Cell Reports, Volume *28*

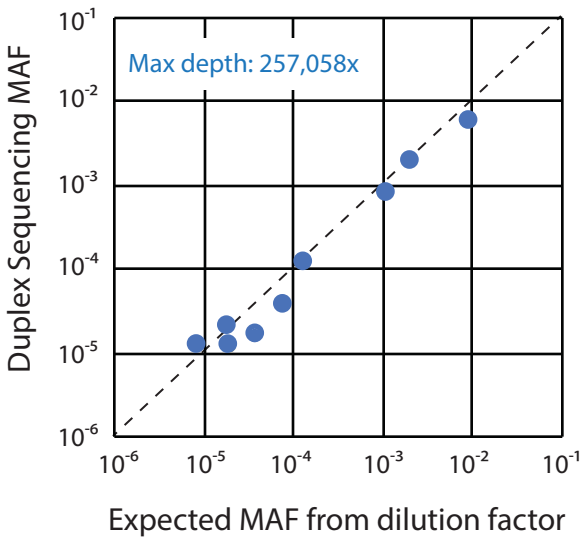
Supplemental Information

Ultra-Sensitive *TP53* Sequencing for Cancer Detection Reveals Progressive Clonal Selection in Normal Tissue over a Century of Human Lifespan

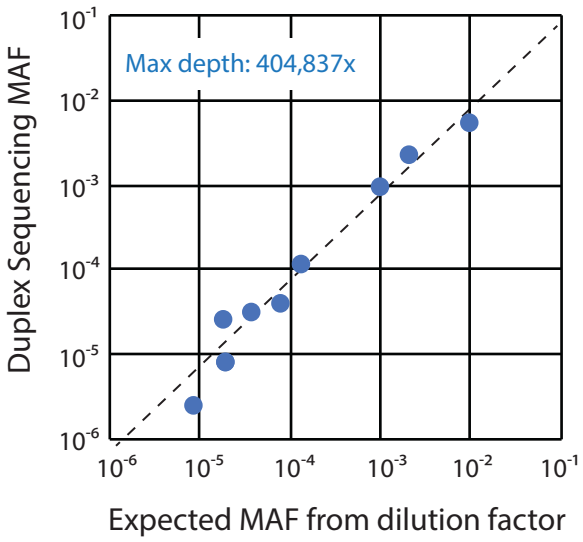
Jesse J. Salk, Kaitlyn Loubet-Senear, Elisabeth Maritschnegg, Charles C. Valentine, Lindsey N. Williams, Jacob E. Higgins, Reinhard Horvat, Adriaan Vanderstichele, Daniela Nachmanson, Kathryn T. Baker, Mary J. Emond, Emily Loter, Maria Tretiakova, Thierry Soussi, Lawrence A. Loeb, Robert Zeillinger, Paul Speiser, and Rosa Ana Risques

SUPPLEMENTAL FIGURES

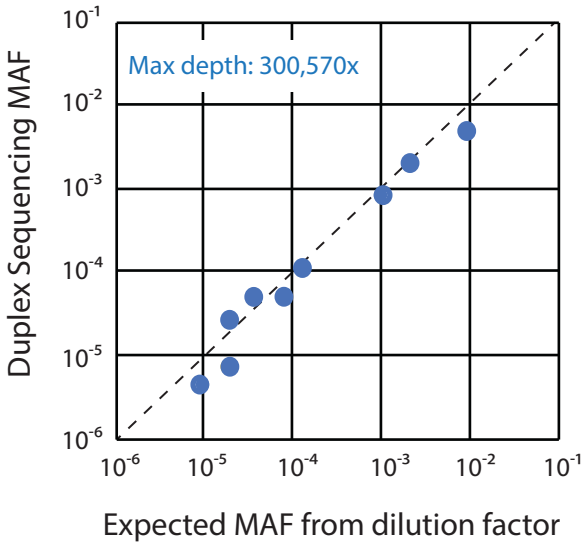
A.



B.



C.



D.

Gene	Mutation	Dilution
TP53	Q136fs	1:100
TP53	R249S	1:500
TP53	R248Q	1:1,000
NRAS	Q61R	1:5,000
TP53	G245S	1:10,000
NRAS	Q61L	1:25,000
TP53	R175H	1:50,000
KRAS	G12D (het.)	1:50,000
KRAS	G12V	1:100,000

Figure S1. Duplex Sequencing spike-in test of reproducibility and accuracy [related to Fig. 1B-E]. (A-C) DNAs from 9 cell lines, each carrying one or more unique single nucleotide mutation in *TP53* or a *RAS* gene, were serially diluted into DNA obtained from the peripheral blood of a healthy 26 year old male donor at levels ranging from 1:100-1:100,000. This single mixture was divided into three portions, with each prepared into Duplex libraries on three different days and sequenced on three independent runs. The total Duplex depth achieved approximately 1-million-fold (i.e. a million independent genome equivalents). All mutations were detected in all runs, with R^2 values of measured vs. expected MAFs respectively being 0.98, 0.96 and 0.95. Larger MAF variations among replicates at higher dilutions reflects the greater impact of stochastic Poisson sampling with lower frequency mutations. (D) Table of each spike-in mutation and the dilution factor used to generate the mixture. Expected MAF values for each data point in A-C were uniformly adjusted from the dilution factor listed in D by known copy number variation/zygosity at the loci of interest, as determined by Duplex Sequencing of pure DNA from each cell line.

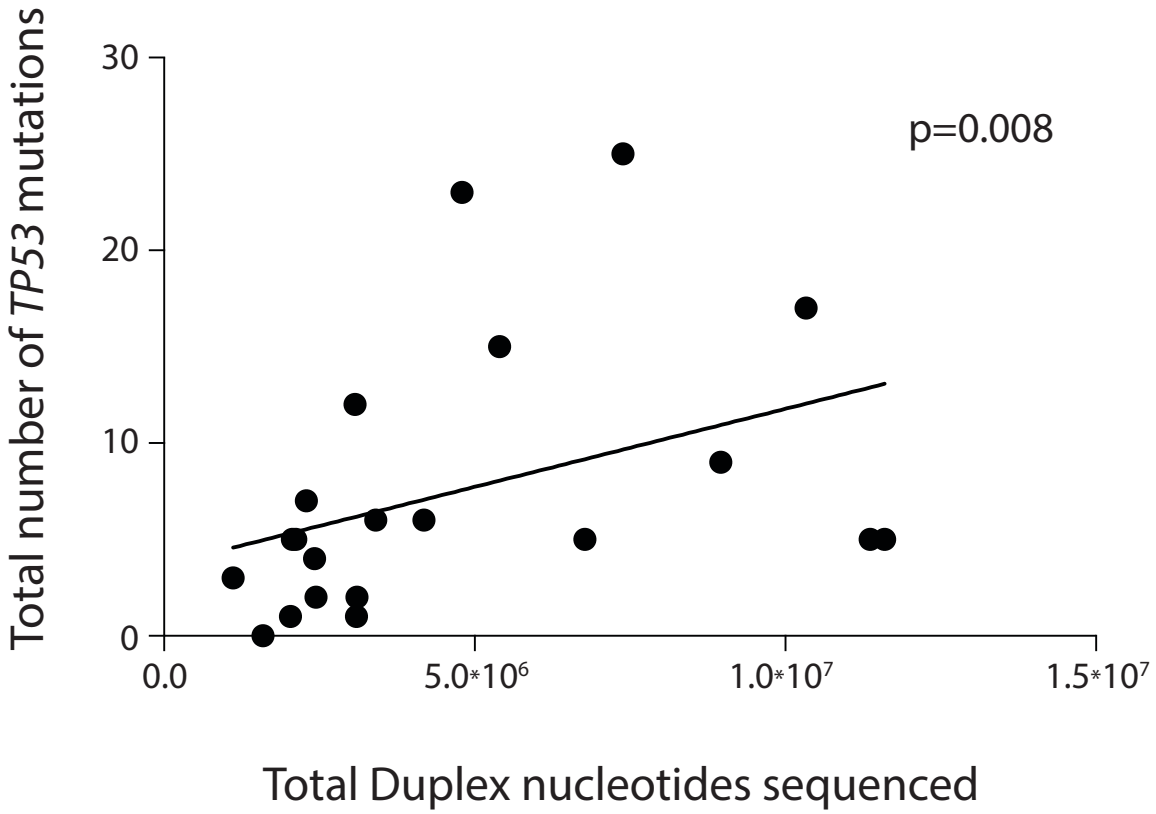


Figure S2. Association between number of independent *TP53* mutations detected and total number of Duplex nucleotides sequenced [related to Fig. 1E-F]. The total number of *TP53* mutations found (including exons and flanking intronic regions) in the 21 uterine lavages of the study was plotted against the total number of Duplex nucleotides sequenced in each sample. More *TP53* mutations were identified in samples with more nucleotides sequenced (p=0.008 by Spearman's correlation test).

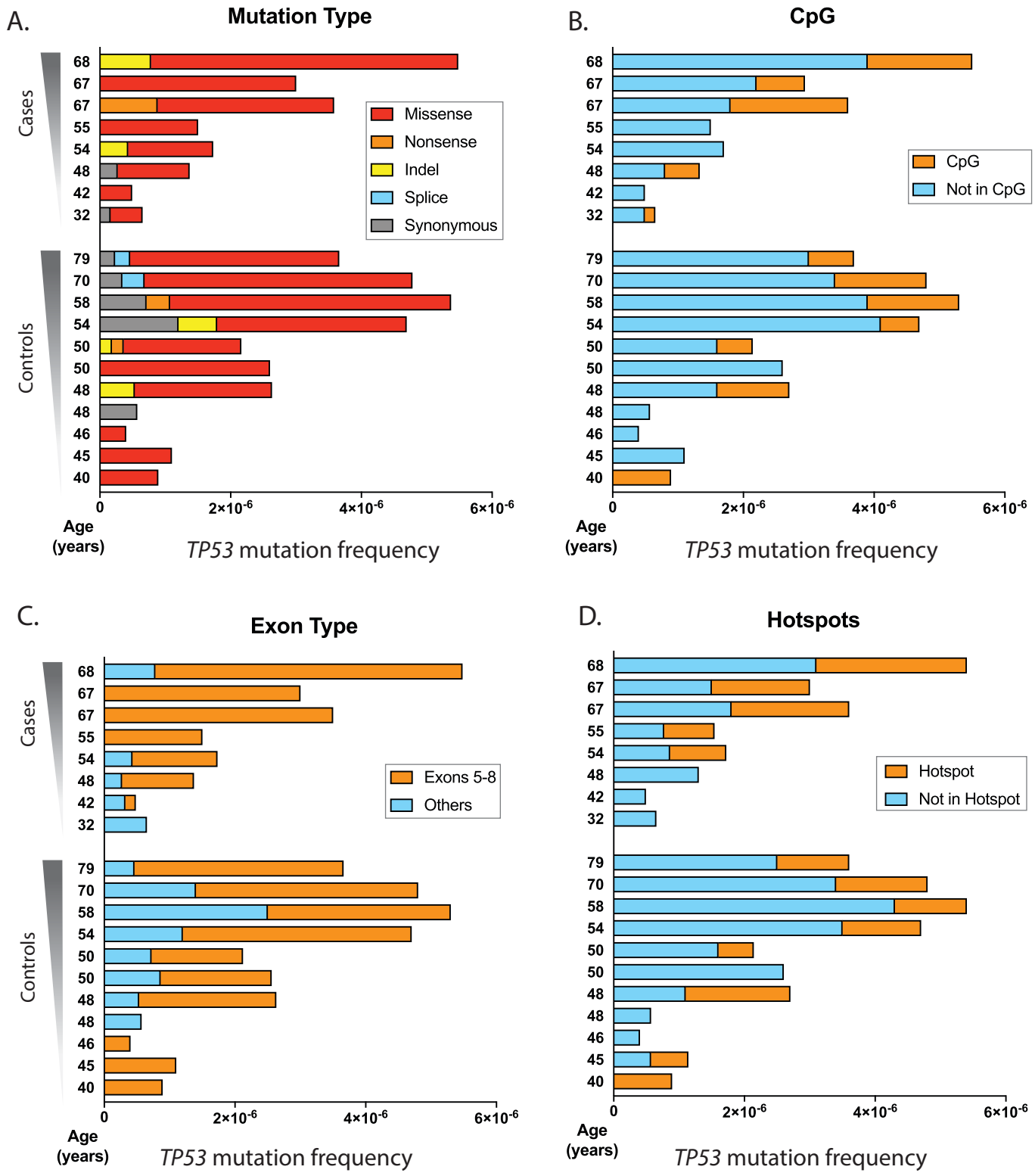


Figure S3. *TP53* mutation frequency and characteristics by age for individual patient lavages in case-control study [related to Fig. 3A-D]. Data is parsed by (A) mutation type, (B) CpG dinucleotide site, (C) exon type and (D) cancer-associated hotspots. Patients are divided in cases (women with ovarian cancer) and controls (women without ovarian cancer) and ordered by age within each group. For each patient, *TP53* mutation frequency was calculated as the number of *TP53* mutations identified in the coding region divided by the total number of Duplex nucleotides sequenced in that region. For each trait, the fraction of mutations corresponding to each of the categories of analysis is indicated by color.

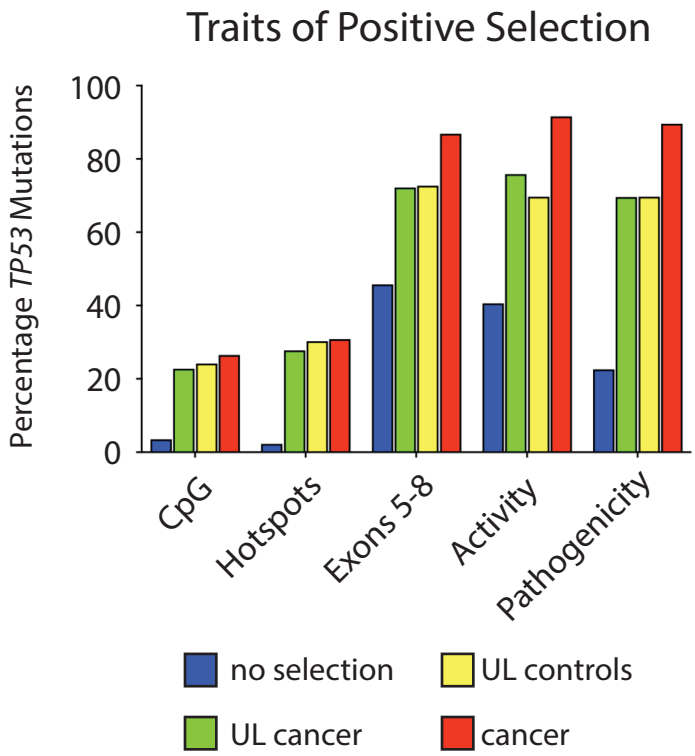


Figure S4. Comparison of traits of positive selection between *TP53* mutations in the UMD cancer database and uterine lavages [related to Fig. 4A]. For each trait, the percentage of observed *TP53* mutations is color-coded for each group. The 'no selection' group includes all possible mutations in the *TP53* coding region (n=3,546). *TP53* background mutations found in uterine lavage from women without ovarian cancer (controls, n= 79) and uterine lavage from women with ovarian cancer (cases, n=33) show similar distribution of mutational traits, which more closely resemble mutations found in cancers (n=71,051) than mutations expected in the absence of selection. UL: Uterine lavage.

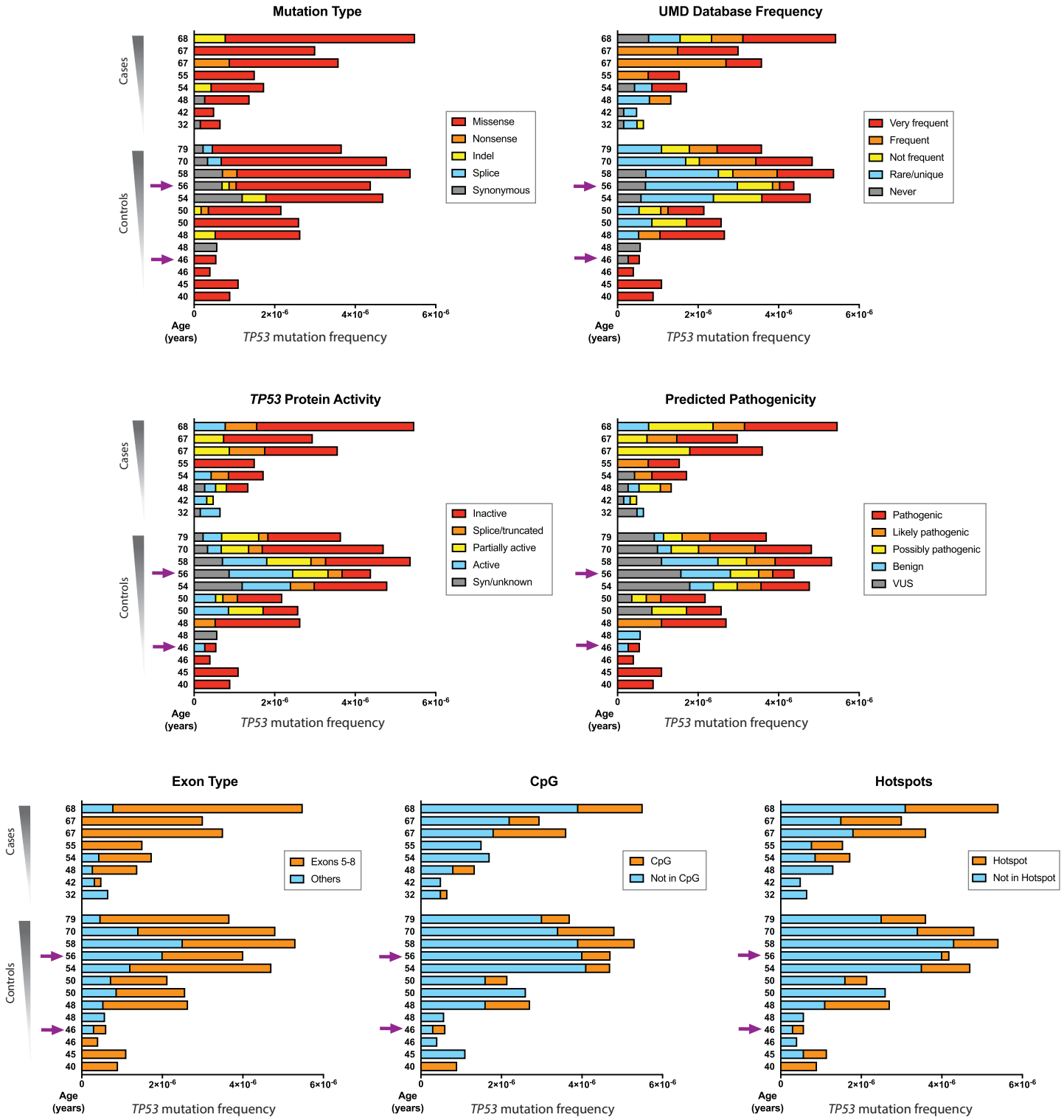


Figure S5. *TP53* mutation frequency and characteristics by age including uterine lavages from the two middle age women in the normal tissue study [related to Figs. 3E-F and 4B]. The two new lavages correspond to a 46 year old woman and a 56 year old woman and are indicated by arrows. *TP53* mutation type, frequency in cancer database, activity, pathogenicity, exon 5-8 location, CpG location, and hotspot location are indicated by color, with warm colors indicating 'cancer-like' features. The *TP53* mutation frequency and the distribution of mutational cancer-like traits in the two new lavages are very similar to the data obtained for women of comparable age in the first part of the study.

TISSUE	MUTATIONS FOUND IN DIFFERENT TISSUES OF SAME INDIVIDUAL					
101 yo	**c.659A>G**	**c.596G>A**	**c.517G>A**	**c.455C>T**	**c.389T>G**	**c.149T>C**
A001-Leukocytes	85/6971	2/11543		44/9587	1/10797	1/11512
A001-Peritoneum (a)	1/4501			1/4939		
A001-Peritoneum (b)	1/5344	3/8248	1/7342	2/6633		
A001-Endometrium (a)	1/5712		1/7803		1/7961	
A001-Endometrium (b)						1/9988

56 yo	**c.151G>T**
A004-Leukocytes	1/13560
A004-Cervix (a)	
A004-Cervix (b)	
A004-Endometrium (a)	
A004-Endometrium (b)	
A004-Myometrium	
A004-FT (a)	
A004-FT (b)	
A004-Ovary (a)	
A004-Ovary (b)	
A004-Uterine lavage	1/4843

46 yo	**c.659A>G**	**c.524G>T**
A006-Leukocytes		1/6665
A006-Peritoneum (a)		
A006-Peritoneum (b)		
A006-Cervix		1/4349
A006-Endometrium		
A006-Myometrium (a)	16/2694	
A006-Myometrium (b)		
A006-FT (a)	1/1918	
A006-FT (b)		
A006-Uterine lavage		
A006-Peritoneal fluid		
A006-ctDNA		

	MAF>1%
	MAF>0.1%

Figure S6. Analysis of mutations shared across multiple tissue samples within the same individual [related to Figs 5B and 6]. For each individual, all analyzed samples are listed and color coded by tissue type. Mutations identified in more than one biopsy are indicated in columns with ratios provided for the biopsies in which the mutation was identified. The ratios indicate the number of Duplex reads with the given mutation divided by the depth of sequencing in that position. Mutations found at MAF>1% and MAF>0.1% are indicated. It should be noted that the first mutation listed for the 101 year old woman and the 46 year old woman is the same and corresponds to codon 220, one of the most common hotspots in *TP53*. FT: fallopian tube.

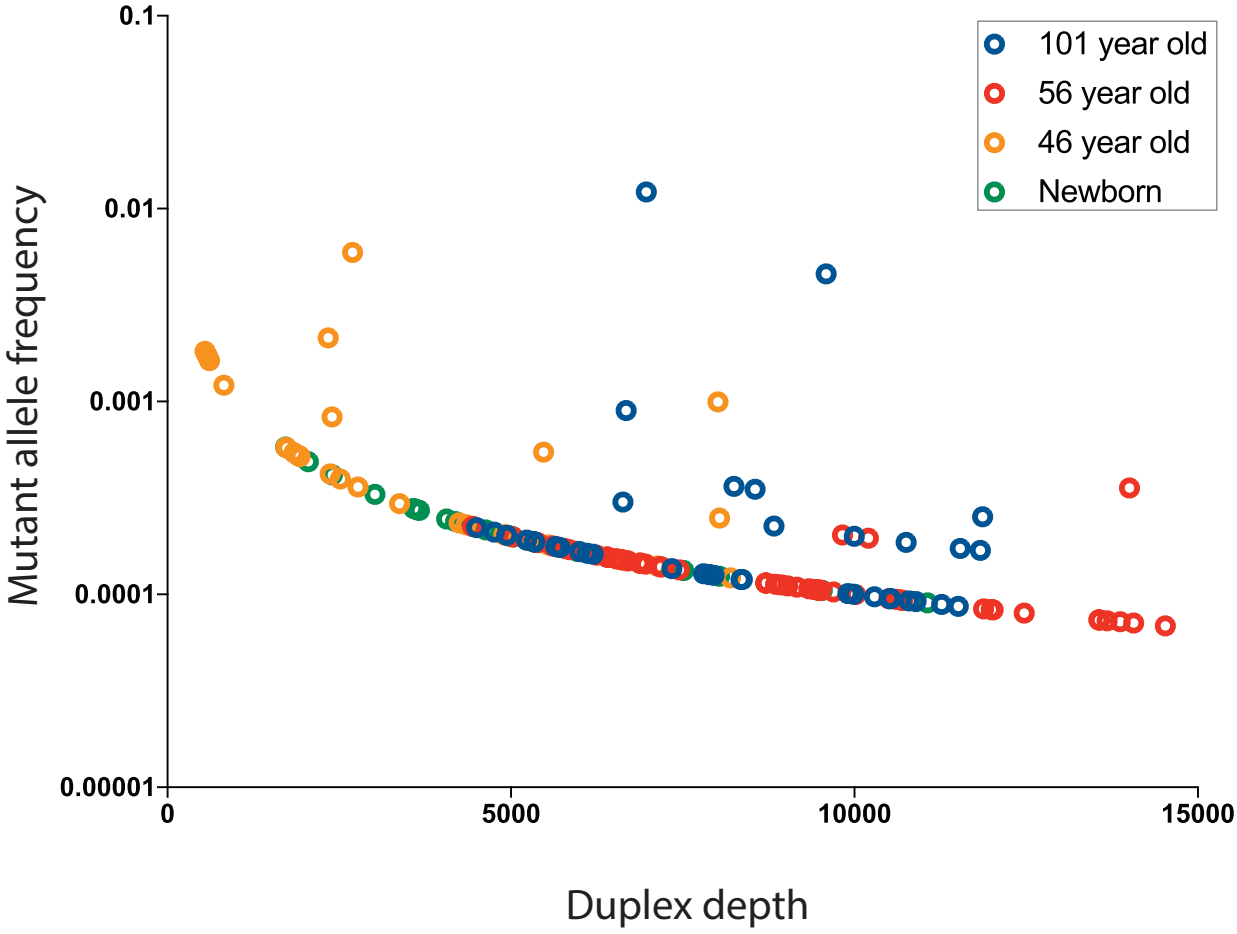
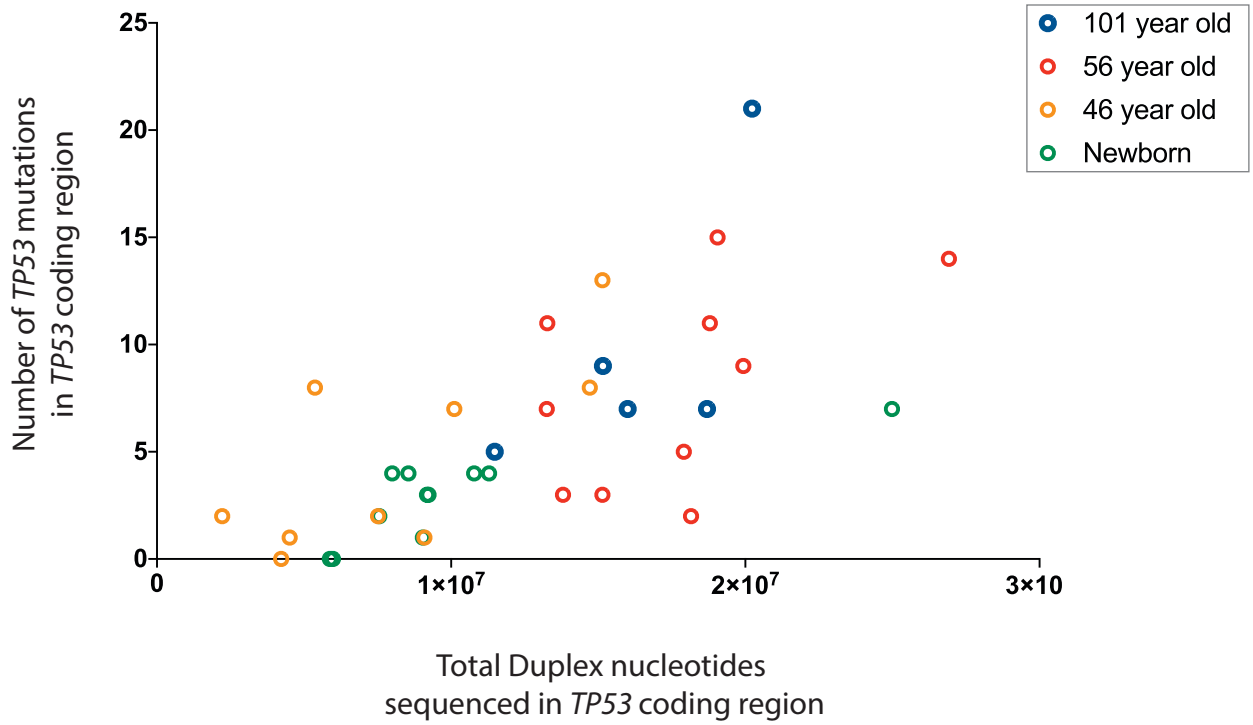


Figure S7. Mutant allele frequency as a function of Duplex Sequencing depth [related to Fig. 5B]. Each dot corresponds to a *TP53* mutation identified in normal tissue. Mutations are color coded by subject. MAF is calculated as the number of times a mutation was observed divided by the depth of sequencing at the given position. Because MAF is inversely associated with depth, mutations identified in biopsies sequenced at a lower depth (mostly from the 46 year old woman and newborn) present with higher MAF.

A.



B.

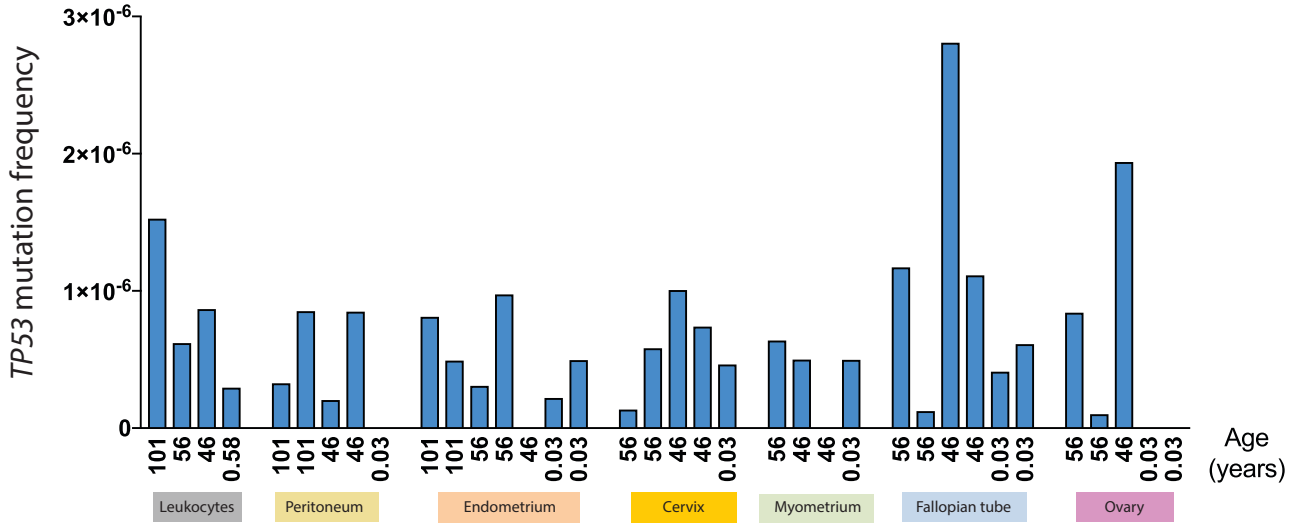


Figure S8. *TP53* mutation frequency by tissue type [related to Figs. 5B, 6, 7A]. (A) Association between number of *TP53* mutations and total number of Duplex nucleotides sequenced in the *TP53* coding region. Dots correspond to samples and are color coded by individual of origin. (B) For each sample, *TP53* mutation frequency was calculated as the number of *TP53* mutations identified in the coding region divided by the total number of Duplex nucleotides sequenced in that region. Subject age is indicated in the X-axis.

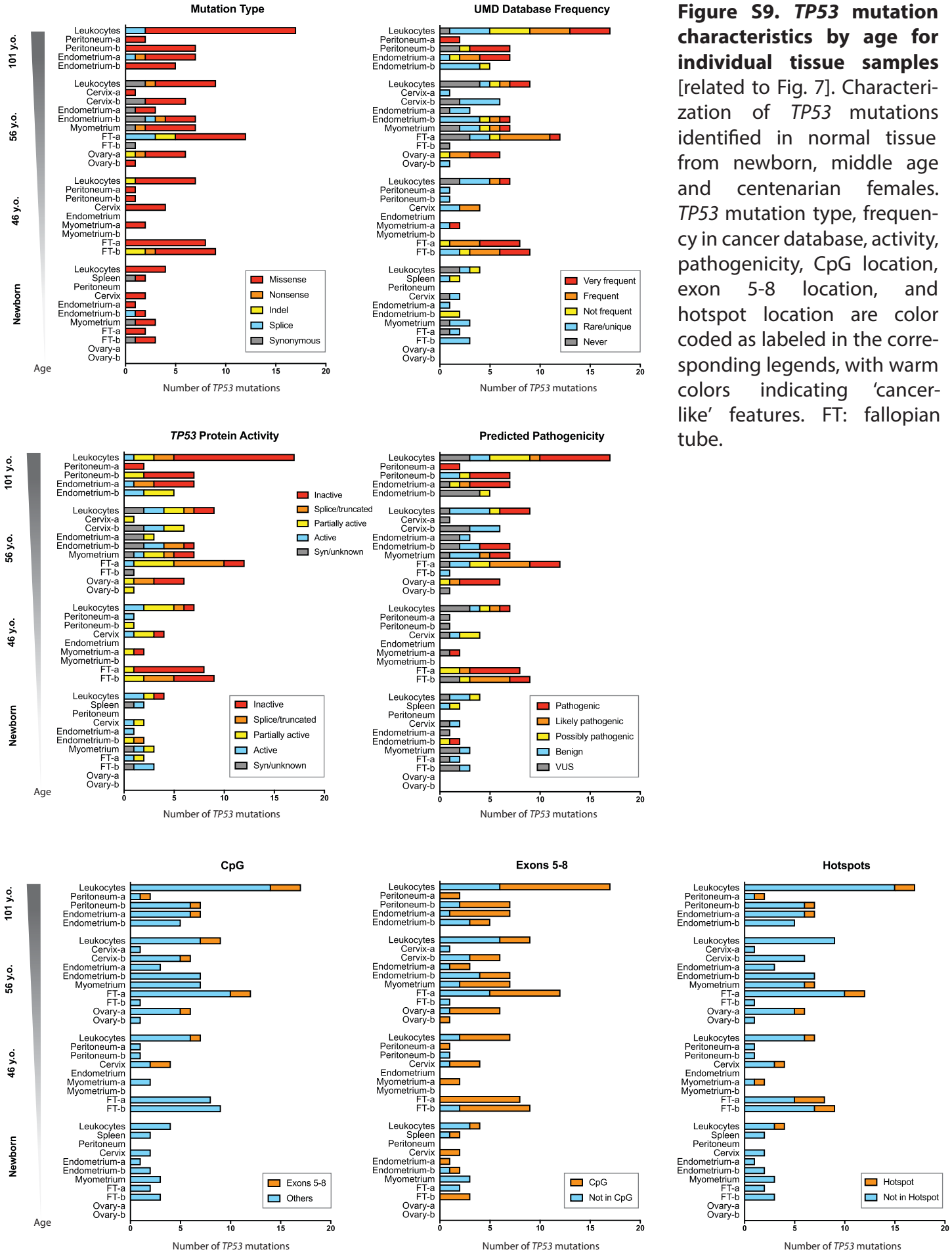


Figure S9. *TP53* mutation characteristics by age for individual tissue samples [related to Fig. 7]. Characterization of *TP53* mutations identified in normal tissue from newborn, middle age and centenarian females. *TP53* mutation type, frequency in cancer database, activity, pathogenicity, CpG location, exon 5-8 location, and hotspot location are color coded as labeled in the corresponding legends, with warm colors indicating 'cancer-like' features. FT: fallopian tube.

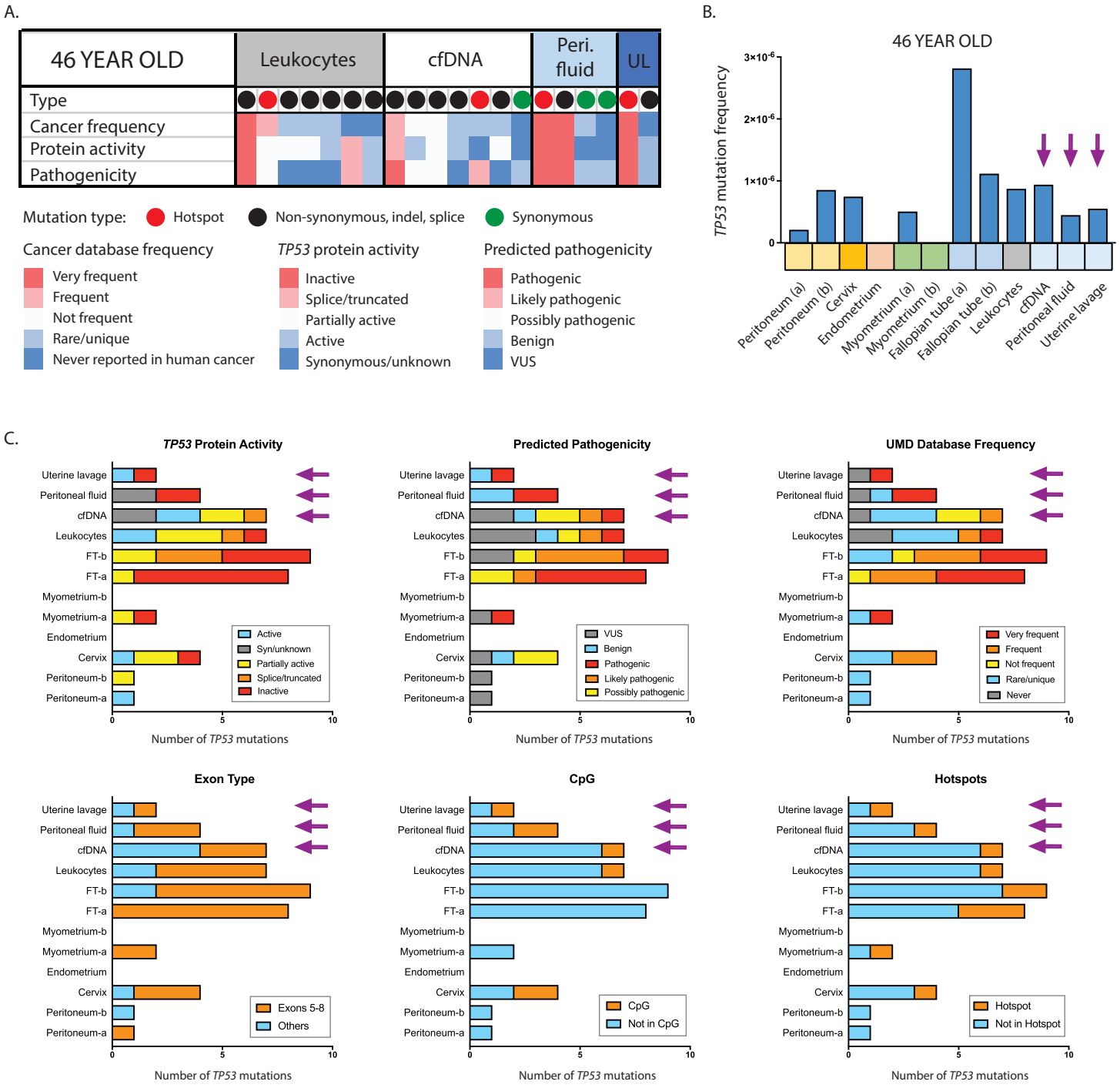


Figure S10. *TP53* mutation characteristics within non-invasively collected body fluids from a 46 year-old woman [related to Figs. 5B and 6]. (A) Heatmap indicating mutation type, frequency in cancer database, impact on protein activity, and predicted pathogenicity for each of the *TP53* mutations identified in leukocytes, cfDNA, peritoneal fluid, and uterine lavage. Categories and color coding are the same as for Figure 6. Each column represents a mutation. (B) Comparison of *TP53* mutation frequency in all tissues collected from the 46 year old woman, including liquid biopsies. (C) Comparison of *TP53* mutational features in all tissues collected from the 46 year old woman, including liquid biopsies (indicated by arrows). FT: fallopian tube.