Table S2. mtDNA sequence variations in single $CD34^+$ cells isolated from group I samples

Sample	Sequence variation	No.	No. of haplotypes	No. of haplotypes defined by nucleotide substitutions
Sample 5	Aggregate sequence	36	1	
	+16086T>Y, 9C/10C	1	2	1
	+16131T>Y, 8C/9C/10C/11C	1	3	2
	+16150C>Y, 9C/10C/11C	1	4	3
	+16181A>A/C, 9C/10C/11C	1	5	4
	+16519T>Y, 9C/10C/11C	1	6	5
	8C/9C/10C/11C	1	7	
	8C/9C/10C/11C/12C	1	8	
	9C/10C	19	9	
	9C/10C/11C/12C	34	10	
Sample 10	Aggregate sequence	34	1	
	+16034G>R, 9C/10C/11C/12C	1	2	1
	+16072C>Y, 9C/10C/11C	1	3	2
	+16081A>R, 16445T>Y, 9C/10C/11C/12C	1	4	3
	+16086T>Y, 9C/10C/11C	1	5	4
	+16124T>C, 9C/10C/11C	1	6	7
	+16129G>R, 9C/10C/11C/12C	1	7	6
	+16131T>C, 9C/10C	2	8	7
	+16131T>C, 9C/10C/11C	1	9	
	+16468T>Y, 9C/10C/11C/12C	1	10	8
	+204T>Y, 9C/10C/11C/12C	1	11	9
	+226T>Y, 9C/10C	1	12	10
	+302A>R, 9C/10C/11C/12C	1	13	11
	+303C>C/A, 9C/10C/11C/12C	1	14	12
	+303insA, 8C/9C/10C/11C	1	15	
	+405T>C, 466T>C, 9C/10C/11C/12C	1	16	13
	+405T>C, 9C/10C/11C/12C	1	17	14
	8C/9C/10C	1	18	
	8C/9C/10C/11C	3	19	
	8C/9C/10C/11C/12C	1	20	
	9C/10C	8	21	
	9C/10C/11C	31	22	
Sample 6	Aggregate sequence	40	1	
	+16018T>Y, 9C/10C/11C/12C	1	2	1
	+16250C>Y, 9C/10C	1	3	2

	+182C>Y, 9C/10C/11C	1	4	3
	+204T>Y, 9C/10C/11C	1	5	4
	+270A, 9C/10C	1	6	5
	+405T>C, 9C/10C/11C	1	7	6
	+418C>Y, 9C/10C/11C/12C	1	8	7
	+445C>Y, 9C/10C/11C/12C	1	9	8
	+513G>A, 9C/10C/11C/12C	1	10	9
	+573insC/non-ins, 9C/10C/11C/12C	1	11	
	+574A>R, 9C/10C/11C/12C	1	12	10
	+(523-524)delAC/non-del, 9C/10C/11C	1	13	
	+(523-524)delAC/non-del, 9C/10C/11C/12C	1	14	
	8C/9C/10C	1	15	
	8C/9C/10C/11C	1	16	
	8C/9C/10C/11C/12C	1	17	
	9C/10C	18	18	
	9C/10C/11C	21	19	
Sample 9	Aggregate sequence	44	1	
	+16095C>Y, 9C/10C/11C/12C	1	2	1
	+16129G>A, 9C/10C/11C/12C	1	3	2
	+16287C>Y, 308C>Y, 9C/10C/11C	1	4	3
	+16301C>Y, 9C/10C	1	5	4
	+217T>C, 9C/10C/11C	1	6	5
	+260G>R, 9C/10C/11C/12C	1	7	6
	+308C>Y, 9C/10C/11C	1	8	7
	+405T>C, 9C/10C/11C	1	9	8
	8C/9C/10C	1	10	
	8C/9C/10C/11C	5	11	
	9C/10C	8	12	
	9C/10C/11C	29	13	

Note: Sequence variation was scored relative to the Cambridge reference sequence [1]. A site heterogeneous for both thymidine (T) and cytidine (C) was abbreviated as Y, and heterogeneous for adenosine (A) and guanosine (G) was abbreviated as R. A site heterogeneous for a transversion or an indel (insertion and deletion) was highlighted by listing all status. We followed the approach in our previous studies [14, 16] to highlight the status of the C-tract length variations in region 303-309 and the dinucleotide AC repeat in region 515-524 in these cells differ from the aggregated sequence by listing all the status, e.g. 9C/10C means region 303-315 has heteroplasmy of CCCCCCCCCTCCCCCC and CCCCCCCCCCTCCCCCC in the cell, 523-524delAC/non-del means co-existing of 4 and 5 repeats of AC in region 515-524 in the cell. +, extra mtDNA changes compared with the aggregate sequence. The aggregate sequence of Sample 5 contains 9C/10C/11C, whereas the other three samples all contain 9C/10C/11C/12C. The sequence variations in single $CD34^+$ cells isolated from group II samples (B-8, B-3, B-6 and B-5) were reported in Yao et al. [15].