, smokiung and age on human *hprt* mutation frequencies and spectra. *Genetics*, **152**, 1065-1077.
5. Hou, S.-M., Fält, S. and Steen, A.-M. (1995) *HPRT* mutant frequency and GSTM1 genotype in non-smoking healthy individuals. *Environmental and Molecular Mutagenesis*, **25**, 97-105.
6. Hou, S.-M., Lambert, B. and Hemminki, K. (1995) Relationship between *hprt* mutant frequency, aromatic DNA adducts and genotypes for GSTM1 and NAT2 in bus maintenance workers. *Carcinogenesis*, **16**, 1913-1917.
7. Viezzer, C., Norppa, H., Clonfero, E., Gabbani, G., Mastrangelo, G., Hirvonen, A. and Celotti, L. (1999) Influence of GSTM1, GSTT1, GSTP1, and EPHX gene polymorphisms on DNA adduct level and *HPRT* mutant frequency in coke-oven workers. *Mutat. Res.*, **431**, 259-269.
8. Vodicka, P., Soucek, P., Tates, A.D., Dusinska, M., Sarmanova, J., Zamecnikova, M., Vodickova, L., Koskinen, M., de Zwart, F.A., Natarajan, A.T. and Hemminki, K. (2001) Association between genetic polymorphisms and biomarkers in styrene-exposed workers. *Mutat. Res.*, **482**, 89-103.

9. Grant, D.J., Hall, I.J., Eastmond, D.A., Jones, I.M. and Bell, D.A. (2004) Bilirubin UDP-glucuronosyltransferase 1A1 (UGT1A1) (2004) gene promoter polymorphisms and HPRT, glycophorin A, and micronuclei mutant frequencies in human blood. *Mutat. Res.*, **2004**, 1-10.
10. Burkhart-Schultz, K.J., Thompson, C.L. and Jones, I.M. (1996) Spectrum of somatic mutations at the hypoxanthine phosphoribosyltransferase (*hprt*) gene of healthy people. *Carcinogenesis*, **17**, 1871-1883.
11. Podlutsky, A., Österholm, A.-M., Hou, S.-M., Hofmaier, A. and Lambert, B. (1998) Spectrum of point mutations in the coding region of the hypoxanthine-guanine phosphoribosyltransferase (*hprt*) gene in human T-lymphocytes in vivo. *Carcinogenesis*, **19**, 557-566.
12. Podlutsky, A., Hou, S.M., Nyberg, F., Pershagen, G. and Lambert, B. (1999) Influence of smoking and donor age on the spectrum of in vivo mutation at the *hprt* locus in T-lymphocytes of healthy adults. *Mutation Res.*, **431**, 325-339.
13. Moore, D.H., James, H., Tucker, D., Jones, I.M., Langlois, R.G., Pleshanov, P., Vorobtsova, I. and Jensen, R. (1997) A study of the effects of exposure on cleanup workers at the Chernobyl nuclear reactor accident using multiple end points. *Radiation Research*, **148**, 463-475.
14. Jones, I.M., Galick, H., Kato, P., Langlois, R.G., Mendelsohn, M.L., Murphy, G.A., Pleshanov, P., Ramsey, M.J., Thomas, C.B., Tucker, J.D., Tureva, L., Vorobtsova, I. and Nelson, D.O. (2002) Three somatic genetic biomarkers and covariates in radiation-exposed Russian clean-up workers of the Chernobyl nuclear reactor, 6 – 13 years after exposure. *Radiat. Res.*, **158**, 424-442.
15. Hou, S.-M., Yang, K., Nyberg, F., Hemminki, K., G, P. and Lambert, B. (1999) *HPRT* mutant frequency and aromatic DNA adduct levels in nonsmoking and smoking lung cancer patients and population controls. *Carcinogenesis*, **20**, 437-444.
16. Jones, I.M., Thomas, C.B., Tucker, B., Thompson, C.L., Pleshanov, P., Vorobtsova, I. and Moore, D.H. (1995) Impact of age and environment on somatic mutation at the *hprt* gene of T lymphocytes in humans. *Mutation research*, **338**, 129-139.
17. Jones, I.M., Thomas, C.B., Haag, K., Pleshanov, P., Vorobstova, I., Tureva, L. and Nelson, D.O. (1999) Total gene deletions and mutant frequency of the *HPRT* gene as

indicators of radiation exposure in Chernobyl Liquidators. *Mutation Research*, **431**, 233-246.

18. Adams, W.T. and Skopek, T.R. (1987) Statistical tests for the comparison of samples from mutational spectra. *J. Mol. Biol.*, **194**, 391-396.
19. Cariello, N.F. and Skopek, T.R. (1993) Analysis of mutations occurring at the human *hprt* locus. *J. Mol. Biol.*, **231**, 41-57.
20. Cariello, N.F., Piegorsch, W.W., Adams, W.T. and Skopek, T.R. (1994) Computer program for the analysis of mutational spectra: application for p53 mutations. *Carcinogenesis*, **15**, 2281-2285.
21. Thomas, C.B., Nelson, D.O., Pleshanov, P., Vorobstova, I., Tureva, L., Jensen, R. and Jones, I.M. (1999) Elevated frequencies of hypoxanthine phosphoribosyltransferase lymphocyte mutants are detected in Russian liquidators 6 to 10 years after exposure to radiation from the Chernobyl nuclear power plant accident. *Mutation Research*, **439**, 105-119.
22. Thomas, C.B., Nelson, D.O., Pleshanov, P. and Jones, I.M. (2002) Induction and decline of *HPRT* mutants and deletions following a low dose radiation exposure at Chernobyl. *Mutation Research*, **499**, 177-187.
23. Österholm, A.-M. and Hou, S.-M. (1998) Splicing mutations at the *HPRT* locus in human T-lymphocytes in vivo. *Environmental and Molecular Mutagenesis*, **32**, 25-32.
24. O'Neill, J.P., Rogan, P.K., Cariello, N. and Nicklas, J.A. (1998) Mutations that alter RNA splicing of the human *HPRT* gene: a review of the spectrum. *Mutation Res*, **411**, 179-214.
25. Cariello, N.F. (1994) Software for the analysis of mutations at the human *hprt* gene (The 6th version of the human *HPRT* mutation database, released in August 1996, Mutabase Software Durham, NC.). *Mutation Res.*, **312**, 173-185.
26. Österholm, A.-M., Bastlová, T., Meijer, A., Podlutsky, A., Zanesi, N. and Hou, S.-M. (1996) Sequence analysis of deletion mutations at the *hprt* locus of human T-lymphocytes: association of a palindromic structure with a breakpoint cluster in exon 2. *Mutagenesis*, **11**, 511-517.

27. Burkhart-Schultz, K.J. and Jones, I.M. (1997) Deletion and insertion in vivo somatic mutations in the hypoxanthine phosporibosyl transferase (*hprt*) gene of human T-lymphocytes. *Environ. Mol. Mutagen*, **30**, 371-384.

28. Hackman, P., Hou, S.-M., Nyberg, F., Pershagen, G. and Lambert, B. (2000) Mutational spectra at the hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) locus in T-lymphocytes of nonsmoking and smoking lung cancer patients. *Mutation research*, **468**, 45-61.

29. Curry, J., Khaidakov, M. and Glickman, B.W. (2000) Russian mutational spectrum differs from that of their Western counterparts. *Human Mutation*, **15**, 439-446.

30. Cariello, N.F., Douglas, G.R., W, N.J.G.H.D., Wilson, J.D. and Soussi, T. (1998) Databases and software for the analysis of mutations in the human p53 gene, human *hprt* gene and both the lacI and lacZ gene in transgenic rodents. *Nuclei Acid Research*, **26**, 198-199.

31. Jackson, A.L. and Loeb, L.A. (2001) The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat Res*, **477**, 7-21.

32. Thilly, W.G. (2003) Have environmental mutagens caused oncomutations in people? *Nature Genet.*, **34**, 255-259.

33. Hemminki, K., Koskinen, M., Rajaniemi, H. and Zhao, C. (2000) DNA adducts, mutations, and cancer 2000. *Regul Toxicol Pharmacol*, **32**, 264-75.

34. Edwards, A., Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H. and Caskey, C.T. (1990) Automated DNA sequencing of the human hprt locus. *Genomics*, **6**, 593-608.

9 FIGURE LEGENDS

Figure 1

The distribution of the 161 independent mutations among the 50 individuals.

Figure 2

Spectra of base substitution mutations in the coding region of the *HPRT* gene in T-lymphocytes from populations in Russia, USA and Sweden.
Russian data are from this work. Data for the USA population are from [10], for Sweden 1 from [11], and for Sweden 2 from [12]. There are a total of 382 mutations; 94 from the Russian population, 94 from the USA population, 87 from Sweden 1 and 107 from Sweden 2 population. No distinction has been made between mutations of different kinds at one and the same site. The numbers for the positions in *HPRT* cDNA starts from A in the first ATG. For practical reasons, only mutated positions are shown, and the sequence has been divided in two parts at the border between exons 3 and 4.

Figure 3

Detail spectrum of single base substitutions in the *HPRT* coding region in T-lymphocytes in populations from Russia, USA and Sweden. Russian data are from this work. Data for the USA population are from [10], for Sweden 1 from [11], and for Sweden 2 from [12]. All different kinds of mutation are shown at all mutated positions. Exon borders are marked by a slash. The position is indicated for every 10th base below the sequence. Where 2 or more mutations of the same kind occurs at one site, this is indicated by the substituted base followed by the number of mutations, e.g. A6 means that six independent substitution to adenine have been recorded at that particular site.

Figure 1
The distribution of the 161 independent mutations among the 50 individuals.

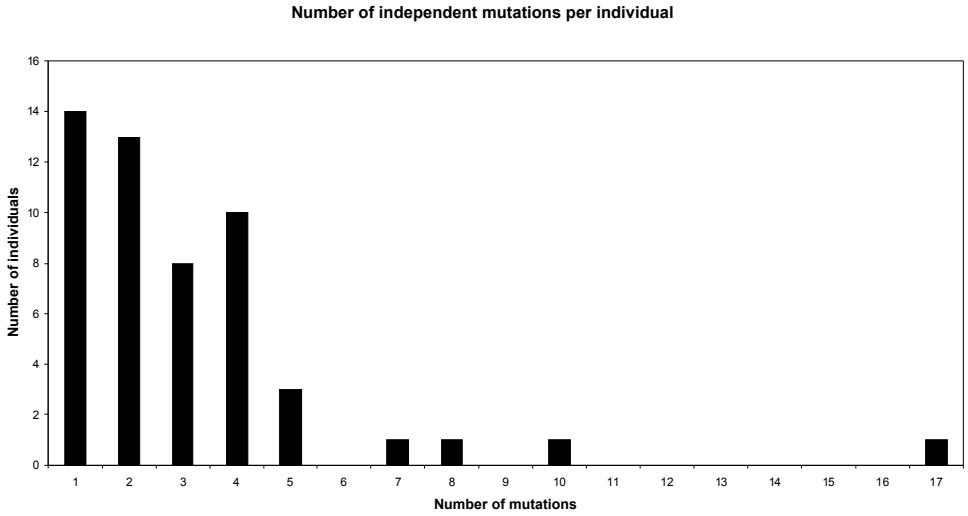
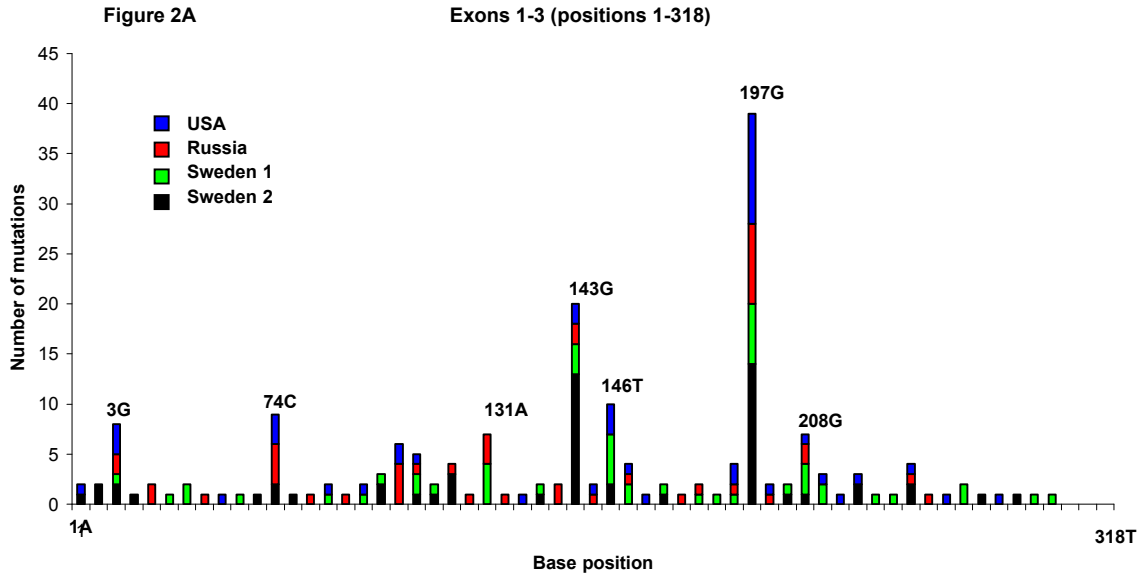


Figure 2.
Spectra of base substitution mutations in the coding region of the *HPRT* gene in peripheral blood T-lymphocytes in populations from Russia, USA and Sweden.



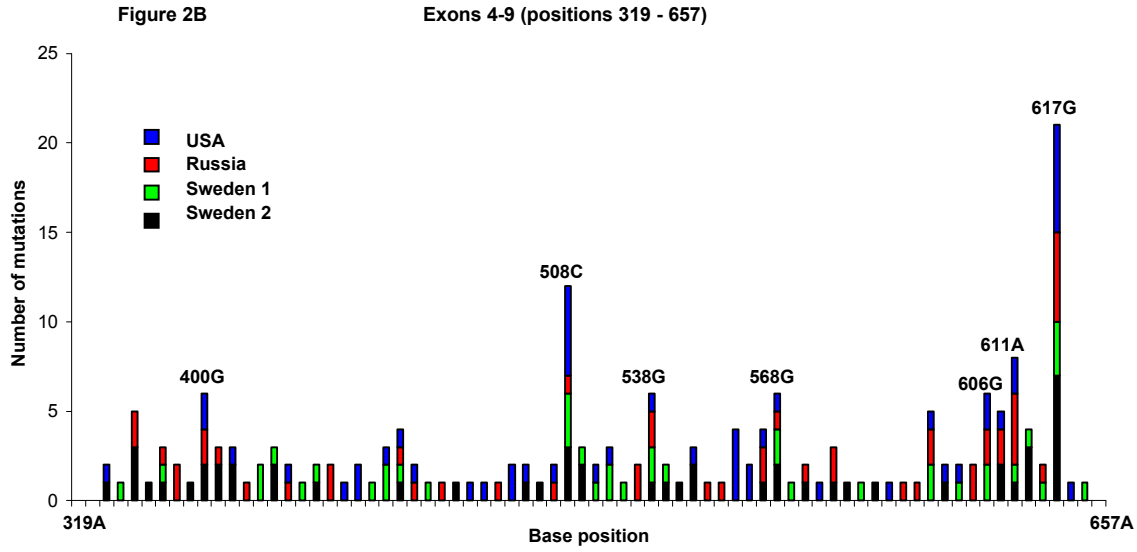


Figure 3

Base substitutions in T-cells in vivo in populations from Russia (red), USA (blue) and Sweden (Podlutsky 98 green and Podlutsky 99 black) superimposed on the human *HPRT* coding sequence. Exon borders are marked by a slash. The position is indicated for every 10th base below the sequence. Where 2 or more mutations of the same kind occurs at one site, this is indicated by the substituted base followed by the number of mutations, e.g. A6 means that six independent substitution to adenine have been recorded at that particular site.

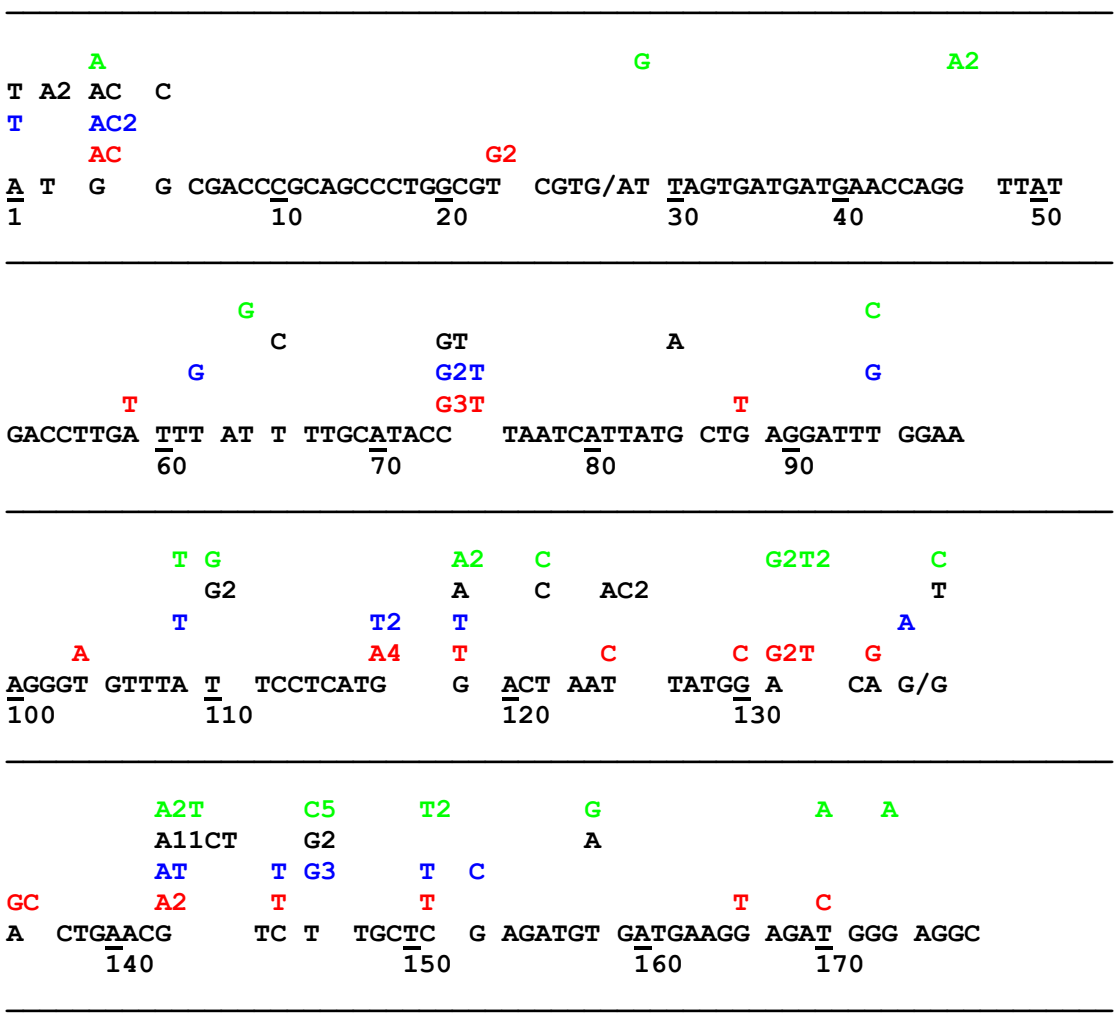


Fig 2B, continued

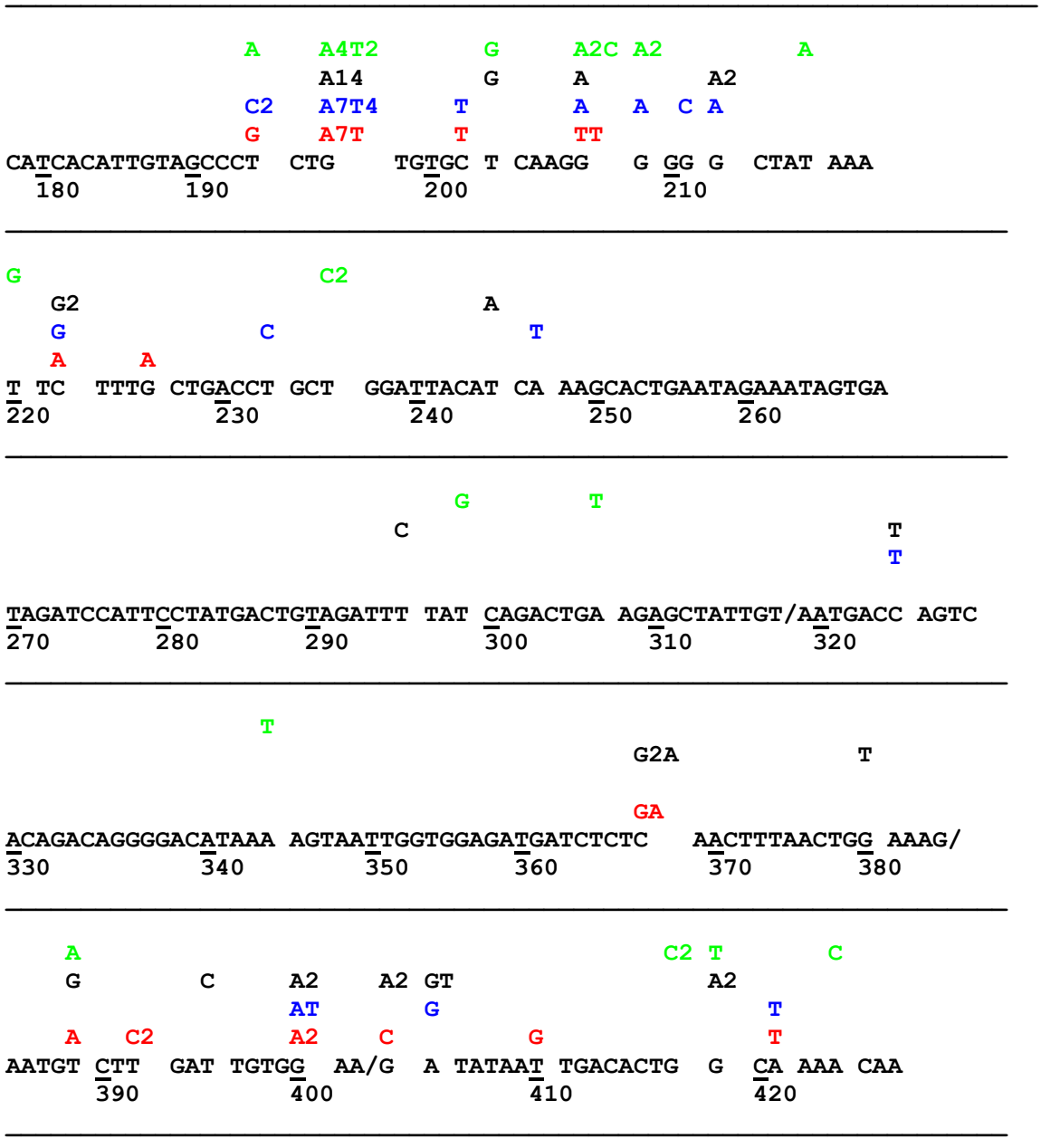


Fig 2B, continued