

S4 Table. List of sample filtering criteria and number of samples removed

	Number of Samples	
Total sequenced by ddGBS + WGS	4,625 (17 WGS only)	
Low read count in FASTQ (< 4 million reads) **performed prior to variant calling	316	
Females samples (only females in sample set)	77	
Purchased from Taconic	4	
Lack of phenotype information	10	
Poor clustering in principal component analysis	54	
Putative sample mix-up	18	
Duplicate sample	7	
Inflated/deflated heterozygosity	34	
≥ 30 samples with $\hat{\pi} \geq 0.1875$ in sample	12	
$\hat{\pi} > 0.6$ with another sample	32	
Final Sample Set (Harlan + Charles River)	4,061	
	CR – 1,780	Har – 2,281