

Fig. S1. The composition of combined set.

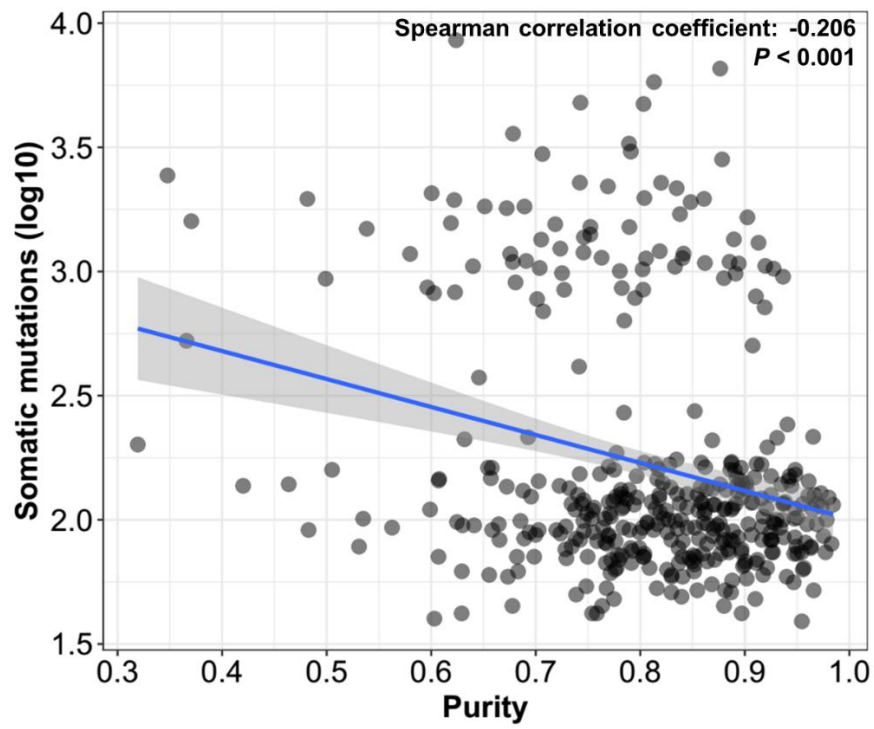


Fig. S2. The correlation between tumor purity and somatic mutations.

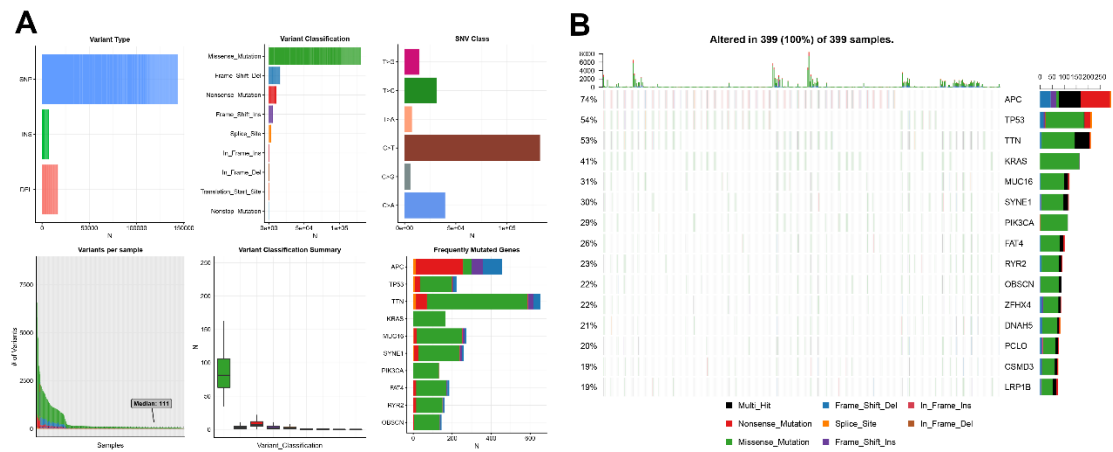


Fig. S3. The mutation profile of colon cancer samples in TCGA dataset.

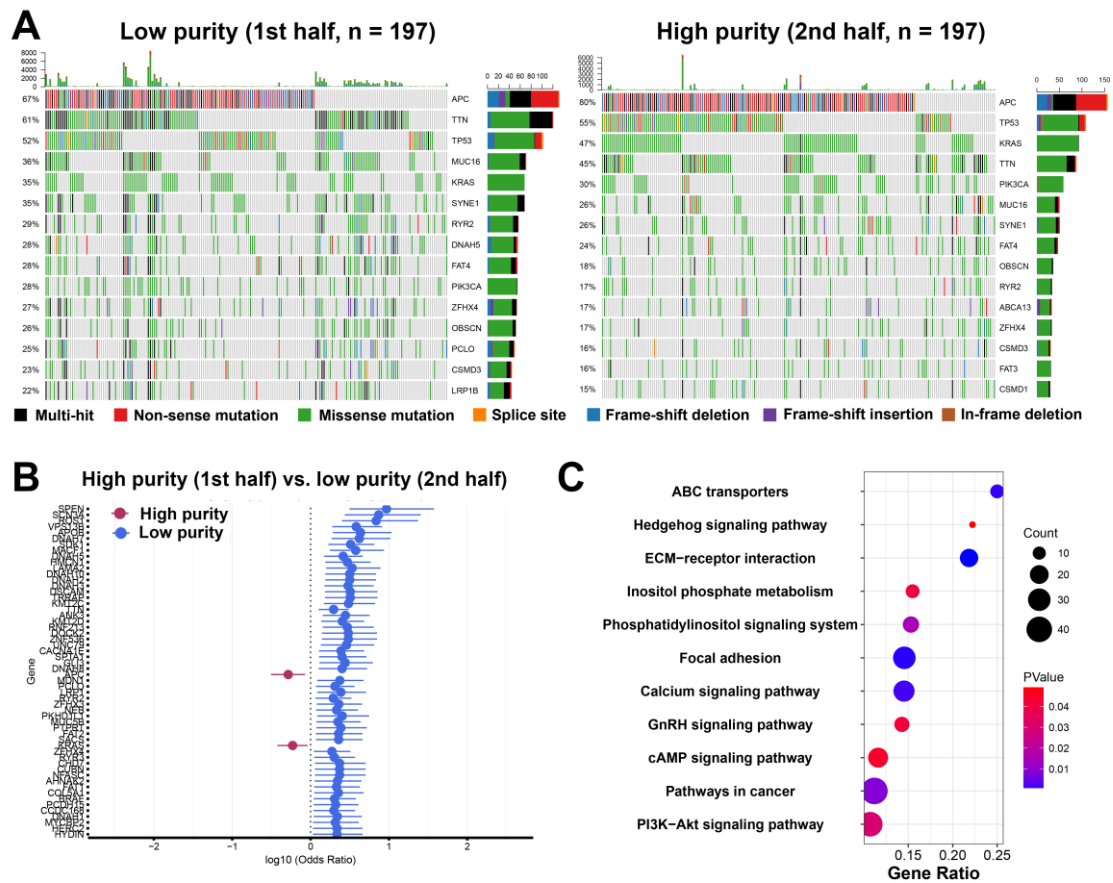


Fig. S4. Mutation profile between low and high purity groups (1st vs. 2nd half) in TCGA dataset. (A) Mutation profile of low and high purity groups. (B) Differentially mutated genes between low and high purity groups. (C) Highly mutated pathways in low purity group.

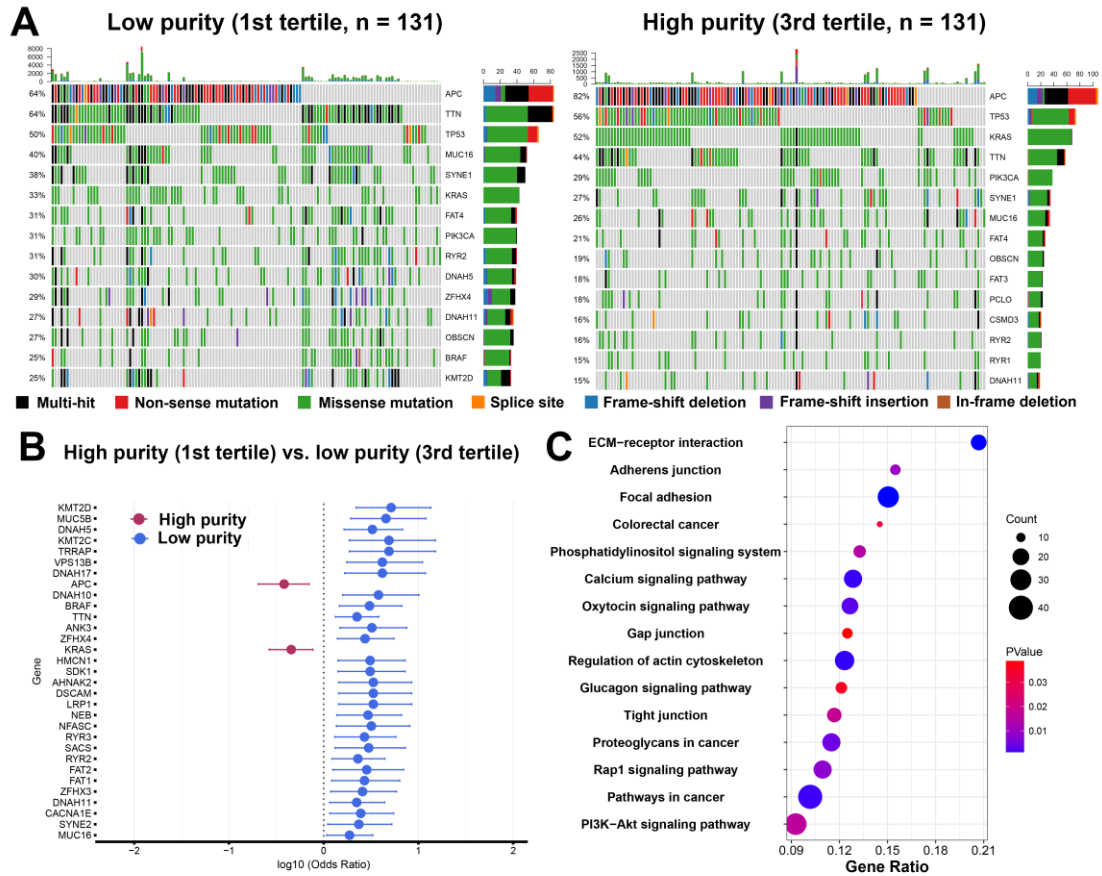


Fig. S5. Mutation profile between low and high purity groups (1st vs. 3rd tertile) in TCGA dataset. (A) Mutation profile of low and high purity groups. (B) Differentially mutated genes between low and high purity groups. (C) Highly mutated pathways in low purity group.

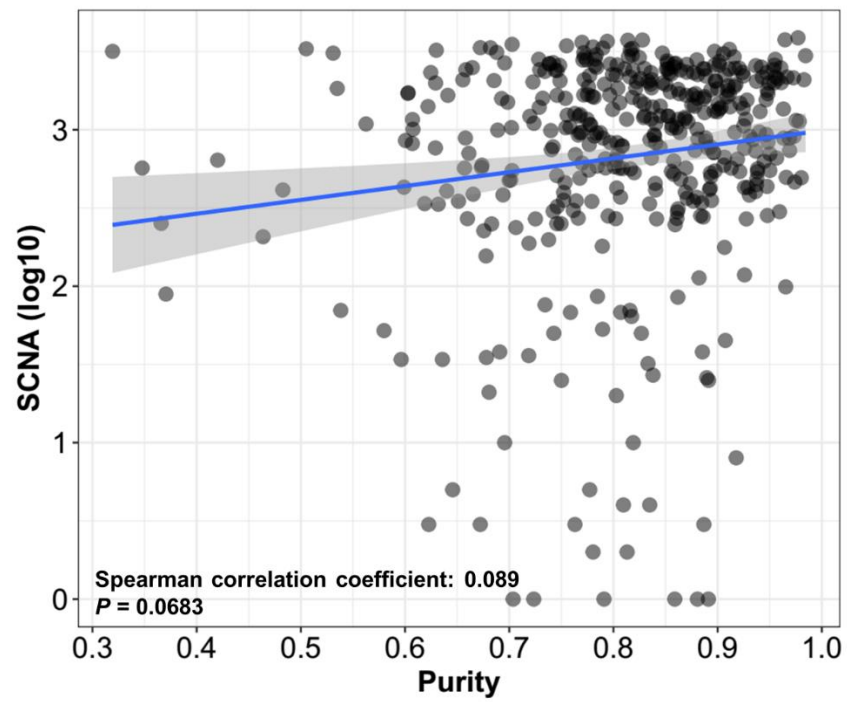


Fig. S6. The correlation between tumor purity and somatic copy number alteration (SCNA).

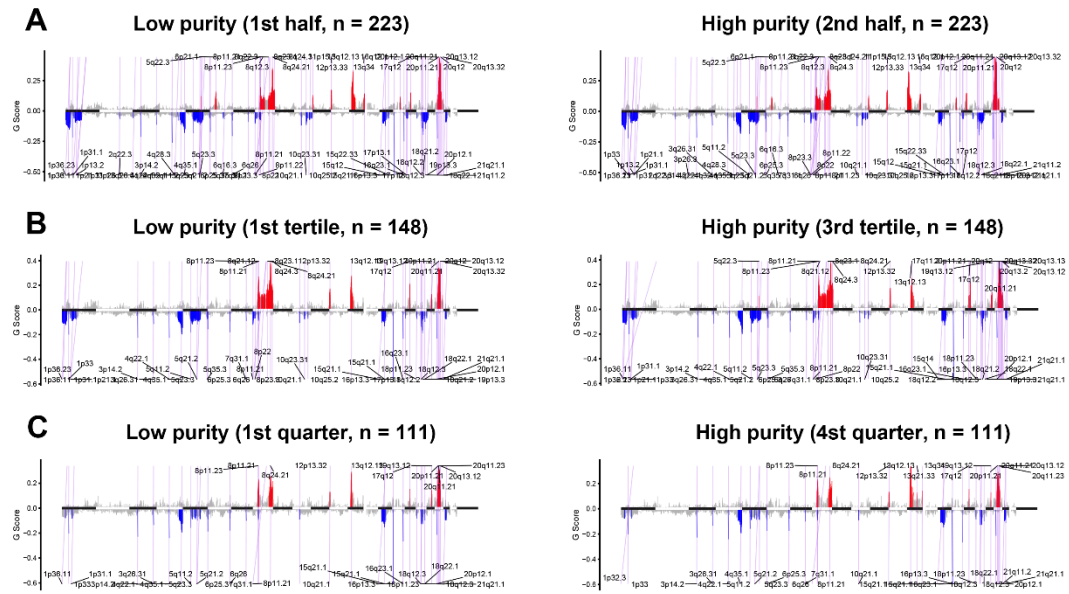


Fig. S7. SCNA profile comparison between low and high purity groups. (A) 1st vs. 2nd half. (B) 1st vs. 3rd tertile. (C) 1st vs. 4th quarter.

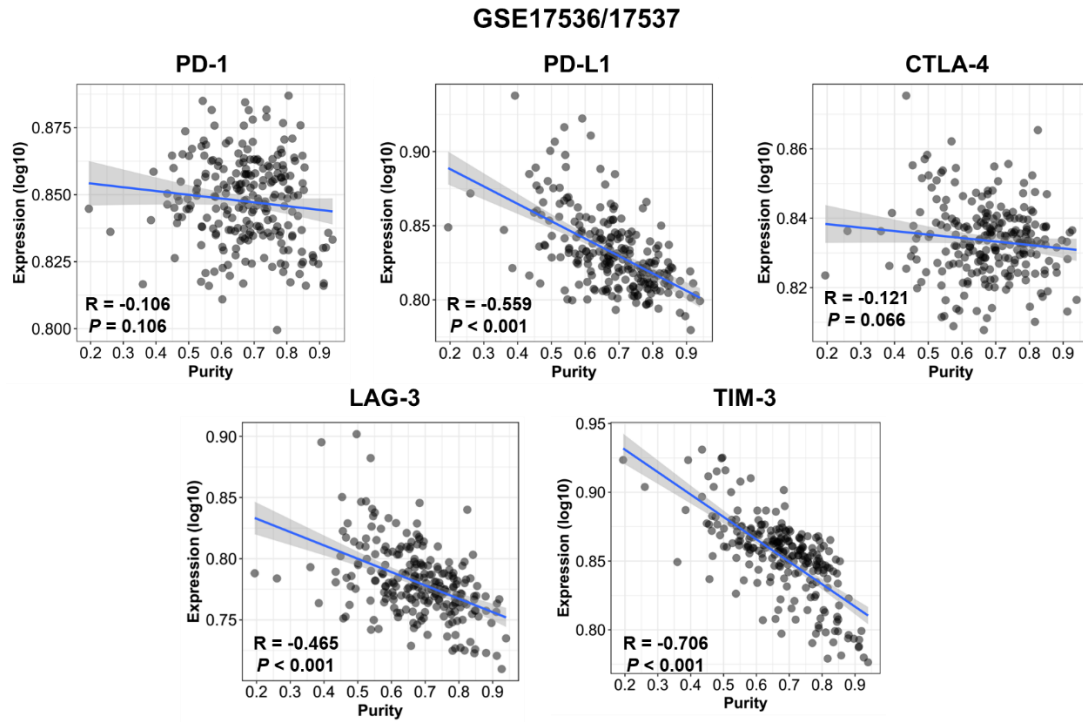


Fig. S8. The expression level of immunotherapy-associated genes were inversely correlated with tumor purity in GSE17536/17537 dataset.

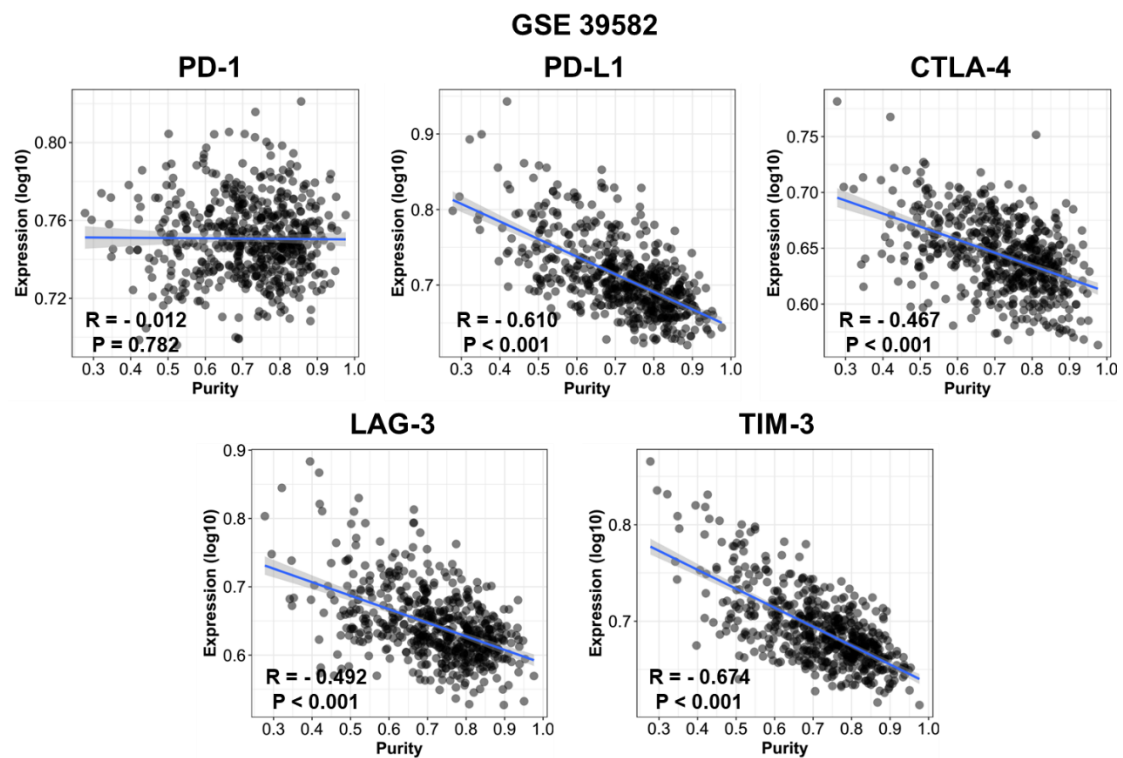


Fig. S9. The expression level of immunotherapy-associated genes were inversely correlated with tumor purity in GSE39582 dataset.

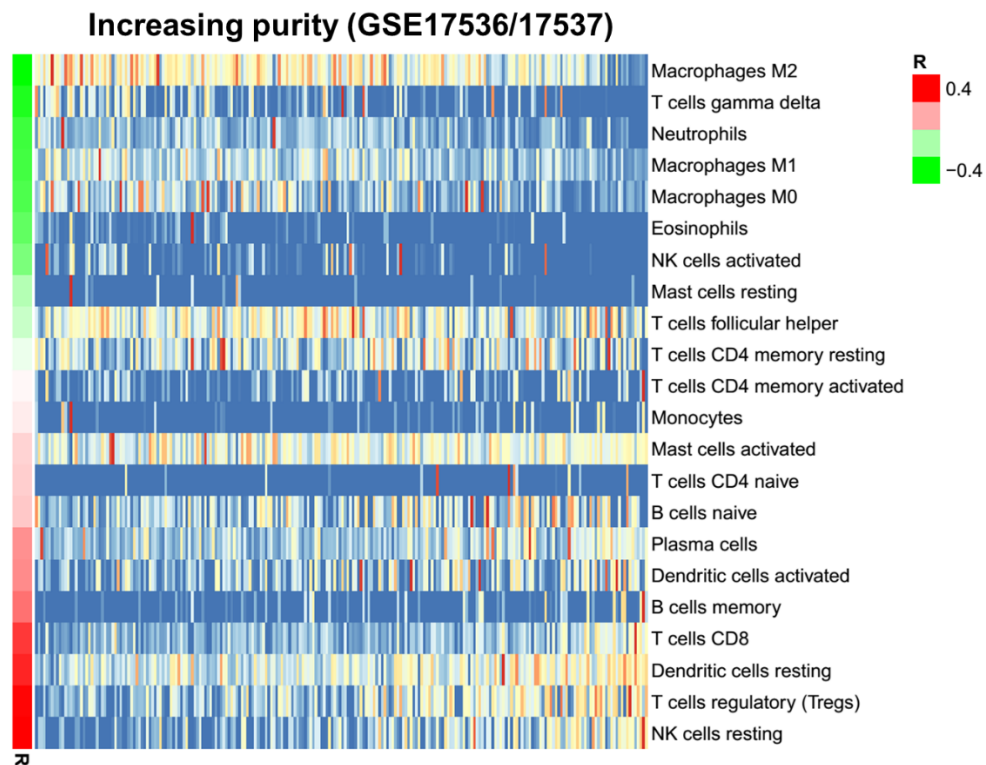


Fig. S10. Correlation between CC purity and different types of immune cells in GSE17536/17537 dataset.

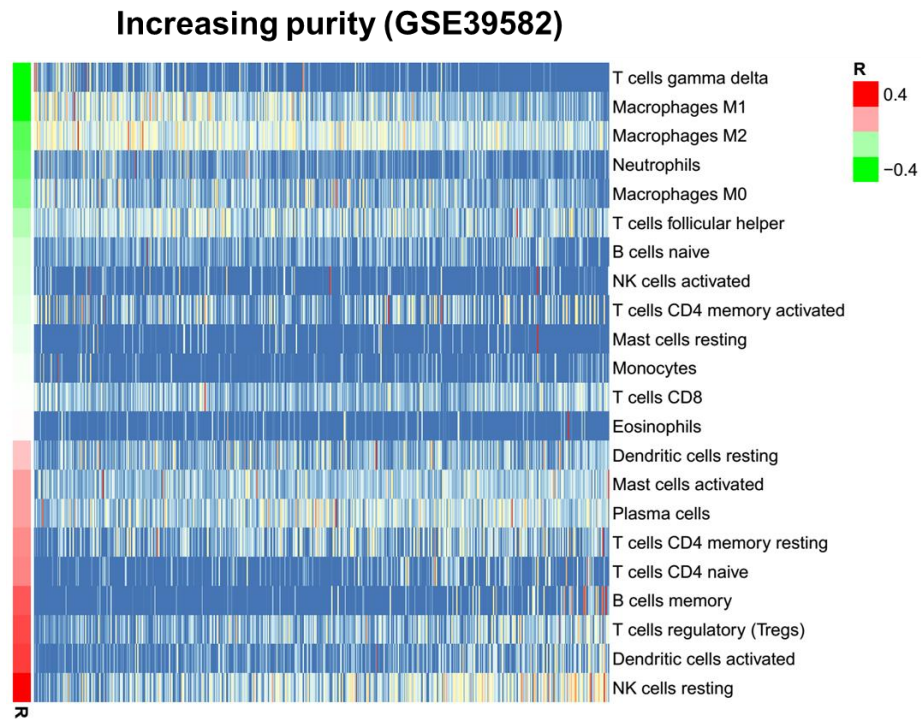


Fig. S11. Correlation between CC purity and different types of immune cells in GSE39582 dataset.

Table S1. Baseline clinicopathological characteristics of enrolled patients from three public datasets.

Dataset	GSE39582 (%)	GSE17536/17537 (%)	TCGA (%)	Combined (%)
All patients	562	232	454	1248
Age				
≤60	157 (27.9)	84 (36.2)	137 (30.2)	378 (30.3)
>60	404 (71.9)	148 (63.8)	317 (69.2)	869 (69.6)
Unknown	1 (0.2)			1 (0.1)
Gender				
Male	307 (54.6)	122 (52.6)	240 (52.9)	669 (53.6)
Female	255 (45.4)	110 (47.4)	214 (47.1)	579 (46.4)
Primary site				
Right-sided colon	222 (39.5)	Unavailable	256 (56.4)	478 (38.3)
Left-sided colon	340 (60.5)		178 (39.2)	518 (41.5)
Unknown			20 (4.4)	252 (20.2)
Stage				
1	33 (5.9)	28 (12.1)	76 (16.7)	137 (11.0)
2	264 (47.0)	72 (31.0)	178 (39.2)	514 (41.2)
3	205 (36.5)	76 (32.8)	128 (28.2)	409 (32.8)
4	60 (10.7)	56 (24.1)	64 (14.1)	180 (14.4)
Unknown			8 (1.8)	8 (0.6)
MMR status				
MMR-proficient	441 (78.5)	Unavailable	283 (62.3)	724 (58.0)
MMR-deficient	74 (13.2)		56 (12.3)	130 (10.4)
Unknown	47 (8.4)		115 (25.3)	394 (31.6)
Chemotherapy				
No	312 (55.5)	Unavailable	240 (52.9)	552 (44.2)
Yes	233 (41.5)		146 (32.2)	379 (30.4)
Unknown	17 (3.0)		68 (15.0)	317 (25.4)
Median follow-up time (months)	52.0	46.7	18.3	34.0

Table S2. Relationship between clinicopathological characteristics and tumor purity in microarray and TCGA dataset.

	Microarray dataset		
Factors	Total (%)	Purity (mean, 95%CI)	P value
All patients	794	0.705 (0.696 - 0.714)	
Age (years)			0.758
≤60	241 (30.4)	0.703 (0.686 - 0.719)	
>60	552 (69.5)	0.706 (0.695 - 0.717)	
Gender			0.418
Male	429 (54.0)	0.708 (0.696 - 0.720)	
Female	365 (46.0)	0.701 (0.687 - 0.714)	
Primary site			<0.001
Right-sided colon	222 (28.0)	0.692 (0.673 - 0.711)	
Left-sided colon	340 (42.8)	0.732 (0.720 - 0.745)	
Stage			0.182
I	61 (7.7)	0.709 (0.681 - 0.736)	
II	336 (42.3)	0.714 (0.700 - 0.728)	
III	281 (35.4)	0.691 (0.675 - 0.708)	
IV	116 (14.6)	0.709 (0.686 - 0.731)	
MMR status			<0.001
pMMR	441 (55.5)	0.734 (0.722 - 0.745)	
dMMR	74 (9.3)	0.655 (0.623 - 0.687)	
	TCGA dataset		
Factors	Total (%)	Purity (mean, 95%CI)	P value
All patients	454	0.811 (0.801 - 0.821)	
Age (years)			0.698
≤60	137 (30.2)	0.808 (0.787 - 0.829)	
>60	317 (69.8)	0.812 (0.800 - 0.824)	
Gender			0.210
Male	240 (52.9)	0.817 (0.802 - 0.831)	
Female	214 (47.1)	0.804 (0.789 - 0.819)	
Primary site			0.008
Right-sided colon	256 (56.4)	0.801 (0.786 - 0.816)	
Left-sided colon	178 (39.2)	0.829 (0.814 - 0.843)	
Stage			0.283
I	76 (16.7)	0.822 (0.802 - 0.843)	
II	178 (39.2)	0.801 (0.784 - 0.819)	
III	128 (28.2)	0.808 (0.787 - 0.828)	
IV	64 (14.1)	0.833 (0.809 - 0.858)	
MMR status			0.780
pMMR	283 (62.3)	0.808 (0.797 - 0.819)	
dMMR	56 (12.3)	0.814 (0.775 - 0.853)	

Table S3. Univariate and multivariate Cox regression analyses for overall survival in patients with adjuvant chemotherapy in microarray dataset.

	Univariate analysis		Multivariate analysis	
Factors	HR (95% CI)	<i>P</i>	HR (95% CI)	<i>P</i>
Age (years)				
Increasing years	1.009 (0.990 - 1.029)	0.353		
Gender				
Male	1 (reference)			
Female	0.818 (0.515 - 1.300)	0.395		
Primary site				
Right-sided colon	1 (reference)			
Left-sided colon	0.767 (0.477 - 1.233)	0.273		
Stage				
II	1 (reference)		1 (reference)	
III	1.097 (0.614 – 1.959)	0.755	1.156 (0.645 – 2.071)	0.627
IV	8.377 (4.113 – 17.061)	<0.001	8.508 (4.178 - 17.324)	<0.001
MMR status				
MMR-proficient	1 (reference)			
MMR-deficient	0.937 (0.341 – 2.575)	0.899		
Purity				
Low purity	1 (reference)		1 (reference)	
High purity	0.608 (0.386 – 0.958)	0.032	0.621 (0.393 – 0.980)	0.041

Table S4. Top pathways enriched in low purity group.

Name of pathway	NES	NOM p-val	FDR q-val	FWER p-val
Lymphocyte chemotaxis	2.015658	< 0.001	< 0.001	< 0.001
Chemokine mediated signaling pathway	1.978716	< 0.001	< 0.001	< 0.001
Lymphocyte migration	1.955421	< 0.001	< 0.001	< 0.001
Regulation of monocyte chemotaxis	1.936586	< 0.001	< 0.001	< 0.001
Collagen fibril organization	1.930819	< 0.001	< 0.001	< 0.001
Lymphocyte costimulation	1.924046	< 0.001	< 0.001	< 0.001
Positive regulation of lymphocyte migration	1.920384	< 0.001	< 0.001	< 0.001
Positive regulation of leukocyte mediated immunity	1.92003	< 0.001	< 0.001	< 0.001
Positive regulation of cytokine secretion	1.919398	< 0.001	< 0.001	0.001
Myd88 dependent toll like receptor signaling pathway	1.914614	< 0.001	< 0.001	0.001
Cytokine secretion	1.913273	< 0.001	< 0.001	0.001
Positive regulation of inflammatory response	1.91126	< 0.001	< 0.001	0.001
Adaptive immune response	1.909809	< 0.001	< 0.001	0.001
Regulation of lymphocyte migration	1.907982	< 0.001	< 0.001	0.001
Regulation of tumor necrosis factor superfamily cytokine production	1.902731	< 0.001	< 0.001	0.003
Positive regulation of tumor necrosis factor superfamily cytokine production	1.901051	< 0.001	< 0.001	0.003
Cellular defense response	1.891724	< 0.001	< 0.001	0.005
Positive regulation of leukocyte chemotaxis	1.883404	< 0.001	< 0.001	0.006
Cellular response to interferon gamma	1.881093	< 0.001	< 0.001	0.007

Table S5. Correlation coefficients of tumor purity and relative proportion of immune cells.

Cell type	TCGA	GSE17536/17537	GSE39582
Macrophages M1	-0.3193	-0.3108	-0.4123
Macrophages M2	-0.3564	-0.4392	-0.2704
Neutrophils	-0.2378	-0.3201	-0.2481
Plasma cells	0.2654	0.2337	0.1668
Dendritic cells activated	0.1763	0.2434	0.3257
Mast cells activated	0.2157	0.1036	0.1611
T cells CD4 memory resting	0.3082	-0.0007	0.2013
NK cells resting	0.0726	0.4961	0.4321
T cells regulatory (Tregs)	0.1498	0.4834	0.3056
B cells naive	0.0991	0.1229	-0.0696
T cells CD8	0.0613	0.3925	0.0001
Monocytes	0.2506	0.0568	-0.0078
Eosinophils	0.0915	-0.2552	0.0075
NK cells activated	0.1102	-0.2093	-0.0605
T cells CD4 memory activated	0.0230	0.0443	-0.0384
Dendritic cells resting	0.0270	0.4221	0.1041
T cells follicular helper	0.0270	-0.0665	-0.1155
T cells gamma delta	-0.0455	-0.3808	-0.4187
B cells memory	-0.1216	0.2883	0.2850
T cells CD4 naive	-0.0368	0.1212	0.2089
Macrophages M0	-0.1680	-0.2905	-0.1894
Mast cells resting	-0.1616	-0.1116	-0.0237