**Supplementary Table S1: Details of the samples included in the study**

**G:\work\NAR_paper\figures\Supplementary Table 1.tif**

NA: Not Applicable, \* Indicates tissue collection time, # FC- Frontal cortex, CC- Corpus Callosum (Brain Samples were collected from NIMHANS Brain Bank, Bangalore, India)

**Supplementary Table S2: Genotype concordance between somatic sites called by Varscan 2 and Mutect for brain samples**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample** | **No. of somatic sites called by Varscan 2** | **Concordance between Varscan and Mutect calls** | **Percentage concordance** |
| Brain\_152 | 32 | 9 | 28.1 |
| Brain\_156 | 132 | 104 | 78.8 |
| Brain\_171 | 90 | 59 | 65.6 |
| Brain\_202 | 117 | 81 | 69.2 |

**Supplementary Table S3: Genotype concordance between Exome sequence and Infinium genotyping**

G:\work\NAR_paper\figures\Supplementary Table 2.tif

**Supplementary Table S4: Comparison of Hi-Seq and Mi-Seq data for the selected loci**

G:\work\NAR_paper\figures\Supplementary Table 3.tif

The rows marked in red indicate the validated loci

**G:\work\NAR_paper\figures\Supplementary Figure 1.tif**

**Supplementary Figure S1:** Overall pipeline followed for data analysis

G:\work\NAR_paper\figures\Supplementary Figure 2.TIF

**Supplementary Figure S2:** Average read depth of all the samples: Blue bars represent average read depth for all the brain samples and red bars represent the same for blood-saliva samples.

G:\work\NAR_paper\figures\Supplementary Figure 3.tif

**Supplementary Figure S3: Number of germline variations per sample pair:** Variations with same genotype in both tissues from same sample are termed as germline variations, hence one bar for each sample pair. Blue bars represent brain samples and red bars represent blood saliva samples

G:\work\NAR_paper\figures\Supplementary Figure 4.tif

**Supplementary Figure S4:** Distribution of somatic variations (A) and germline variations (B) across different genomic regions (others: Downstream, upstream, intergenic, splicing, exonic, intronic, UTR’s of non- coding RNAs)

G:\work\NAR_paper\figures\Supplementary Figure 5.tif

**Supplementary Figure S5**: Distribution of synonymous and non-synonymous variations in all blood-saliva samples A) for somatic variations and B) for germline variations

G:\work\NAR_paper\figures\Supplementary Figure 6.tif

**Supplementary Figure S6**: Distribution of amino acid changes in total somatic variations and only G:C>T:A somatic variations.

G:\work\NAR_paper\figures\Supplementary Figure 7.tif

**Supplementary Figure S7:** Base change distribution in somatic (A) and germline variations (B) of blood-saliva samples

G:\work\NAR_paper\figures\Supplementary Figure 8.tif

**Supplementary Figure S8:** Distribution of genotypes for variations A) for somatic variations (excluding G:C>T:A changes) in brain B) for all classes of germline variations in brain samples C) somatic variation in blood-saliva samples

G:\work\NAR_paper\figures\Supplementary Figure 9.tif

**Supplementary Figure S9: The neighbouring sequence context for G>T transversions**: A) A total of 124 sequences of 11 nucleotides each were used to generate the frequency plot. For each sequence the ‘G’ in the middle was found to have a somatic transversion (G>T) in our data. The motif shows clear enrichment of ‘A’ immediately 3’ to the ‘G’. B)From the reference genome sequence (Chromosome 1, hg19) 11 contiguous nucleotides were randomly selected with a ‘G’ in the middle (6th) position. A total of 1000 sets were made keeping 200 sequences of 11 nucleotides in each set and frequency plots were generated for each of these set of 200. The figure shows a representative plot. Unlike the upper panel here the bias for 3’ A is not observed.