Open Access

Asian Australas. J. Anim. Sci.
Vol. 27, No. 9 : 1345-1354 September 2014
http://dx.doi.org/10.5713/ajas.2013.13857

www.ajas.info
pISSN 1011-2367 eISSN 1976-5517

Copy Number Deletion Has Little Impact on Gene Expression Levels in Racehorses

Kyung-Do Park[1,a], Hyeongmin Kim[2,a], Jae Yeon Hwang[2], Chang-Kyu Lee[2], Kyoung-Tag Do[1], Heui-Soo Kim[3], Young-Mok Yang[4], Young-jun Kwon[5], Jaemin Kim[5], Hyeon Jeong Kim[6], Ki-Duk Song[1], Jae-Don Oh[1], Heebal Kim[2,6], Byung-Wook Cho[7], Seoae Cho[6], and Hak-Kyo Lee[1,*]

[1] Genomic Informatics Center, Hankyong National University, Anseong 456-749, Korea

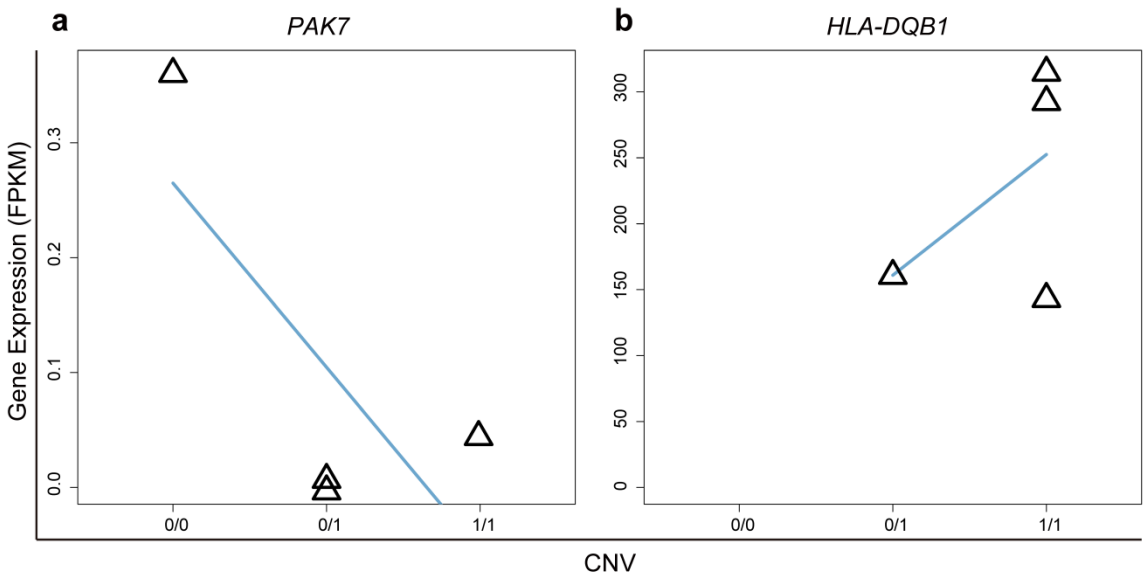
SUPPLEMENTARY DATA

TABLE OF CONTENTS

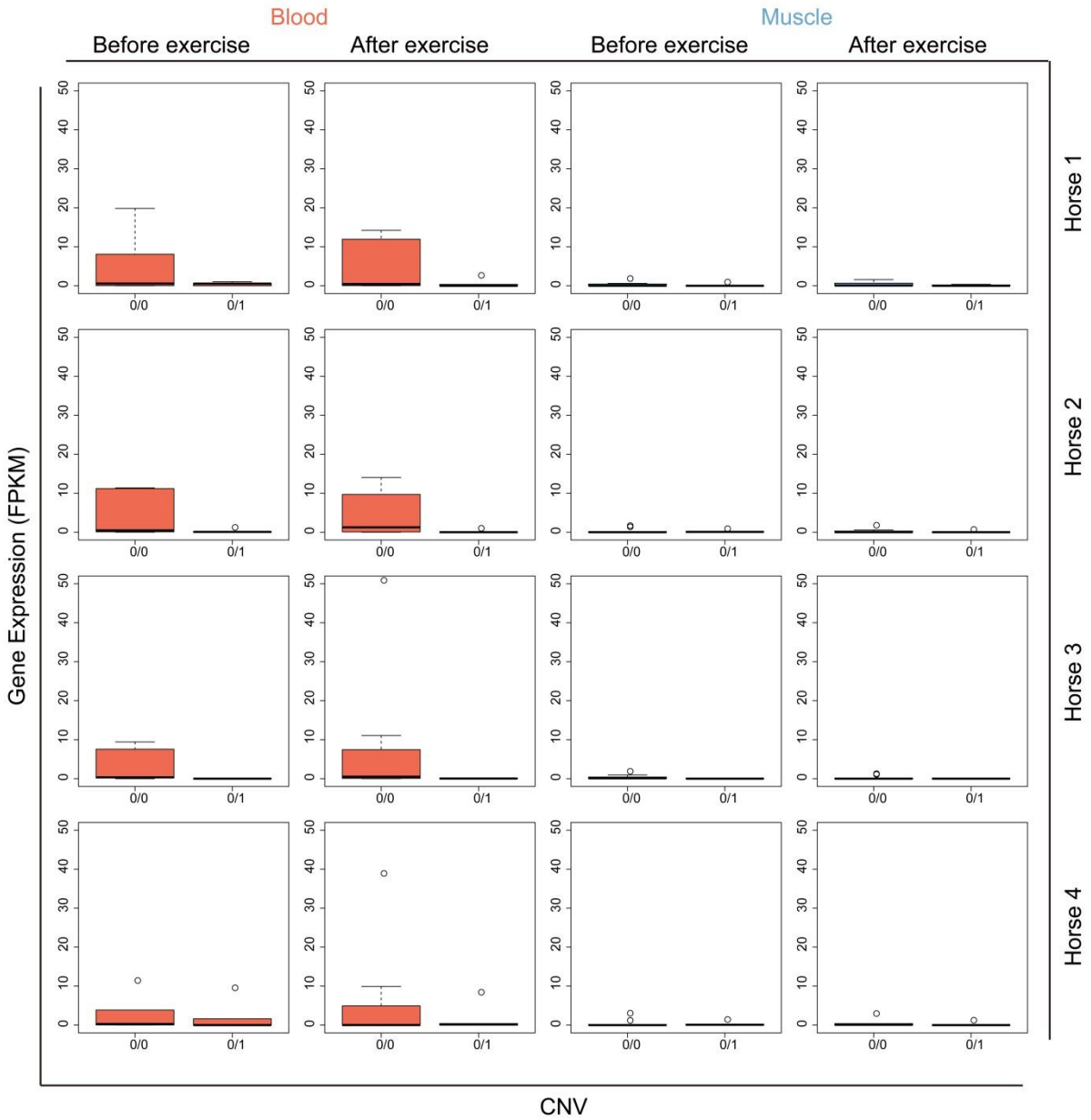
Supplementary Figures

Supplementary Tables

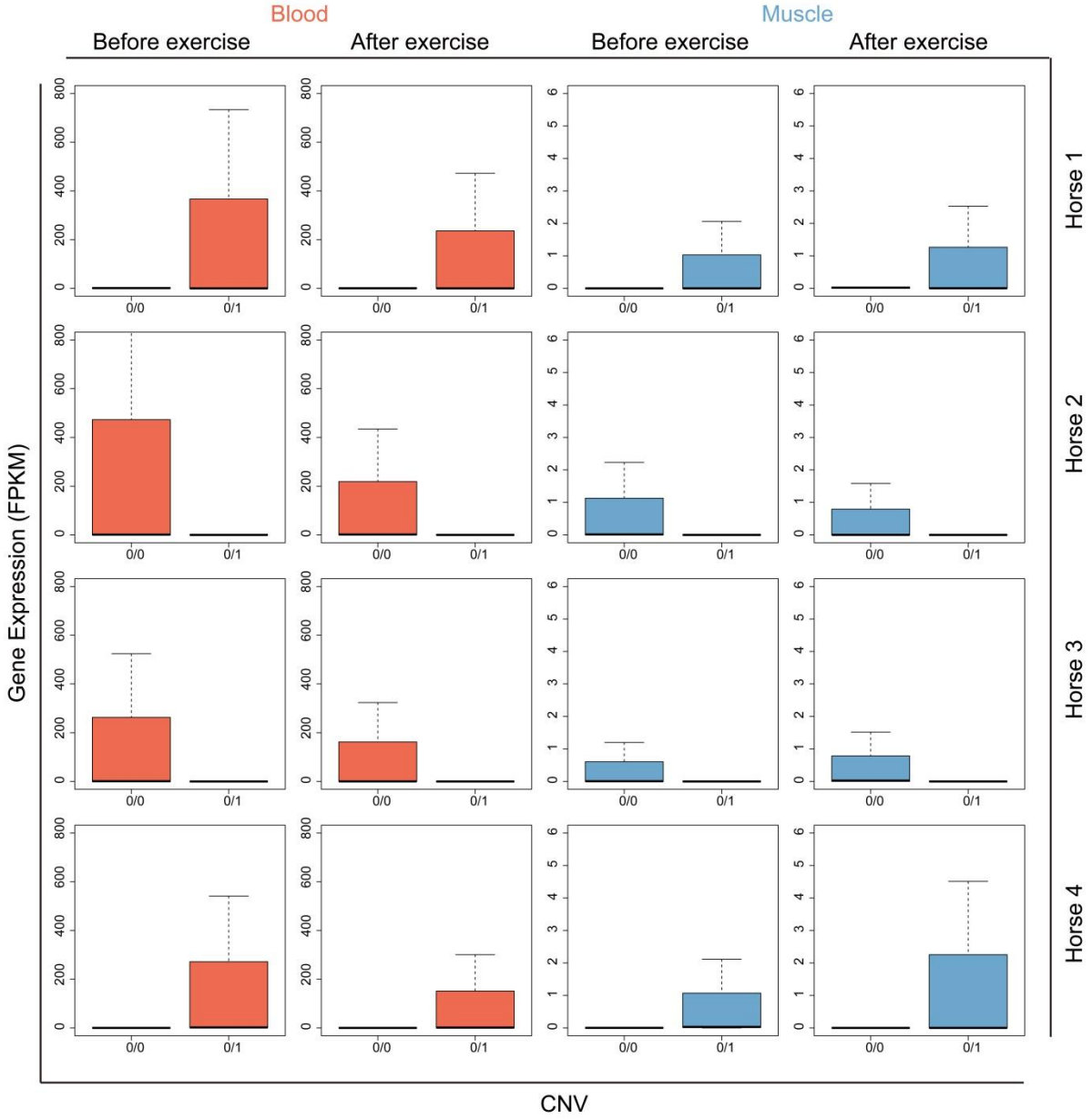
Supplementary Figures



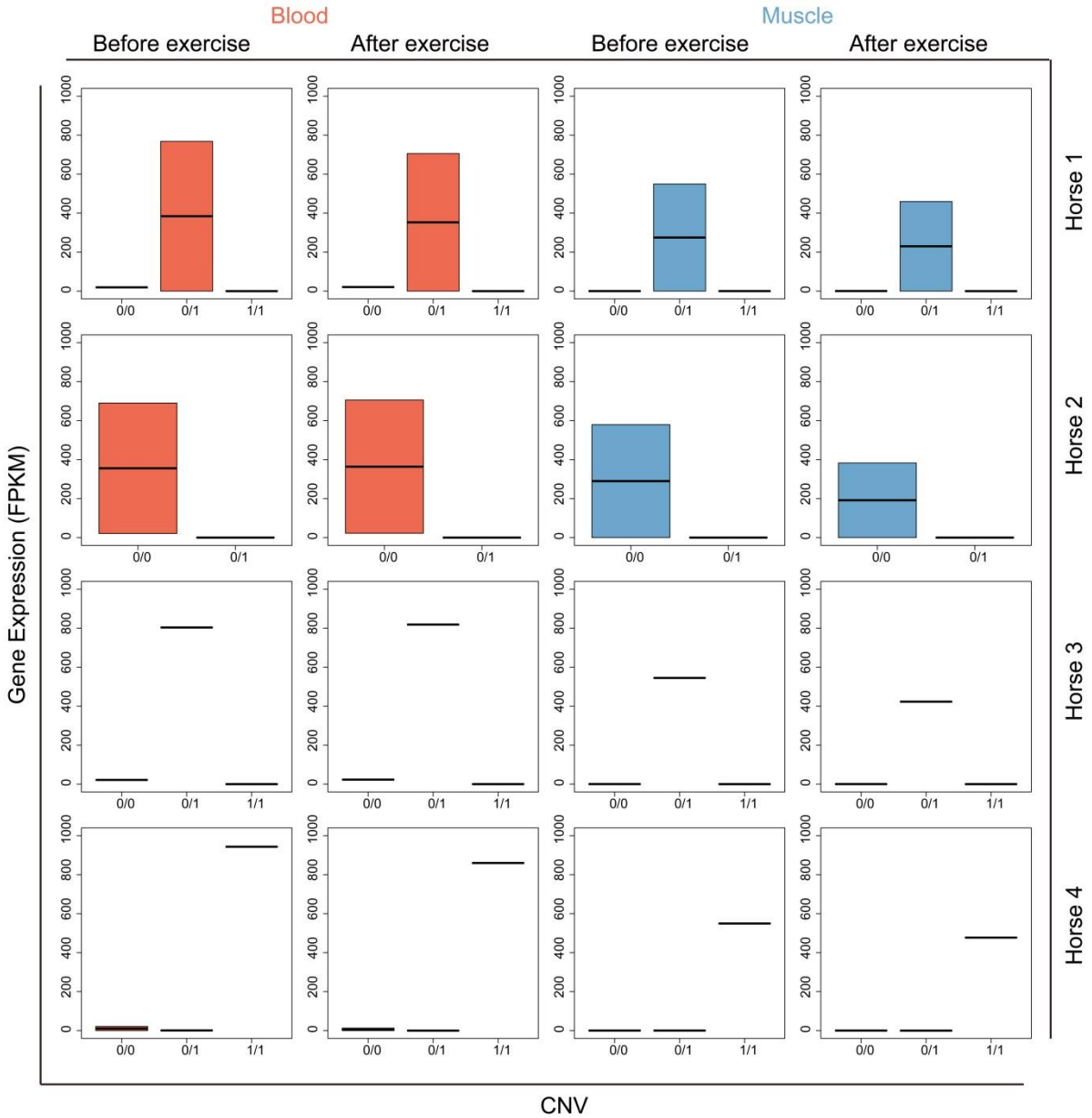
Supplementary Figure 1. Deletion status of CNVs and (a) *PAK7 (*b) *HLA-DQB1* expressions. Deletion status of CNVs and *PAK7*, *HLA-DQB1* expressions in the blood of four Thoroughbred horses before exercise are shown. In each plot, the *x*-axis shows the deletion status of CNVs and the *y*-axis shows the gene expression level (FPKM). The deletion status of CNVs is denoted by 0/0 for no deletion, 0/1 for one allele deletion, and 1/1 for two allele deletions at the CNV region. The linear regression line (blue solid line) for each gene is shown.



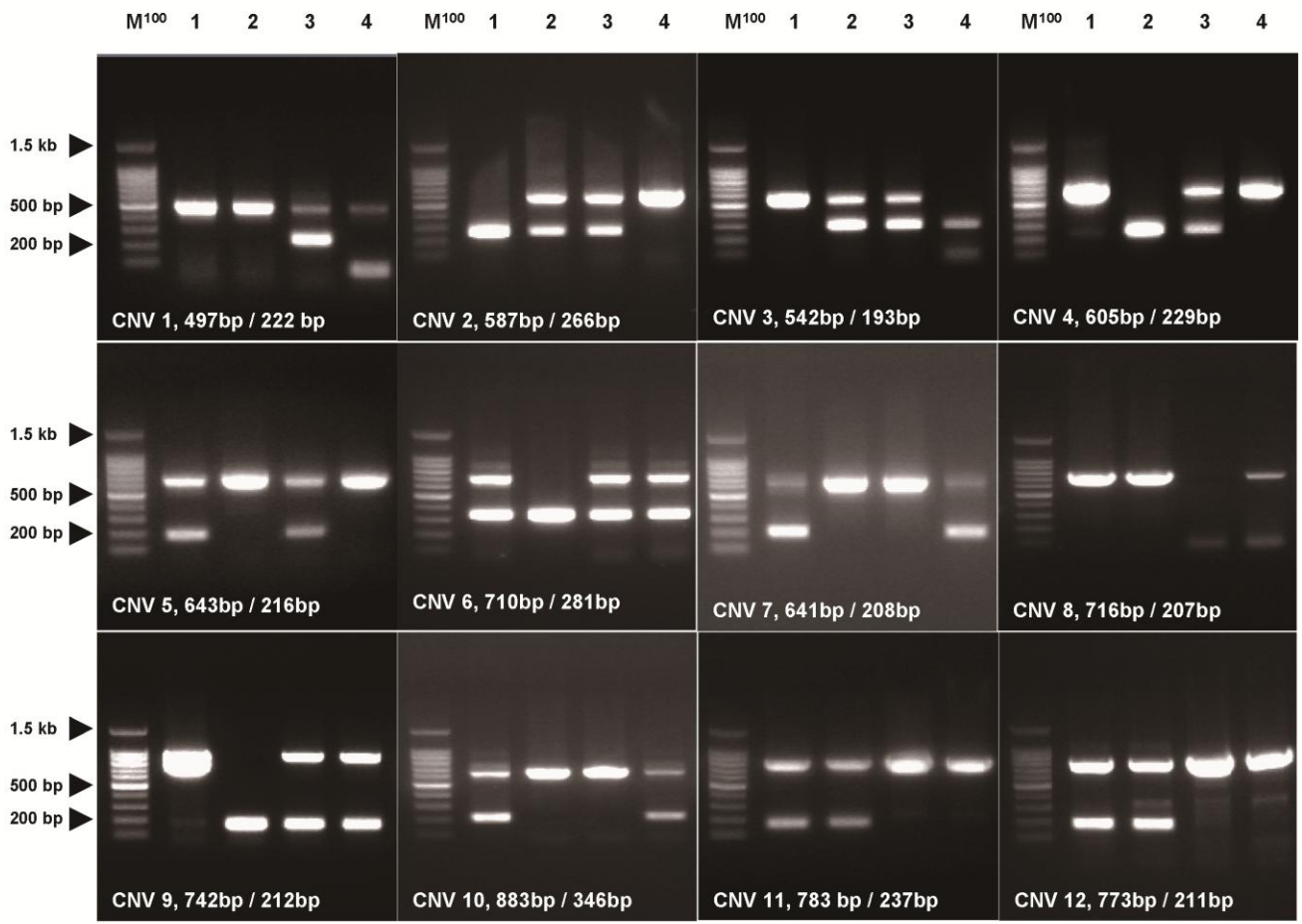
Supplementary Figure 2. The skeletal Box and Whisker plots show the relationships between CNVs and 'inside' position linked gene expression in blood and muscle before and after exercise of four Thoroughbred horses. The 'inside' means that the genes are located in CNV regions. In each box plot, x-axis shows the deletion status of CNVs and y-axis shows the gene expression level (FPKM). The deletion status of CNVs are denoted by 0/0 for no deletion at the population or individual based CNV region, 0/1 for one allele deletion at the CNV region and 1/1 for two allele deletion at the CNV region. The sampling conditions in four horses are shown on the top of the figure, which are blood (red rectangles in the box plot) and muscle (blue rectangles in the box plot) before and after exercise. On the right side of the panel individual horse number was shown.



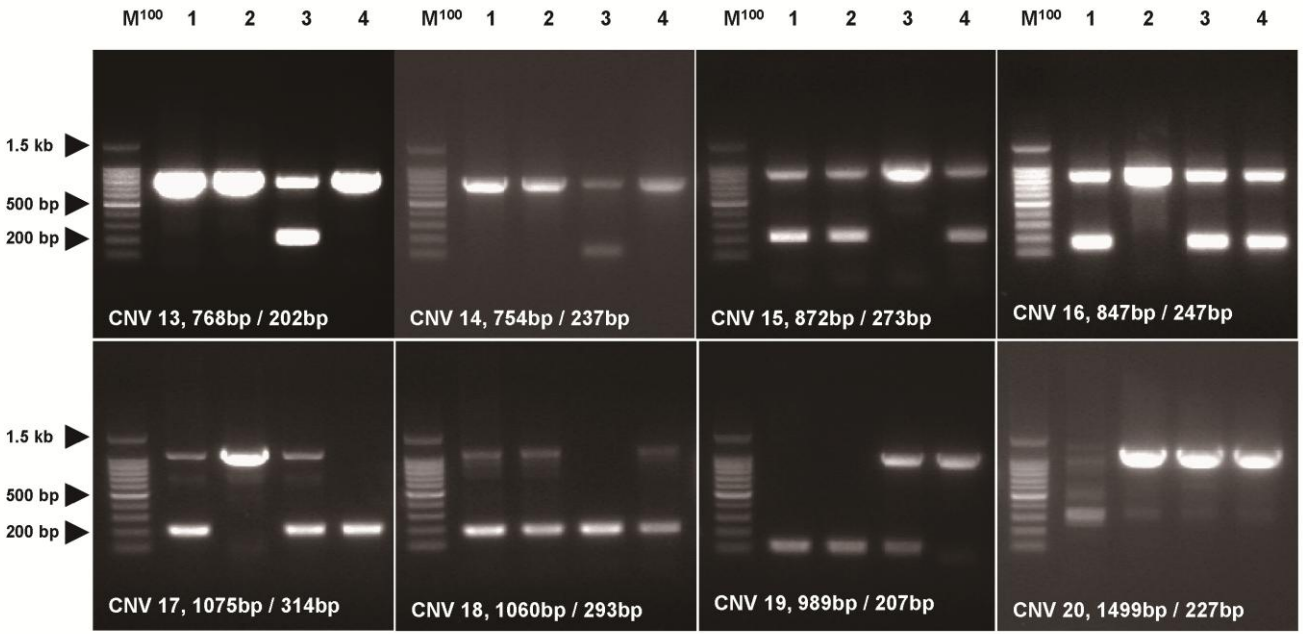
Supplementary Figure 3. The skeletal Box and Whisker plots show the relationships between CNVs and 'front' position linked gene expression in blood and muscle before and after exercise of four Thoroughbred horses. The 'front' means that the genes span the front part of the CNV regions. In each box plot, x-axis shows the deletion status of CNVs and y-axis shows the gene expression level (FPKM). The deletion status of CNVs are denoted by 0/0 for no deletion at the population or individual based CNV region, 0/1 for one allele deletion at the CNV region and 1/1 for two allele deletion at the CNV region. The sampling conditions in four horses are shown on the top of the figure, which are blood (red rectangles in the box plot) and muscle (blue rectangles in the box plot) before and after exercise. On the right side of the panel individual horse number was shown.



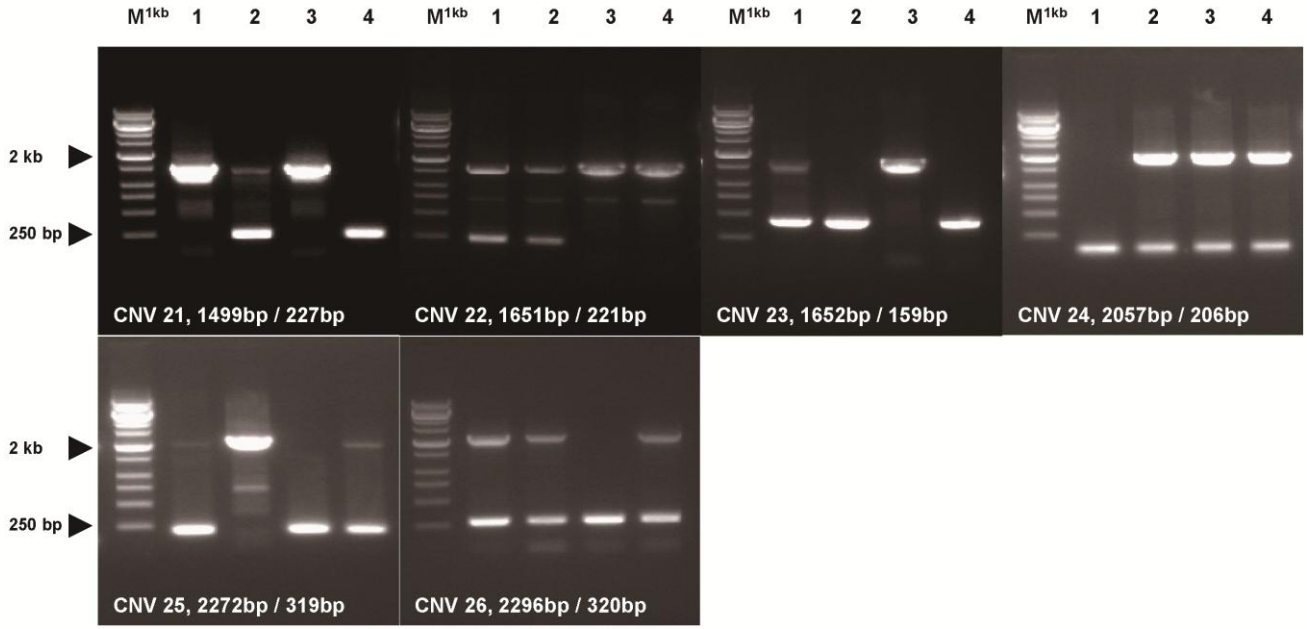
Supplementary Figure 4. The skeletal Box and Whisker plots show the relationships between CNVs and 'rear' position linked gene expression in blood and muscle before and after exercise of four Thoroughbred horses. The 'rear' means that the genes span the rear part of the CNV regions. In each box plot, x-axis shows the deletion status of CNVs and y-axis shows the gene expression level (FPKM). The deletion status of CNVs are denoted by 0/0 for no deletion at the population or individual based CNV region, 0/1 for one allele deletion at the CNV region and 1/1 for two allele deletion at the CNV region. The sampling conditions in four horses are shown on the top of the figure, which are blood (red rectangles in the box plot) and muscle (blue rectangles in the box plot) before and after exercise. On the right side of the panel individual horse number was shown.



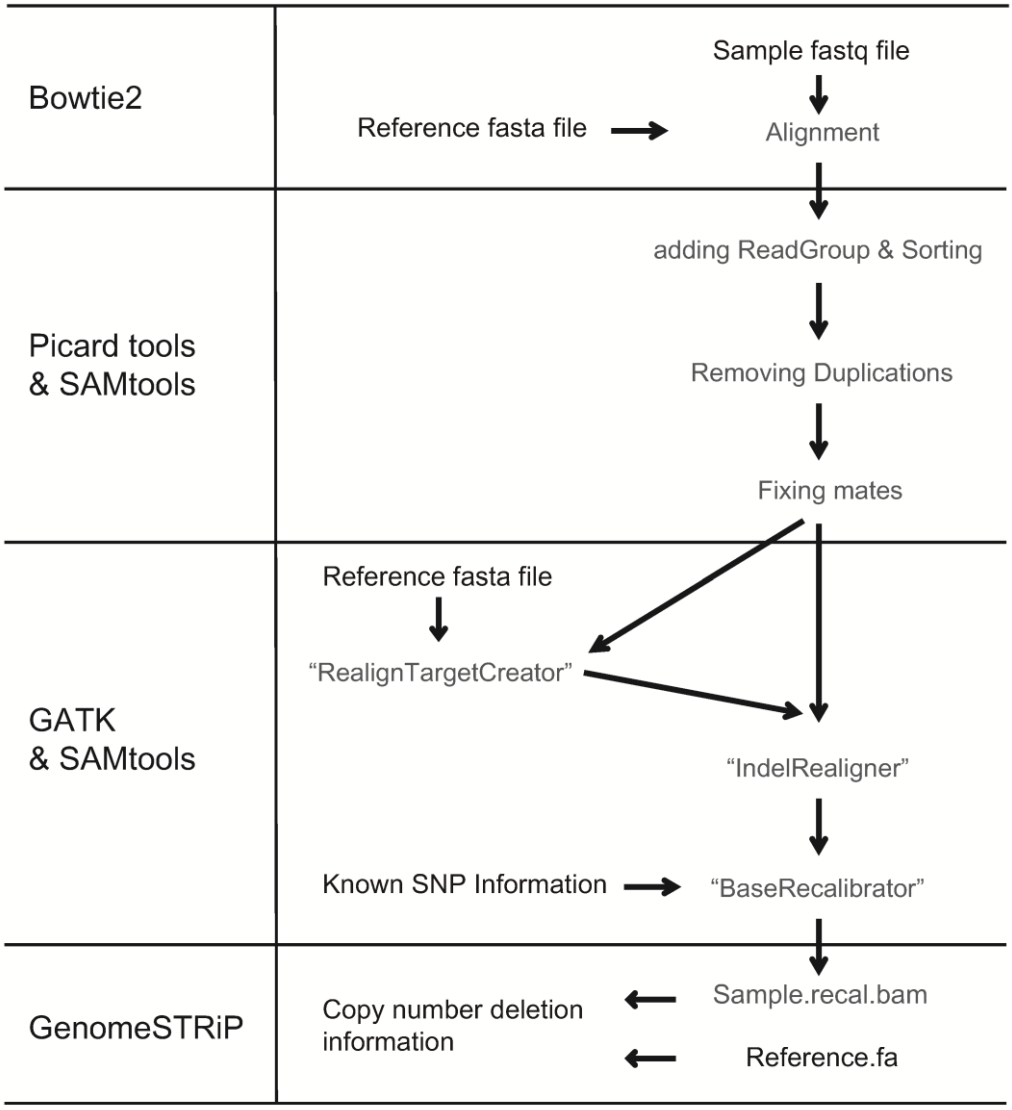
Supplementary Figure 5-1. PCR analysis of each examined CNV region. With designed primer pairs, PCR was performed on genomic DNA from four Thoroughbred racing horses (lane 1-4). M_{100} and M_{1kb} represent 100 bp- and 1 kb-size marker, respectively. Each predicted size of CNV present and deleted amplicon was marked on the right down of each gel loaded figure.



Supplementary Figure 5-2. PCR analysis of each examined CNV region. With designed primer pairs, PCR was performed on genomic DNA from four Thoroughbred racing horses (lane 1-4). M_{100} and M_{1kb} represent 100 bp- and 1 kb-size marker, respectively. Each predicted size of CNV present and deleted amplicon was marked on the right down of each gel loaded figure.



Supplementary Figure 5-3. PCR analysis of each examined CNV region. With designed primer pairs, PCR was performed on genomic DNA from four Thoroughbred racing horses (lane 1-4). M_{100} and M_{1kb} represent 100 bp- and 1 kb-size marker, respectively. Each predicted size of CNV present and deleted amplicon was marked on the right down of each gel loaded figure.



Supplementary Figure 6. DNA re-sequencing data process pipeline before Genome STRiP for CNV identification.

Supplementary Tables

Supplementary Table 1. The number of linear regression lines between CNVs and gene expressions.

	Slope	Blood		Muscle	
		Before exercise	After exercise	Before Exercise	After exercise
Cover*	Positive	69	56	68	78
	Negative	68	66	78	63
	Zero	19	34	10	15
	#NA**	73	73	73	73
	Total	229	229	229	229
	Slope	Blood		Muscle	
		Before exercise	After exercise	Before Exercise	After exercise
Inside*	Positive	2	2	4	2
	Negative	5	5	2	4
	Zero	2	2	3	3
	#NA**	5	5	5	5
	Total	14	14	14	14
	Slope	Blood		Muscle	
		Before exercise	After exercise	Before Exercise	After exercise
Front*	Positive	0	2	1	1
	Negative	3	0	1	1
	Zero	0	1	1	1
	#NA**	2	2	2	2
	Total	5	5	5	5
	Slope	Blood		Muscle	
		Before exercise	After exercise	Before Exercise	After exercise
Rear*	Positive	1	1	1	2
	Negative	1	0	2	0
	Zero	1	2	0	1
	#NA**	1	1	1	1
	Total	4	4	4	4

* The group includes genes that cover CNV regions is designated by ‘Cover’, the group includes genes that located in CNV region is designated by ’Inside’, the group includes genes that span the front part of the CNV region is designated by ’Front’ and the group includes genes that span the rear part of the CNV region is designated by ’Rear’.

** The slope of linear regression is infinite when the four horses have same CNV type at CNV region, which is

designated by “#NA’.

Supplementary Table 2. The genes show a consistent tendency in gene expression - CNVs slope of all four conditions and involved KEGG pathway.

Slope*	ENSEMBL GENE ID	HGNC symbol	KEGG pathway highest terms	KEGG pathway upper terms	KEGG pathway terms
Cover Negative	ENSECAG00000006577	*TBCK*			
	ENSECAG00000011240	*CNTN1*	Environmental Information Processing	Signaling Molecules and Interaction	Cell adhesion molecules (CAMs)**
	ENSECAG00000013878	*ANK1*			
	ENSECAG00000018274	*FAM172A*			
	ENSECAG00000019161	*DIAPH3*	Cellular Processes	Cell Motility	Regulation of actin cytoskeleton
	ENSECAG00000019334	*SLC24A3*			
	ENSECAG00000019687				
	ENSECAG00000020220	*PAK7*	Cellular Processes	Cell Motility	Regulation of actin cytoskeleton
				Cell Communication	Focal adhesion
			Organismal Systems	Development	Axon guidance
				Immune System	T cell receptor signaling pathway
			Environmental Information Processing	Signal Transduction	ErbB signaling pathway
			Human Diseases	Cancers	Renal cell carcinoma
	ENSECAG00000021157	*GARNL3*			
	ENSECAG00000022594	*SCEL*			
	ENSECAG00000022742	*APBB1IP*			
	ENSECAG00000022845	*PTPRM*	Environmental Information Processing	Signaling Molecules and Interaction	Cell adhesion molecules (CAMs)**
			Cellular Processes	Cell Communication	Adherens junction
	ENSECAG00000023254	*SUCLG2*	Metabolism	Carbohydrate Metabolism	Propanoate metabolism
					Citrate cycle (TCA cycle)
	ENSECAG00000024191	*TRAPPC9*			
	ENSECAG00000025096	*TMEM117*			
Cover Positive	ENSECAG00000000018	*KANSL1L*			
	ENSECAG00000004645	*PARK2*	Human Diseases	Neurodegenerative Diseases	Parkinson's disease
			Genetic Information Processing	Folding, Sorting and Degradation	Ubiquitin mediated proteolysis
	ENSECAG00000006492	*HLA-DQB1*	Human Diseases	Immune Diseases	Asthma
					Allograft rejection
					Autoimmune thyroid disease
					Systemic lupus erythematosus

					Graft-versus-host disease
				Cardiovascular Diseases	Viral myocarditis
				Endocrine and Metabolic Diseases	Type I diabetes mellitus
			Organismal Systems	Immune System	Intestinal immune network for IgA production
					Antigen processing and presentation
			Environmental Information Processing	Signaling Molecules and Interaction	Cell adhesion molecules (CAMs)
	ENSECAG00000006519	*SMG6*			
	ENSECAG00000008100	*CEP350*			
	ENSECAG00000008985	*NBEA*			
	ENSECAG00000012126	*RASAL2*			
	ENSECAG00000012429	*DCDC2*			
	ENSECAG00000012815	*TNPO3*			
	ENSECAG00000014475	*STK39*			
	ENSECAG00000022821	*ZNF570*			
	ENSECAG00000022893	*MAN2A1*	Metabolism	Glycan Biosynthesis and Metabolism	N-Glycan biosynthesis
Inside Negative	ENSECAG00000003638				

* The group includes genes that cover CNV regions is designated by ‘Cover’ and the group includes genes that located within CNV region is designated by ’Inside’. When the slope of linear regression of all four conditions which are ‘blood before exercise’, ‘blood after exercise’, ‘muscle before exercise’ and ‘muscle after exercise’ is positive, which means more deletion makes more gene expression, it is designated by ‘Positive’ and in the opposite case it is designated by ‘Negative’

**The enriched KEGG pathway (counts ≥2, EASE ≥0.1) is denoted by double asterisk.

Supplementary Table 3. Summary statistics of RNA sequencing read mapping.

Sample ID*	Filtered out reads	Total number of reads	Total size of read mapping (Gbp)	Number of junctions
BF1B	1202 (0.005%)	26577778	2.39	127091
BF1P	985 (0.004%)	26111110	2.35	122878
BF2B	1030 (0.004%)	26111110	2.35	128232
BF2P	1153 (0.004%)	26111110	2.35	127227
BF3B	982 (0.004%)	26111110	2.35	131639
BF3P	1175 (0.004%)	26577778	2.39	123907
BS3B	2533 (0.010%)	25833299	2.32	131578
BS3P	1435 (0.005%)	26133334	2.35	124569
MF1B	887 (0.003%)	28953336	2.61	109266
MF1P	880 (0.003%)	28218561	2.54	107565
MF2B	896 (0.003%)	28606040	2.57	109875
MF2P	771 (0.003%)	28953336	2.61	100956
MF3B	1221 (0.004%)	28953336	2.61	115153
MF3P	612 (0.002%)	28595004	2.57	96347
MS3B	938 (0.003%)	28069445	2.53	100137
MS3P	1331 (0.005%)	28069445	2.53	98567

*First ‘B’ is for blood and ‘M’ is for muscle. ‘F2’ is horse1, ‘F1’ is horse2, ‘F3’ is horse3 and ‘S3’ is horse4. Last ‘B’ is for ‘before exercise’ and ‘P’ is for ‘after exercise’.

Supplementary Table 4. Primer information used for validating extracted CNV regions.

CNV No.		Primer sequence (5´ to 3´)	Anealing temperature	No. of Chromosome	CNV location		CNV size	PCR product size (bp)	
					Start (nt)	End (nt)		Non-deleted	Deleted
CNV1	F	CACTGCATACCATAATCCACGAGATG	62 ℃	31	3949660	3949935	275	497	222
	R	TTTGCTGTTTCTGCTGTTTTTGTG							
CNV2	F	AAGCCCAAAAAGGTAGGGGTTACA	62 ℃	28	509953	510274	321	587	266
	R	GCCTGCCAACAGAGAAGATGGTT							
CNV3	F	TGAGAAGATAGTAACAACAGGGAGCAA	62 ℃	1	132713782	132714131	349	542	193
	R	TTGGAACAAGATGGCAAAATACCC							
CNV4	F	TGAAGGAGGGTCAAAAAGAAAAGAAA	62 ℃	18	74178345	74178721	376	605	229
	R	CCAGGAGGAAGCAGGTGAAGAGA							
CNV5	F	AAACGAGTGGGACGCAGTAAAAAC	62 ℃	6	861580	862007	427	643	216
	R	TCTGTGGTGGCATCTCTTCTTTCA							
CNV6	F	TGAGGAAGTAGCTGGGATGGAAC	62 ℃	29	10951559	10951988	429	710	281
	R	TCAGTACCAAATGAAGGGAGAAGACC							
CNV7	F	ACTTTCCCATCCAGACTTCCCCA	62 ℃	3	43601326	43601759	433s	641	208
	R	TATTCCTTCCCCAGACAGCAACC							
CNV8	F	TCTGTGCAGCGTTATCTCCTGTGT	62 ℃	5	96903626	96904135	509	716	207
	R	GCAGCTTCTTGTGTGTTTTTCAGG							
CNV9	F	CCTTGCTTCCTTCCCCAGTTTTT	62 ℃	20	32960247	32960777	530	742	212
	R	TCCCCTCTCTGTGTTCAATGGTG							
CNV10	F	GCAACAAAACGCCAGCATAAGG	62 ℃	18	47807410	47807947	537	883	346
	R	GGAATAGTGATAAAGTCAAGGTGGGG							
CNV11	F	CCTGGTGGTGAATGGATGCTAAA	62 ℃	2	92154131	92154677	546	783	237

	R	GCAAAACACTGCTCCCAATGCTC							
CNV12	F	CGTTCATCACCATGCCAGCC	62 ℃	29	10260417	10260979	562	773	211
	R	CGGGTTATGTAAAACTTGGTCCTTG							
CNV13	F	TACTTGGGCTCCATTCAACCTCC	62 ℃	3	75906746	75907312	566	768	202
	R	TAGCATGTCCAGCCCGTACTCAA							
CNV14	F	ACCAGATAAGCAGTGGGAGGATTT	62 ℃	9	79283047	79283628	581	754	173
	R	TCAGCAGGAGGGGAAGGAAGA							
CNV15	F	GCAATCAGCCCTGGGTTCTG	62 ℃	8	56785912	56786511	599	872	273
	R	TGCTTCTTATTAAGCACACATTCGGA							
CNV16	F	TTCAAAGAAACTTGGGGAGCCTG	62 ℃	19	35910806	35911406	600	847	247
	R	CGCTACAATATGGATGAGCCCTG							
CNV17	F	GGCAGTTTGAAGATAAGTCTGCTGG	64 ℃	9	22543035	22543796	761	1075	314
	R	ATGCAGGCCAGAGGGAAGTGA							
CNV18	F	TGAGGCTGTCCTGCAACAAAGAG	64 ℃	21	6627644	6628411	767	1060	293
	R	TCTGTGGAGCGTGGTTAAGAGGG							
CNV19	F	ATGGAATCAGAACAGCCCAGGAA	64 ℃	4	82105543	82106325	782	989	207
	R	TGAGGCTAAGGGAAGGCACTCTG							
CNV20	F	CCTGCCCCGACCCAATAAAA	58 ℃	4	83850967	83851917	950	1105	155
	R	AGCCAGCCAGAGTGGTGAGGA							
*CNV21	F	ACAGTCATTCTCAGCCCCTTGCT	64 ℃	9	11723069	11724341	1272	1499	227
	R	TTCAGTGACATTGGGTAACGGACA							
CNV22	F	TCCCATTGTGGTTTTAGTTTGGACTT	60 ℃	21	14199115	14200545	1430	1651	221
	R	CCTTTGGTAGGAGGAGGGGG							
CNV23	F	TGGTTGCCTTGAGCCTGAGAAG	62 ℃	15	65164871	65166364	1493	1652	159

	R	AGTCACCTGCCCACTTTGCTGT							
CNV24	F	GCCTCATCCCTAATCTCTTCCCC	60 ℃	22	33442484	33444335	1851	2057	206
	R	GGCCTCGTGTCTTTGGATTTTGT							
CNV25	F	TCGATGTAGTCACAAATGGCAGGA	64 ℃	14	82298279	82300232	1953	2272	319
	R	AACTGCGTATGTGGGGATGGTTT							
CNV26	F	AAAGGAGGGAATGGGGATAGGAA	62 ℃	5	13318923	13320899	1976	2296	320
	R	CCAATGTTTCTGGGTTTTGAGGA							

*The primer pairs producing over 1kb region in CNV present allele (CNV 21 to CNV 26) was amplified with longer extention condition (72℃ for 1min 40sec) compared to primer paris amplifying less than 1 kb region (CNV 1 to CNV 20, 72℃ for 1min)

Supplementary Table 5. Mapping rate and genome coverage of 18 Thoroughbred horses.

Sample		Mapping rate							Genome coverage		
		Total number of reads	Concordantly 1 time	Concordantly >1 time	Discordantly 1 time	1 time with single reads*	>1 time with single reads*	Overall mapping rate	Reference length (bp)	Matched length (bp)	Coverage
4 Thoroughbred horses (SRP017702)	HORSE1	157137184	125054630	27476129	557862	1743266	1951251	154935879.5	2367053447	2312788259	0.98
		100.00%	79.58%	17.49%	0.36%	0.55%	0.62%	98.60%			
	HORSE2	151012115	109064309	25632903	10176908	2638165	5049374	148717889.5	2367053447	2320687186	0.98
		100.00%	72.22%	16.97%	6.74%	0.87%	1.67%	98.48%			
	HORSE3	148371662	115594429	25752630	2588733	1863084	2543036	146138852	2367053447	2319883129	0.98
		100.00%	77.91%	17.36%	1.74%	0.63%	0.86%	98.50%			
	HORSE4	141392469	111521219	24155804	1936798	1715175	2191156	139566986.5	2367053447	2321797613	0.98
		100.00%	78.87%	17.08%	1.37%	0.61%	0.77%	98.71%			
14 Thoroughbred horses (SRA053569)	NGS1	221234170	169743657	32157454	2005982	7268293	3422650	209252564.5	2367053447	2330056676	0.98
		100.00%	76.73%	14.54%	0.91%	1.64%	0.77%	94.58%			
	NGS2	225716953	174828510	28832594	2579791	7009823	3409808	211450710.5	2367053447	2327685469	0.98
		100.00%	77.45%	12.77%	1.14%	1.55%	0.76%	93.68%			
	NGS3	229936248	178607008	29964418	2895609	7967904	3759179	217330576.5	2367053447	2326890963	0.98
		100.00%	77.68%	13.03%	1.26%	1.73%	0.82%	94.52%			
	NGS4	223786527	169856826	31499097	2906547	8446841	4099115	210535448	2367053447	2329166757	0.98
		100.00%	75.90%	14.08%	1.30%	1.89%	0.92%	94.08%			
	NGS5	220001399	169625446	30501074	2624217	7399155	3589195	208244912	2367053447	2328576671	0.98
		100.00%	77.10%	13.86%	1.19%	1.68%	0.82%	94.66%			
	NGS6	245242121	190552860	33379461	2911065	7548974	3616236	232425991	2367053447	2328854903	0.98
		100.00%	77.70%	13.61%	1.19%	1.54%	0.74%	94.77%			

NGS7	255719715	190504335	36696560	5198256	11842825	6427277	241534202	2367053447	2328175571	0.98
	100.00%	74.50%	14.35%	2.03%	2.32%	1.26%	94.45%			
NGS8	203687722	157087041	27738084	1948774	6645217	2976130	191584572.5	2367053447	2327153598	0.98
	100.00%	77.12%	13.62%	0.96%	1.63%	0.73%	94.06%			
NGS9	208505749	162313811	28633972	2243384	6270115	3143508	197897978.5	2367053447	2326505656	0.98
	100.00%	77.85%	13.73%	1.08%	1.50%	0.75%	94.91%			
NGS10	200887652	158694981	28112280	3813662	5602430	3436949	195140612.5	2367053447	2329506207	0.98
	100.00%	79.00%	13.99%	1.90%	1.39%	0.86%	97.14%			
NGS11	246618258	187432836	34421705	4097626	8668977	4913330	232743320.5	2367053447	2328826699	0.98
	100.00%	76.00%	13.96%	1.66%	1.76%	1.00%	94.37%			
NGS12	254051029	196701224	35599856	3978850	8541394	4548670	242824962	2367053447	2329960006	0.98
	100.00%	77.43%	14.01%	1.57%	1.68%	0.90%	95.58%			
NGS13	244207405	192915298	34735280	2916637	7067929	3475827	235839093	2367053447	2329718870	0.98
	100.00%	79.00%	14.22%	1.19%	1.45%	0.71%	96.57%			
NGS14	240711458	189008118	35472466	3923502	7330309	3971448	234054964.5	2367053447	2330244489	0.98
	100.00%	78.52%	14.74%	1.63%	1.52%	0.82%	97.23%			

*The number of single reads mapped was presented. It was counted as 0.5 of read pair for the overall mapping rate.