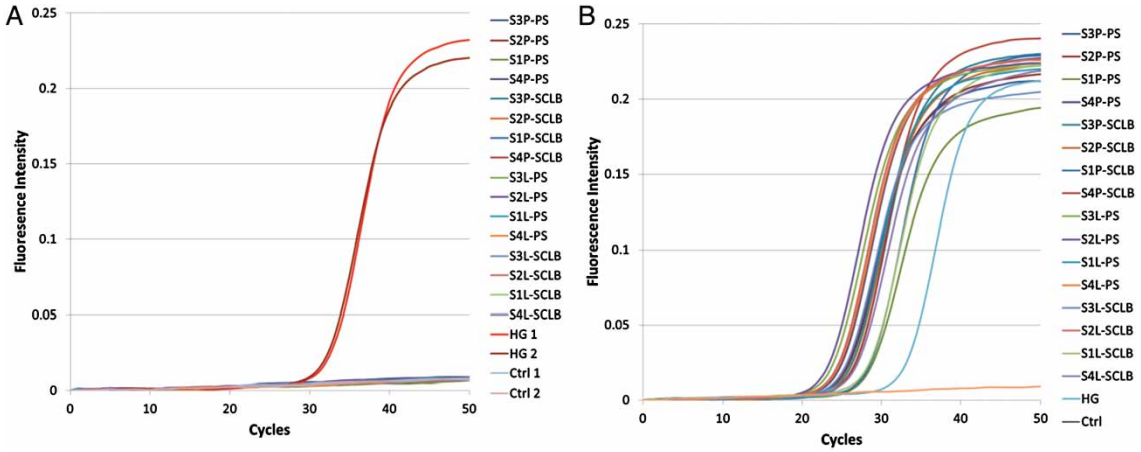

Figure S1. PCR assessment of DNA contamination and RNA integrity. A. Isolated RNAs and two human genomic cDNA samples were used as template. B. The synthesized cDNAs were used as template. The primer pair spans the *PRM1* intron. S1~S4 are the samples from four subjects; 'P' denotes pelleted storage. 'L' denotes liquefied storage; PS denotes PureSperm purification and SCLB is somatic cell lysis buffer purification; HG denotes human genome DNA; Ctrl denotes negative control.

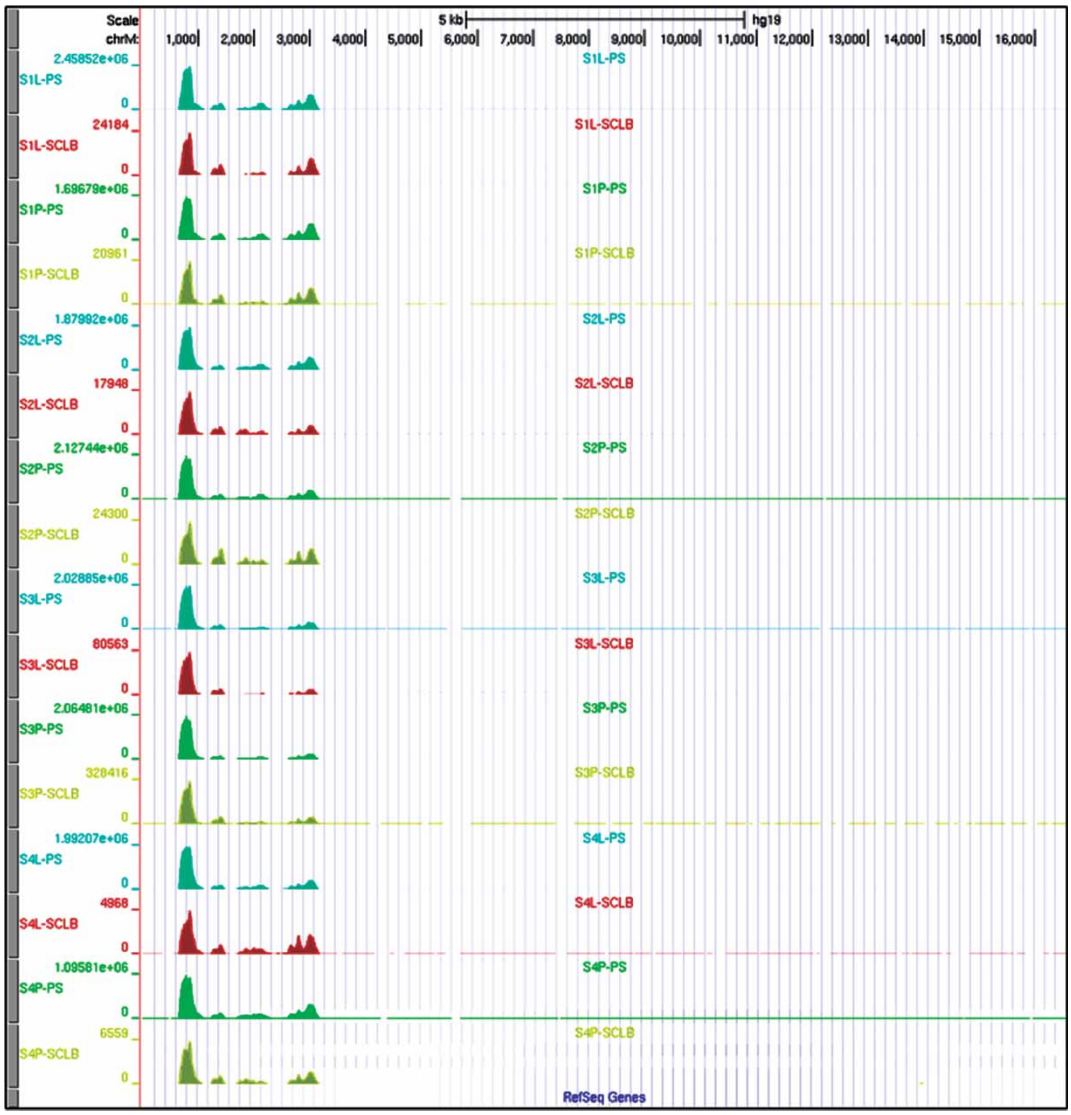

Figure S2. Distribution of reads mapped to mitochondrial genome. The scale for each sample is based on the maximum read count with the displayed region. In the mitochondrial genome, the coordinate of 12S rRNA is 648 – 1,601 and 16S rRNA is 1,671 – 3,229. S1~S4 are the samples from four subjects; 'P' denotes pelleted storage. 'L' denotes liquefied storage; PS denotes PureSperm purification and SCLB is somatic cell lysis buffer purification.

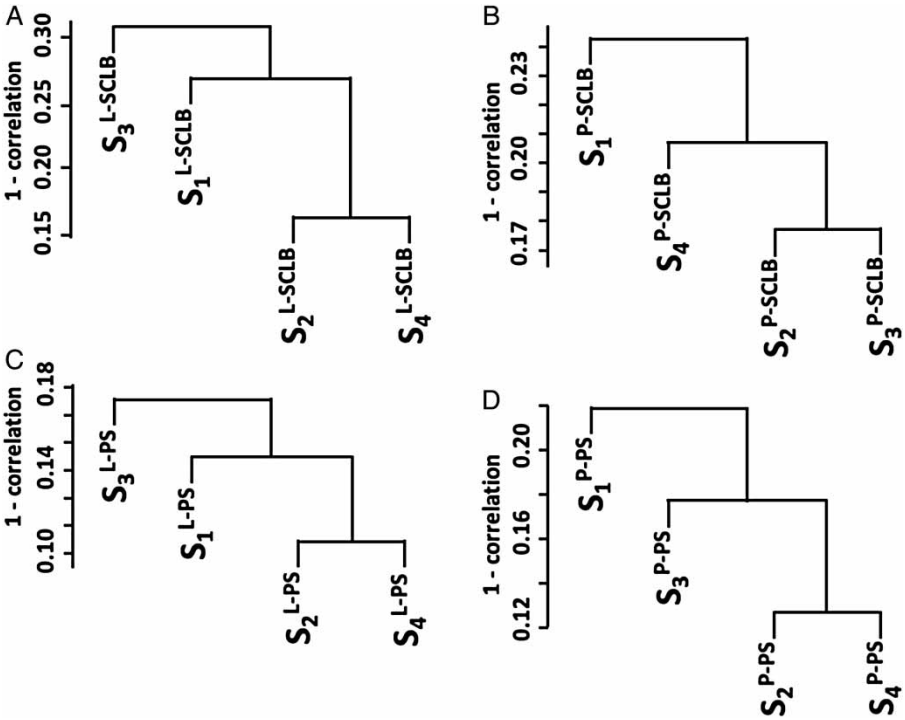

Figure S3. Unsupervised clustering by subjects. In each method, the transcript profile generated from four subjects were clustered. The Y-axis is the 1 - Pearson correlation coefficient. Panel A indicates SCLB purification from liquefied storage (L) samples; Panel B indicates SCLB purification from pelleted storage (P) samples; Panel C indicates PureSperm purification from liquefied storage (L) samples; and Panel D indicate PureSperm purification from pelleted storage (P) samples.

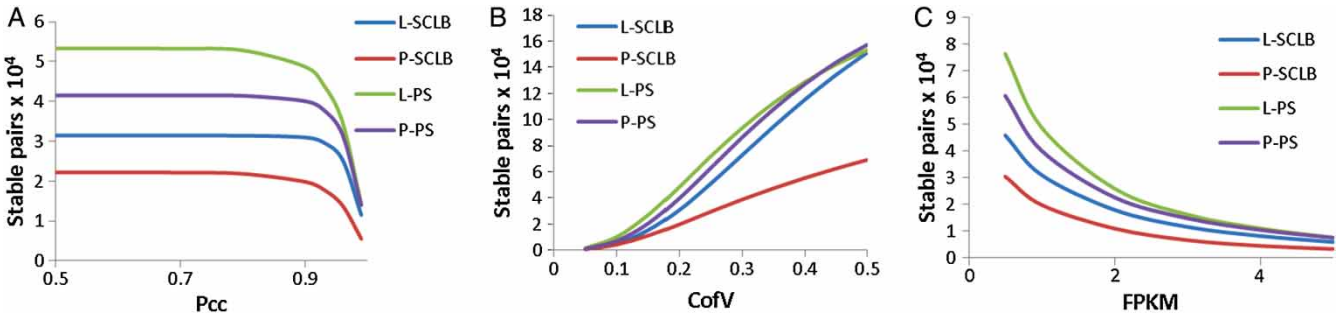

Figure S4. Number of stable transcript pairs as a function of different criteria. The short reads uniquely aligned to reference genome without mitochondrial sequence were used as input to calculate each gene's FPKM values. The Y-axis is the number of stable transcript pairs. The X-axis is: Pcc in panel A, Cofv in panel B; and FPKM in panel C. The other criteria are also required for stable transcripts identification: In panel A, CofV≤0.2, FPKM≥1; in panel B, Pcc≥0.9, FPKM≥1; in panel C, Pcc≥0.9, CofV≤0.2. The Spearman correlation coefficient=1 for the stable pairs in all three panels.

Table S1. The number of sperm from each sample (in million cells)

Subjects	Pellet storage (P)	Liquefied storage (L)
S_1	333	380
S_2	290	160
S_3	265	510
S_4	310	140

Table S2. RNA-seq statistics summary

Samples	Total reads	Aligned reads	% of align	Unique aligned reads	Average length (bp)	STDEV (bp)
$S_1^{P\text{-}SCLB}$	16.9	14.2	83.8%	10.0	85	35.2
$S_2^{P\text{-}SCLB}$	14.5	13.2	90.9%	11.2	93	40.3
$S_3^{P\text{-}SCLB}$	14.7	13.4	90.8%	11.9	86	33.6
$S_4^{P\text{-}SCLB}$	17.4	14.8	85.4%	12.0	90	39.4
$S_1^{L\text{-}SCLB}$	17.9	14.4	80.5%	11.5	92	37.6
$S_2^{L\text{-}SCLB}$	21.2	19.3	91.1%	15.5	93	38.8
$S_3^{L\text{-}SCLB}$	42.0	37.2	88.5%	31.3	93	39.1
$S_4^{L\text{-}SCLB}$	19.1	16.9	88.7%	14.0	96	43.4
$S_1^{P\text{-}PS}$	29.2	27.1	92.8%	22.6	110	37.9
$S_2^{P\text{-}PS}$	32.4	29.6	91.3%	26.1	108	41.1
$S_3^{P\text{-}PS}$	26.7	25.0	93.7%	22.3	111	39.7
$S_4^{P\text{-}PS}$	24.9	22.9	91.7%	19.8	106	40.5
$S_1^{L\text{-}PS}$	35.9	34.8	97.0%	28.8	103	38.2
$S_2^{L\text{-}PS}$	27.8	26.4	95.1%	23.3	106	39.9
$S_3^{L\text{-}PS}$	25.8	23.7	92.1%	20.8	105	39.9
$S_4^{L\text{-}PS}$	27.0	26.0	96.3%	22.5	105	38.7

The read counts in the table are in millions of reads. "Total reads" indicates the number of reads generated from sequencer; "Aligned reads" indicate the number of reads that have been aligned to reference genome by aligner; "Unique aligned reads" indicates the number of reads that were uniquely aligned to reference genome. "Average length" and "STDEV" are the average length of mapped fragments and their standard deviation. S1∼S4 are four different subjects. P-PS denotes pelleted storage with PureSperm purification; L-PS denotes liquefied storage with PureSperm purification; P-SCLB denotes pelleted storage and SCLB purification; L-SCLB denotes liquefied storage and SCLB purification.