SRP072112

In the published article (PMID: 27749841; DOI: 10.1038/ng.3683), there were 13 patients in analysis (ESCC01 to ESCC13).

"In total, 51 tumor regions and 13 matched morphologically normal esophageal tissues (four tumor regions and one matched normal tissue per case, with the exception of ESCC04, which only had three tumor regions) were sequenced"

But SRP072112 totally included 55 samples from 11 patients (ESCC03 to ESCC13, no ESCC01 and ESCC02).

The SRR3270839 (ESCC04-N) and SRR3270842 (ESCC04-T3) were not from the same individual in match test, thus SRR3270842 was removed from analysis, which was also excluded in the original published article (ESCC04 only had three tumor regions, ESCC04-T1, ESCC04-T2 and ESCC04-T4).

Run name	**Sample name**
SRR3270839	ESCC04-N
SRR3270843	ESCC04-T4
SRR3270842	ESCC04-T3 (removed)
SRR3270841	ESCC04-T2
SRR3270840	ESCC04-T1

SRP179388

The following samples were removed from analysis due to absence of normal control.

Run name	**Sample name**
SRR8442528	patient3_T1
SRR8442527	patient3_T2
SRR8442544	patient7_T1
SRR8442521	patient7_T2
SRR8442520	patient8_T1
SRR8442522	patient8_T2

SRP327447

There were totally 46 pairs of tumor and matched normal samples of WES (as follow). The other files were low coverage WGS data for CNV analysis, which were not used in our analysis.

Tumor	Normal	Patients
SRR15068498	SRR15068322	P104
SRR15068497	SRR15068321	P107
SRR15068402	SRR15068320	P12
SRR15068391	SRR15068319	P126
SRR15068380	SRR15068318	P128
SRR15068369	SRR15068317	P15
SRR15068358	SRR15068316	P16
SRR15068347	SRR15068315	P17
SRR15068336	SRR15068495	P19
SRR15068325	SRR15068494	P20
SRR15068496	SRR15068493	P21
SRR15068485	SRR15068492	P22
SRR15068474	SRR15068491	P23
SRR15068463	SRR15068490	P24
SRR15068452	SRR15068489	P26
SRR15068441	SRR15068488	P27
SRR15068430	SRR15068487	P28
SRR15068419	SRR15068486	P30
SRR15068408	SRR15068484	P31
SRR15068403	SRR15068483	P32
SRR15068401	SRR15068482	P37
SRR15068400	SRR15068481	P39
SRR15068399	SRR15068480	P4
SRR15068398	SRR15068479	P44
SRR15068397	SRR15068478	P47
SRR15068396	SRR15068477	P48
SRR15068395	SRR15068476	P49
SRR15068394	SRR15068475	P5
SRR15068393	SRR15068473	P52
SRR15068392	SRR15068472	P54
SRR15068390	SRR15068471	P56
SRR15068389	SRR15068470	P57
SRR15068388	SRR15068469	P61
SRR15068387	SRR15068468	P62
SRR15068386	SRR15068467	P63

SRR15068385	SRR15068466	P74
SRR15068384	SRR15068465	P75
SRR15068383	SRR15068464	P76
SRR15068382	SRR15068462	P79
SRR15068381	SRR15068461	P80
SRR15068379	SRR15068460	P82
SRR15068378	SRR15068459	P84
SRR15068377	SRR15068458	P87
SRR15068376	SRR15068457	P89
SRR15068375	SRR15068456	P9
SRR15068374	SRR15068455	P94

SRP099292

The non-tumor samples (NTDs, LDs, HDs) samples and LD or HD) were not used.
The low coverage WGS data for CNV analysis were also not used in our analysis.
The different reads files from the same library were combined to one sample, such as the fastq files from SRR5306379 and SRR7944324 were combined to P1_C2 sample before mapping to genome.
This dataset was separated into SRP099292_S (single tumor, 36 patients) and SRP099292_M (multi-tumor,9 patients). The used samples were listed as follow.

SRP099292_S

Tumor sample	Normal sample	patients
P11_C1	P11_B	P11
P12_C2	P12_B	P12
P14_C1	P14_B	P14
P15_C1	P15_B	P15
P17_C1	P17_B	P17
P18_C1	P18_B	P18
P19_C1	P19_B	P19
P2_C1	P2_B	P2
P21_C1	P21_B	P21
P3_C1	P3_B	P3
P8_C1	P8_N	P8
SP1_C	SP1_B	SP1
SP10_C	SP10_B	SP10
SP11_C	SP11_B	SP11
SP13_C	SP13_B	SP13
SP14_C	SP14_B	SP14
SP15_C	SP15_B	SP15

SP16_C	SP16_B	SP16
SP19_C	SP19_B	SP19
SP2_C	SP2_B	SP2
SP20_C	SP20_B	SP20
SP21_C	SP21_B	SP21
SP22_C	SP22_B	SP22
SP23_C	SP23_B	SP23
SP24_C	SP24_B	SP24
SP25_C	SP25_B	SP25
SP28_C	SP28_B	SP28
SP3_C	SP3_B	SP3
SP30_C	SP30_B	SP30
SP4_C	SP4_B	SP4
SP5_C	SP5_B	SP5
SP6_C	SP6_B	SP6
SP7_C	SP7_B	SP7
SP8_C	SP8_B	SP8
SP9_C	SP9_B	SP9

SRP099292_M

Tumor samples	Normal sample	patients
P1_C1, P1_C2, P1_C3, P1_C4	P1_B	P1
P13_C1, P13_C3	P13_B	P13
P16_C2; P16_C3; P16_C4; P16_C8	P16_B	P16
P24_C1, P24_C2, P24_C3	P24_B	P24
P25_C1, P25_C2	P25_N1	P25
P4_C1, P4_C2	P4_B	P4
P5_C1, P5_C2	P5_B	P5
P6_C1, P6_C2	P6_B	P6
P9_C1, P9_C3	P9_B	P9

SRP127593 (and SRP127837)

The sequence data from one cohort were separated into two projects, PRJNA421846(SRP127593) and PRJNA427604(SRP127837). The SRP127593 included ESCC samples while the SRP127837 included control blood samples. Their matching relationships were obtained from the supplementary table of published (PMID: 32561500 DOI: 10.1016/j.ijrobp.2020.06.015). But in our match test, we noticed 8 sample pairs were not from the same individual, which were marked in the following list.

Tumor runs (SRP127593)	Normal runs (SRP127837)	Name	Note
SRR6423648	SRR6435665	13A	
SRR6423654	SRR6435685	48A	
SRR6423652	SRR6435663	27A	
SRR6423650	SRR6435664	17A	Not match
SRR6423713	SRR6435654	100A	
SRR6423710	SRR6435693	53A	
SRR6423708	SRR6435681	69A	
SRR6423608	SRR6435669	23A	Not match
SRR6423609	SRR6435666	65A	
SRR6423606	SRR6435671	47A	
SRR6423630	SRR6435682	22A	
SRR6423731	SRR6435668	9A	
SRR6423727	SRR6435677	11A	
SRR6423628	SRR6435676	6A	
SRR6423624	SRR6435670	8A	
SRR6423661	SRR6435667	51A	
SRR6423663	SRR6435674	58A	
SRR6423664	SRR6435679	57A	
SRR6423613	SRR6435678	63A	
SRR6423618	SRR6435672	78A	
SRR6423689	SRR6435675	60A	Not match
SRR6423691	SRR6435680	76A	Not match
SRR6423636	SRR6435686	41A	
SRR6423635	SRR6435688	39A	
SRR6423633	SRR6435661	33A	Not match
SRR6423704	SRR6435662	25A	
SRR6423703	SRR6435687	46A	
SRR6423701	SRR6435691	35A	
SRR6423698	SRR6435689	44A	
SRR6423696	SRR6435690	45A	Not match
SRR6423682	SRR6435657	103A	
SRR6423676	SRR6435659	118A	Not match
SRR6423589	SRR6435660	64A	
SRR6423582	SRR6435656	79A	
SRR6423583	SRR6435673	77A	
SRR6423585	SRR6435655	81A	
SRR6423720	SRR6435683	98A	
SRR6423602	SRR6435684	32A	
SRR6423647	SRR6435692	50A	Not match
SRR6423732	SRR6435658	87B	

SRP033394

Only WES samples were included.

Totally 19 pairs of samples were included. The patient of **ESCC-D21** were listed in their published article (PMID: 24686850, DOI: 10.1038/ng.2935), but could not be found in this public dataset.