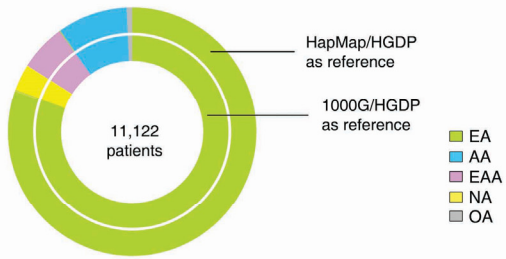


**Supplemental Information**

**Integrated Analysis of Genetic Ancestry  
and Genomic Alterations across Cancers**

**Jiao Yuan, Zhongyi Hu, Brandon A. Mahal, Sihai D. Zhao, Kevin H. Kensler, Jingjiang Pi, Xiaowen Hu, Youyou Zhang, Yueying Wang, Junjie Jiang, Chunsheng Li, Xiaomin Zhong, Kathleen T. Montone, Guoqiang Guan, Janos L. Tanyi, Yi Fan, Xiaowei Xu, Mark A. Morgan, Meixiao Long, Yuzhen Zhang, Rugang Zhang, Anil K. Sood, Timothy R. Rebbeck, Chi V. Dang, and Lin Zhang**

**A****B**

**Figure S1, related to Figure 1. Genetic ancestry of TCGA specimens inferred by EIGENSTRAT.**

**(A)** Distribution of genetic ancestry groups of the TCGA patients using different reference cohorts. The proportion of genetic ancestry groups (EA, AA, EAA, NA and OA) categorized by EIGENSTRAT is represented by a circle plot. The inner layer represents the result using the 1000 Genomes Project as reference populations. The outer layer represents the result using the HapMap Project as reference populations. For both analyses, the HGDP Project (American panel) was used as a reference for Native American. **(B)** Genetic ancestry of TCGA specimens inferred by EIGENSTRAT. Bar plots show the number of TCGA patients categorized into each of the four genetic ancestry groups (EA, NA, EAA and AA) as estimated by EIGENSTRAT across the TCGA cohort. SIRE information is color-coded by green (White), pink (Asian), blue (Black), orange (AI/AN), and grey (unavailable). The proportion of SIRE is also represented with a circle plot.

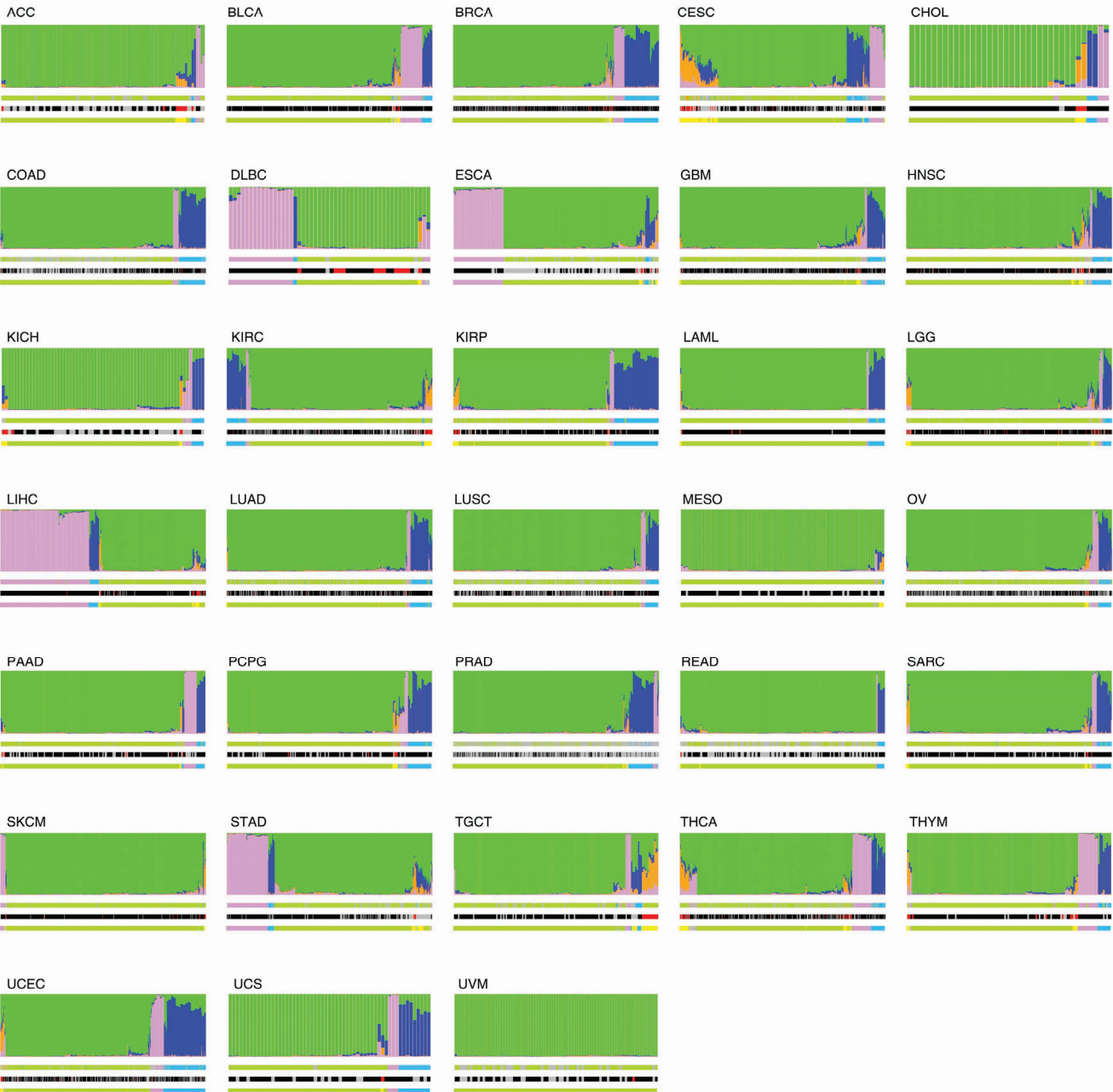
Ancestral reference ■ European ■ East Asian ■ West African ■ Native American



Self-identified race ■ White ■ Asian ■ Black ■ AI/AN ■ Unavailable

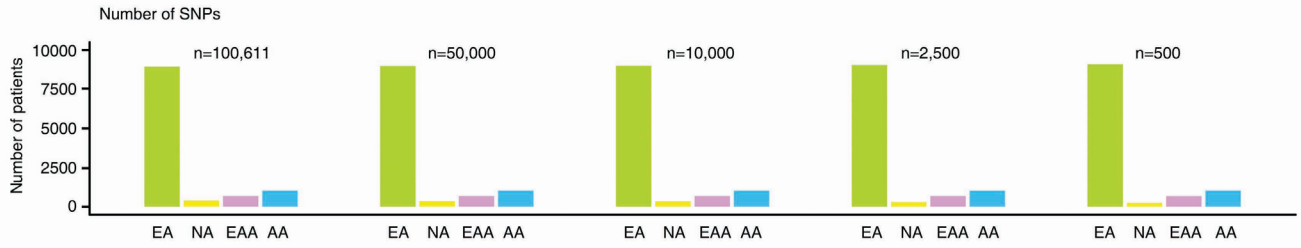
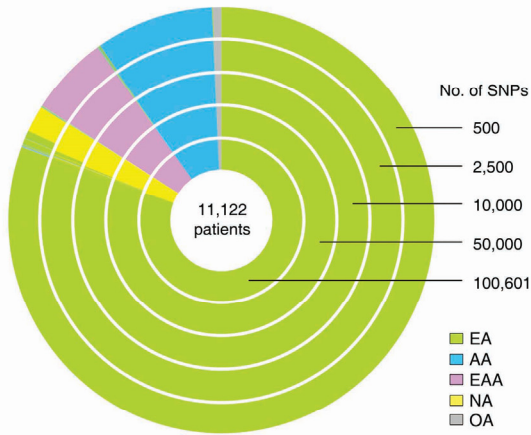
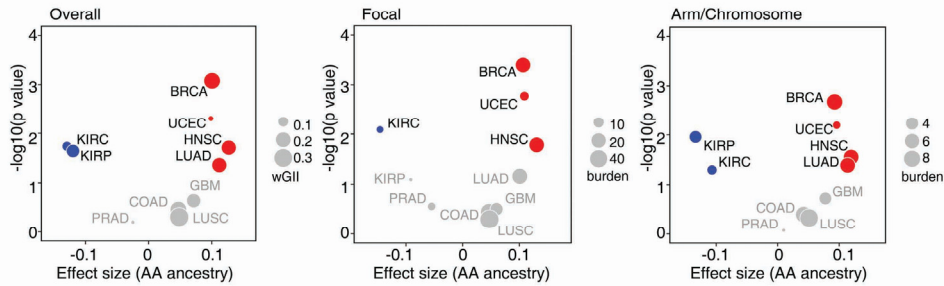
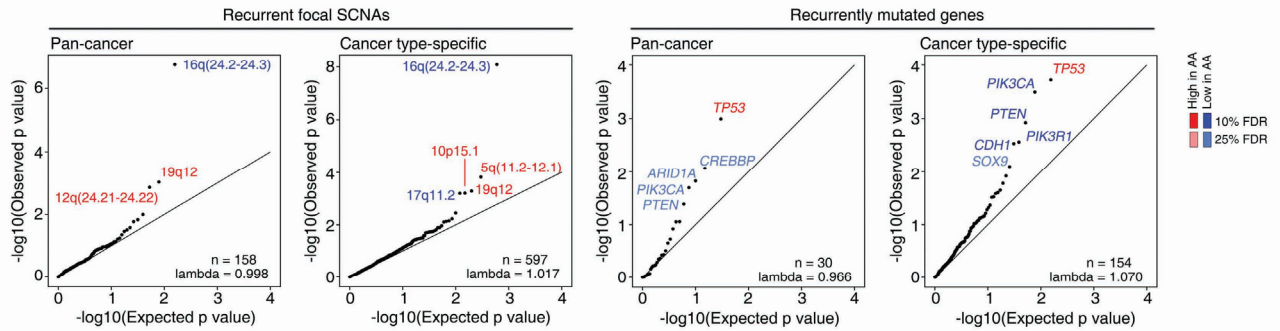
Self-identified ethnicity ■ Hispanic ■ non-Hispanic ■ Unavailable

Genetic ancestry ■ EA ■ EAA ■ AA ■ NA ■ OA



**Figure S2, related to Figure 1. Genetic ancestry of TCGA specimens inferred by STRUCTURE.**

Each color represents one of the ancestral reference groups. Each patient is represented by a column partitioned into different colors corresponding to his/her genetic ancestry composition. Patients are ordered following a hierarchical clustering by Ward's methods on distance matrix calculated as cosine dissimilarity of genetic composition. SIRE and genetic ancestry categorization as estimated by EIGENSTRAT for each patient are shown in the same order at the bottom.

**A****B****C****D**

**Figure S3, related to Figures 1, 4 and 5. AA genetic ancestry and association with genomic features.** (A and B) Distribution of genetic ancestry groups of the TCGA patients using different numbers of the SNPs. Bar plots show the results of genetic ancestry prediction based on down-sampled sets of SNPs (A). The number of the SNPs (n) used in each set is labeled in each bar plot. Circle plots show the proportion of genetic ancestry groups (EA, AA, EAA, NA and OA) categorized by EIGENSTRAT (B). Each layer represents the result applying a set of SNPs with a designated size. The innermost layer represents the result using the maximum number of SNPs we obtained (n=100,601). The outermost layer represents the result using a minimal number of SNPs we tested (n=500). (C) AA genetic ancestry and global somatic copy number alterations by regression analysis. Volcano plots of  $-\log_{10}$  (p value) against effect size represent the influence of AA ancestry on SCNA scores across 10 cancer types. Each circle corresponds to a cancer type with size proportional to median burden of SCNA: weighted genomic instability index at overall level, weighted sum of SCNA events at focal level or arm/chromosomal level. Significance (y axis) and effect size (x axis) were calculated by linear regression with clinical factors as covariates. SCNA scores were rank-scaling transformed as a conservative measure to avoid results driven by outliers. Positive effect size corresponded to elevation of SCNA score by AA ancestry and negative values to reduction. The cancer types with elevated or reduced SCNA scores by AA ancestry (FDR < 10%) are shown in red or blue, respectively. Cancer types with non-significant results are shown in grey. (D) QQ plots compares the distribution of the observed p values ( $-\log$  scale) with an expected uniform distribution under the null (sloping line). In each plot, a dot represents the association of AA genetic ancestry with a recurrent focal SCNA or a recurrently mutated gene (at pan-cancer and cancer type-specific levels, respectively). Dots for recurrent focal SCNAs and recurrently mutated genes with significantly different rates of alteration in AA patients (compared with EA patients) are colored in red or blue (red, higher rates in AAs; blue, lower rates in AAs).

**Table S1, related to Figure 1. Summary of TCGA patients by self-identified race/ethnicity (SIRE) and genetic ancestry.**

Primary site	Cancer type	Abbreviations	Number	Self-identified race						Self-identified ethnicity			Genetic ancestry***				
				White	Black	Asian	AI/AN*	NH/OPI**	Unavailable	Hispanic	non-Hispanic	Unavailable	EA	AA	EAA	NA	Others
Bone Marrow	Acute Myeloid Leukemia	LAML	200	181	15	2	0	0	2	3	194	3	180	16	2	2	0
Adrenal Gland	Adrenocortical carcinoma	ACC	92	78	1	2	0	0	11	8	40	44	82	2	3	5	0
Bladder	Bladder Urothelial Carcinoma	BLCA	412	327	23	44	0	0	18	9	371	32	331	22	43	14	2
Brain	Brain Lower Grade Glioma	LGG	515	474	21	8	1	0	11	32	448	35	452	24	9	25	5
Breast	Breast invasive carcinoma	BRCA	1,098	757	183	61	1	0	96	39	884	175	825	183	57	25	8
Cervix	Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	304	209	30	19	8	2	36	24	168	112	193	33	21	57	0
Bile Duct	Cholangiocarcinoma	CHOL	36	31	2	3	0	0	0	2	33	1	30	2	2	2	0
Colon	Colon adenocarcinoma	COAD	459	214	58	11	1	0	175	4	270	185	386	59	12	2	0
Esophagus	Esophageal carcinoma	ESCA	185	114	5	46	0	0	20	6	88	91	125	6	45	8	1
Brain	Glioblastoma multiforme	GBM	599	504	51	13	0	0	31	13	489	97	512	51	7	19	10
Head and Neck	Head and Neck squamous cell carcinoma	HNSC	526	450	48	11	2	0	15	26	463	37	437	54	6	23	6
Kidney	Kidney Chromophobe	KICH	66	58	4	2	0	0	2	4	32	30	56	4	2	3	1
Kidney	Kidney renal clear cell carcinoma	KIRC	534	463	56	8	0	0	7	26	356	152	449	55	9	20	1
Kidney	Kidney renal papillary cell carcinoma	KIRP	290	206	61	6	2	0	15	12	242	36	208	63	6	12	1
Liver	Liver hepatocellular carcinoma	LIHC	377	187	17	161	2	0	10	18	340	19	178	18	163	17	1
Lung	Lung adenocarcinoma	LUAD	518	391	52	8	1	0	66	7	386	125	451	53	9	5	0
Lung	Lung squamous cell carcinoma	LUSC	504	351	31	9	0	0	113	8	319	177	459	32	11	1	1
Lymph Nodes	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	50	29	1	18	0	0	2	12	36	2	30	1	17	1	1
Pleura	Mesothelioma	MESO	87	85	1	1	0	0	0	0	73	14	84	0	0	2	1
Ovary	Ovarian serous cystadenocarcinoma	OV	592	485	34	20	3	1	49	10	331	251	515	38	17	13	9
Pancreas	Pancreatic adenocarcinoma	PAAD	185	162	7	11	0	0	5	5	137	43	161	10	11	3	0
Adrenal Gland	Pheochromocytoma and Paraganglioma	PCPG	179	148	20	6	1	0	4	5	138	36	144	22	3	6	4
Prostate	Prostate adenocarcinoma	PRAD	498	147	7	2	0	0	342	0	152	346	414	61	11	10	2
Rectal	Rectum adenocarcinoma	READ	167	81	6	1	0	0	79	1	82	84	158	6	1	2	0
Soft Tissue	Sarcoma	SARC	261	228	18	6	0	0	9	5	223	33	226	19	6	10	0
Skin	Skin Cutaneous Melanoma	SKCM	470	447	1	12	0	0	10	11	446	13	449	1	12	8	0
Stomach	Stomach adenocarcinoma	STAD	443	278	13	89	0	1	62	5	318	120	310	15	89	26	3
Testis	Testicular Germ Cell Tumors	TGCT	150	119	6	4	0	0	21	12	111	27	125	4	4	16	1
Thymus	Thymoma	THYM	124	103	6	13	0	0	2	10	100	14	98	8	12	5	1
Thyroid	Thyroid carcinoma	THCA	505	333	27	52	1	0	92	38	364	103	366	34	47	41	17
Uterine	Uterine Carcinosarcoma	UCS	57	44	9	3	0	0	1	1	43	13	44	9	3	1	0
Uterine	Uterine Corpus Endometrial Carcinoma	UCEC	559	374	108	20	4	9	44	15	376	168	393	114	37	13	2
Eye	Uveal Melanoma	UVM	80	55	0	0	0	0	25	1	52	27	80	0	0	0	0
<b>Total</b>			11,122	8,113	922	672	27	13	1,375	372	8,105	2,645	8,951	1,019	677	397	78

\*AI/AN: American Indian and Alaska Native

\*\*NH/OPI: Native Hawaiian and Other Pacific Islander

\*\*\*EA: European American; AA: African American; EAA: East Asian American; NA: Native American; OA: Others



**Table S2, related to Figure 1. Summary of the reference populations from the HapMap and HGDP projects.**

<b>Reference Population</b>	<b>Abbreviations</b>	<b>Database</b>	<b>Number of unrelated individuals</b>
<b>African ancestry in Southwest USA</b>	ASW	HapMap	52
<b>Utah residents with Northern and Western European ancestry from the CEPH collection</b>	CEU	HapMap	112
<b>Han Chinese in Beijing, China</b>	CHB	HapMap	137
<b>Chinese in Metropolitan Denver, Colorado</b>	CHD	HapMap	106
<b>Gujarati Indians in Houston, Texas</b>	GIH	HapMap	97
<b>Japanese in Tokyo, Japan</b>	JPT	HapMap	113
<b>Luhya in Webuye, Kenya</b>	LWK	HapMap	99
<b>Mexican ancestry in Los Angeles, California</b>	MXL	HapMap	54
<b>Maasai in Kinyawa, Kenya</b>	MKK	HapMap	105
<b>Toscani in Italia</b>	TSI	HapMap	102
<b>Yoruba in Ibadan, Nigeria</b>	YRI	HapMap	140
<b>Colombians</b>	HGDP	HGDP	7
<b>Karitiana</b>	HGDP	HGDP	14
<b>Maya</b>	HGDP	HGDP	21
<b>Pima</b>	HGDP	HGDP	14
<b>Surui</b>	HGDP	HGDP	8

**Table S3, related to Figure 3. Summary of genetic ancestry of TCGA patients based on primary sites and cancer types.**

Cancer Type	Abbreviations	Primary Site	Racial and Ethnic Groups							Genetic ancestry					
			Black	Asian	American Indian	Alaska Native	Native Hawaiian	Other Pacific Islander	Hispanic/Latino	EA	AA	EAA	NA	Others	
Acute Myeloid Leukemia	LAML	Bone Marrow									180	16	2	2	0
Adrenocortical carcinoma	ACC	Adrenal Gland									82	2	3	5	0
Bladder Urothelial Carcinoma	BLCA	Bladder									331	22	43	14	2
Brain Lower Grade Glioma	LGG	Brain									452	24	9	25	5
Breast invasive carcinoma	BRCA	Breast	H.M.								825	183	57	25	8
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	Cervix	H.M.						H.I.		193	33	21	57	0
Cholangiocarcinoma	CHOL	Bile Duct		H.I./H.M.							30	2	2	2	0
Colon adenocarcinoma	COAD	Colon	H.I./H.M.								386	59	12	2	0
Esophageal carcinoma	ESCA	Esophagus									125	6	45	8	1
Glioblastoma multiforme	GBM	Brain									512	51	7	19	10
Head and Neck squamous cell carcinoma	HNSC	Head and Neck	H.M.								437	54	6	23	6
Kidney Chromophobe	KICH										56	4	2	3	1
Kidney renal clear cell carcinoma	KIRC										449	55	9	20	1
Kidney renal papillary cell carcinoma	KIRP	Kidney	H.I.		H.M.		H.M.				208	63	6	12	1
Liver hepatocellular carcinoma	LIHC	Liver		H.I./H.M.					H.I.	H.I.	178	18	163	17	1
Lung adenocarcinoma	LUAD										451	53	9	5	0
Lung squamous cell carcinoma	LUSC	Lung	H.I./H.M.								459	32	11	1	1
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	Lymph Nodes									30	1	17	1	1
Mesothelioma	MESO	Pleura									84	0	0	2	1
Ovarian serous cystadenocarcinoma	OV	Ovary									515	38	17	13	9
Pancreatic adenocarcinoma	PAAD	Pancreas	H.I./H.M.								161	10	11	3	0
Pheochromocytoma and Paraganglioma	PCPG	Adrenal Gland									144	22	3	6	4
Prostate adenocarcinoma	PRAD	Prostate	H.I./H.M.								414	61	11	10	2
Rectum adenocarcinoma	READ	Rectal	H.I./H.M.								158	6	1	2	0
Sarcoma	SARC	Soft Tissue	H.I./H.M.								226	19	6	10	0
Skin Cutaneous Melanoma	SKCM	Skin									449	1	12	8	0
Stomach adenocarcinoma	STAD	Stomach	H.M.	H.I.							310	15	89	26	3
Testicular Germ Cell Tumors	TGCT	Testis							H.M.		125	4	4	16	1
Thymoma	THYM	Thymus									98	8	12	5	1
Thyroid carcinoma	THCA	Thyroid		H.M.							366	34	47	41	17
Uterine Carcinosarcoma	UCS										44	9	3	1	0
Uterine Corpus Endometrial Carcinoma	UCEC	Uterine	H.M.								393	114	37	13	2
Uveal Melanoma	UVM	Eye									80	0	0	0	0

**H.I. = Higher Incidence**

**H.M. = Higher Mortality**

**Table S4, related to Figure 3. Summary of cancer cell lines by primary site and genetic ancestry.**

Primary site	Sample number	Genetic ancestry				
		EA	AA	EAA	NA	Others
Adrenal gland	1	1	0	0	0	0
Autonomic ganglia	38	23	2	11	2	0
Biliary tract	6	1	0	5	0	0
Bone	55	42	2	8	3	0
Breast	65	46	11	4	2	2
Central nervous system	64	46	1	17	0	0
Cervix	11	8	1	2	0	0
Endometrium	26	9	0	17	0	0
Haematopoietic and lymphoid tissue	226	138	17	70	1	0
Kidney	41	27	2	12	0	0
Large intestine	53	43	2	6	1	1
Liver	19	1	3	13	2	0
Lung	222	166	22	31	3	0
Oesophagus	36	8	2	26	0	0
Ovary	60	36	3	20	1	0
Pancreas	44	27	1	16	0	0
Placenta	2	2	0	0	0	0
Pleura	25	21	3	1	0	0
Prostate	10	9	1	0	0	0
Salivary gland	3	1	0	2	0	0
Skin	86	79	0	7	0	0
Small intestine	1	0	1	0	0	0
Soft tissue	24	17	2	3	2	0
Stomach	37	2	2	31	2	0
Testis	3	1	0	2	0	0
Thyroid	15	10	0	5	0	0
Upper aerodigestive tract	49	33	1	14	0	1
Urinary tract	26	18	2	6	0	0
Vulva	3	3	0	0	0	0

**Table S5, related to Figure 5. List of recurrent focal SCNAs with significantly different alteration frequencies between AA and EA patients by pan-cancer meta-analysis.**

<b>Genomic Location</b>	<b>Cytoband</b>	<b>SCNA</b>	<b>Z score</b>	<b>p value</b>	<b>FDR</b>
chr12:114801336-117274075	12q24.21-12q24.22	Loss	3.200	1.376E-03	7.246E-02
chr16:87332621-90354753	16q24.2-16q24.3	Loss	-5.220	1.785E-07	2.820E-05
chr19:30211823-30453022	19q12	Gain	3.309	9.373E-04	7.246E-02

**Table S6, related to Figure 5. List of genes potentially contributing to disparity identified by pan-cancer meta-analysis.**

Gene Symbol	Accession	Gene Type	SCNA	SCNA Type	Expression (Upper Quartile)	Differential Expression			Expression Correlation With CN		
						Z score	p value	FDR	Correlation Coefficient	p value	FDR
<i>CCNE1</i>	ENSG00000105173	protein_coding	chr19: 30211823- 30453022	Amplification	7.704676357	5.021993784	5.11378E-07	4.18843E-06	0.494313599	1.77605E-19	6.70951E-19
<i>UR11</i>	ENSG00000105176	protein_coding	chr19: 30211823- 30453022	Amplification	18.86857091	2.721394124	0.006500721	0.016442999	0.734760829	1.2001E-108	1.3601E-107
<i>MED13L</i>	ENSG00000123066	protein_coding	chr12:114801336-117274075	Deletion	12.35400924	-6.73478932	1.64168E-11	3.00057E-10	0.374490726	1.051E-153	1.6243E-152
<i>C12orf49</i>	ENSG00000111412	protein_coding	chr12:114801336-117274075	Deletion	10.08157592	-3.282528399	0.001028806	0.003402974	0.316011068	8.90546E-63	7.96804E-62
<i>RP11-178L8.7</i>	ENSG00000270006	antisense	chr16: 87332621- 90354753	Deletion	1.188055195	6.960455116	3.39175E-12	1.45845E-10	0.352835137	1.87016E-61	1.58964E-60
<i>FBXO31</i>	ENSG00000103264	protein_coding	chr16: 87332621- 90354753	Deletion	5.451229183	2.128747435	0.033275162	0.072447188	0.515907356	2.21191E-89	2.35015E-88
<i>KLHDC4</i>	ENSG00000104731	protein_coding	chr16: 87332621- 90354753	Deletion	3.31894905	6.577066838	4.79819E-11	7.50263E-10	0.509914469	1.39267E-36	8.45548E-36
<i>BANP</i>	ENSG00000172530	protein_coding	chr16: 87332621- 90354753	Deletion	2.017602873	4.880555477	1.05787E-06	7.91106E-06	0.276979827	3.18943E-05	6.77754E-05
<i>ZFPM1</i>	ENSG00000179588	protein_coding	chr16: 87332621- 90354753	Deletion	1.426646968	4.168047718	3.0722E-05	0.000120095	0.241763989	0.006869537	0.010624476
<i>ZC3H18</i>	ENSG00000158545	protein_coding	chr16: 87332621- 90354753	Deletion	6.55897406	4.020107264	5.81716E-05	0.000217511	0.590765487	3.92514E-35	2.22424E-34
<i>CYBA</i>	ENSG00000051523	protein_coding	chr16: 87332621- 90354753	Deletion	55.33277649	4.675498216	2.93241E-06	1.68125E-05	0.306103285	9.77277E-27	4.61492E-26
<i>MVD</i>	ENSG00000167508	protein_coding	chr16: 87332621- 90354753	Deletion	8.332245193	6.725948805	1.74452E-11	3.00057E-10	0.48220506	4.91719E-33	2.61226E-32
<i>RNF166</i>	ENSG00000158717	protein_coding	chr16: 87332621- 90354753	Deletion	5.224803804	6.792837	1.0995E-11	2.47888E-10	0.266463998	0.000259989	0.000491091
<i>CTU2</i>	ENSG00000174177	protein_coding	chr16: 87332621- 90354753	Deletion	5.008474608	5.504419297	3.70387E-08	4.29243E-07	0.499546825	1.12829E-34	6.1874E-34
<i>CDT1</i>	ENSG00000167513	protein_coding	chr16: 87332621- 90354753	Deletion	6.578018314	4.753322083	2.00101E-06	1.32375E-05	0.208472881	0.006937158	0.010624476
<i>APRT</i>	ENSG00000198931	protein_coding	chr16: 87332621- 90354753	Deletion	67.29373777	2.540567294	0.011067279	0.026810874	0.503246149	0	0
<i>GALNS</i>	ENSG00000141012	protein_coding	chr16: 87332621- 90354753	Deletion	4.644139944	4.824766249	1.40168E-06	1.00453E-05	0.581512505	3.68996E-21	1.45882E-20
<i>ACSF3</i>	ENSG00000176715	protein_coding	chr16: 87332621- 90354753	Deletion	2.807687084	3.254320066	0.001136641	0.003620413	0.546780018	1.17029E-26	5.377E-26
<i>SPG7</i>	ENSG00000197912	protein_coding	chr16: 87332621- 90354753	Deletion	7.472638553	6.785985275	1.15297E-11	2.47888E-10	0.550033775	2.60473E-22	1.08001E-21
<i>RPL13</i>	ENSG00000167526	protein_coding	chr16: 87332621- 90354753	Deletion	173.5009749	4.720760297	2.34965E-06	1.49609E-05	0.323475147	2.21184E-39	1.56672E-38
<i>CHMP1A</i>	ENSG00000131165	protein_coding	chr16: 87332621- 90354753	Deletion	43.63303189	5.887622567	3.91791E-09	5.61567E-08	0.613930753	5.0302E-126	6.1081E-125
<i>SPATA33</i>	ENSG00000167523	protein_coding	chr16: 87332621- 90354753	Deletion	2.020762246	4.706304775	2.52247E-06	1.49609E-05	0.459968073	2.47117E-24	1.07718E-23
<i>CDK10</i>	ENSG00000185324	protein_coding	chr16: 87332621- 90354753	Deletion	9.85344778	7.327055761	2.35264E-13	1.34885E-11	0.375486545	3.07529E-18	1.08916E-17
<i>SPATA2L</i>	ENSG00000158792	protein_coding	chr16: 87332621- 90354753	Deletion	7.248753798	5.492382523	3.96547E-08	4.29243E-07	0.406840364	8.02062E-41	5.92828E-40
<i>VPS9D1</i>	ENSG00000075399	protein_coding	chr16: 87332621- 90354753	Deletion	7.85143401	6.801386314	1.03617E-11	2.47888E-10	0.497279634	2.00461E-29	1.0023E-28
<i>VPS9D1-AS1</i>	ENSG00000261373	antisense	chr16: 87332621- 90354753	Deletion	3.691291536	2.347809593	0.018884172	0.044494214	0.148718772	3.9492E-11	1.19886E-10
<i>ZNF276</i>	ENSG00000158805	protein_coding	chr16: 87332621- 90354753	Deletion	3.041370684	6.786878195	1.14586E-11	2.47888E-10	0.358089867	1.65126E-28	8.02039E-28
<i>SPIRE2</i>	ENSG00000204991	protein_coding	chr16: 87332621- 90354753	Deletion	2.565449299	4.712757916	2.44386E-06	1.49609E-05	0.260311128	1.31954E-59	1.0682E-58
<i>TCF25</i>	ENSG00000141002	protein_coding	chr16: 87332621- 90354753	Deletion	18.37152486	5.243169137	1.57842E-07	1.50827E-06	0.570198258	3.4199E-233	8.3054E-232
<i>MC1R</i>	ENSG00000258839	protein_coding	chr16: 87332621- 90354753	Deletion	1.262305911	3.899896344	9.62339E-05	0.000344838	0.292449286	1.0603E-156	1.8025E-155
<i>DEF8</i>	ENSG00000140995	protein_coding	chr16: 87332621- 90354753	Deletion	8.194492002	4.476476639	7.5885E-06	3.73597E-05	0.56813541	1.26543E-87	1.26543E-86
<i>AFG3L1P</i>	ENSG00000223959	unitary_pseudogene	chr16: 87332621- 90354753	Deletion	1.677837022	5.340435832	9.27234E-08	9.38143E-07	0.322808651	4.39212E-11	1.30993E-10
<i>DBNDD1</i>	ENSG00000003249	protein_coding	chr16: 87332621- 90354753	Deletion	12.7264118	2.801436825	0.00508756	0.013672818	0.426457025	1.57259E-10	4.5312E-10
<i>GAS8</i>	ENSG00000141013	protein_coding	chr16: 87332621- 90354753	Deletion	6.143787541	3.037299671	0.00238708	0.006730783	0.481141005	1.1066E-18	4.0026E-18

**Table S7, related to Figure 5. List of recurrent focal SCNAs with significantly different alteration frequencies between AA and EA patients by cancer type-specific analysis.**

Cancer Type	GISTIC Peak	Cytoband	SCNA Type	GISTIC Q value	Alteration Proportion (Copy Number Change>0.25)	z value	p value	FDR
BRCA	chr5:58155654-59787985	5q11.2-5q12.1	Deletion	1.1588E-12	0.205154639	3.78851229	0.00015155	0.0452383
BRCA	chr10:5013517-5039563	10p15.1	Amplification	0.000041502	0.213402062	3.42189295	0.00062187	0.07552452
BRCA	chr16:88525832-90354753	16q24.2-16q24.3	Deletion	5.7308E-08	0.570103093	-5.7640184	8.2134E-09	4.9034E-06
UCEC	chr17:29413917-29708443	17q11.2	Deletion	2.4684E-08	0.182952183	-3.4172663	0.00063253	0.07552452
BRCA	chr19:30072177-30511416	19q12	Amplification	2.8236E-08	0.183505155	3.47174712	0.00051708	0.07552452

**Table S8, related to Figure 5. List of genes potentially contributing to disparity for each cancer type by cancer type-specific analysis.**

Cancer Type	Gene Symbol	Accession	Gene Type	SCNA	SCNA Type	Expression (Upper Quartile)	Differential Expression			Expression Correlation With CN		
							Z score	p value	FDR	Correlation Coefficient	p value	FDR
BRCA	<i>ZC3H18</i>	ENSG00000158545	protein_coding	chr16:88525832-90354753	Deletion	5.82064841	7.429693033	1.0885E-13	8.76836E-13	0.699991088	3.2487E-159	1.267E-157
BRCA	<i>CYBA</i>	ENSG00000051523	protein_coding	chr16:88525832-90354753	Deletion	26.37238309	7.987129513	1.38117E-15	1.36966E-14	0.330587732	7.45911E-29	2.42421E-28
BRCA	<i>MVD</i>	ENSG00000167508	protein_coding	chr16:88525832-90354753	Deletion	6.82803507	9.912385445	3.67759E-23	1.45878E-21	0.563221736	4.40569E-91	4.29555E-90
BRCA	<i>RNF166</i>	ENSG00000158717	protein_coding	chr16:88525832-90354753	Deletion	4.012862075	9.188798807	3.97252E-20	6.75329E-19	0.429461076	1.60097E-49	7.20438E-49
BRCA	<i>CTU2</i>	ENSG00000174177	protein_coding	chr16:88525832-90354753	Deletion	4.192217239	9.960287392	2.27404E-23	1.35306E-21	0.574611418	1.4664E-95	1.55971E-94
BRCA	<i>PIEZO1</i>	ENSG00000103335	protein_coding	chr16:88525832-90354753	Deletion	15.16202238	5.019176703	5.18934E-07	1.62508E-06	0.532687152	6.71022E-80	5.23397E-79
BRCA	<i>RP5-1142A6.9</i>	ENSG00000260121	antisense	chr16:88525832-90354753	Deletion	1.271929141	2.627215512	0.008608679	0.01552171	0.294912816	4.92112E-23	1.339E-22
BRCA	<i>CDT1</i>	ENSG00000167513	protein_coding	chr16:88525832-90354753	Deletion	7.1527411	9.108717101	8.33618E-20	1.24001E-18	0.395103291	1.5992E-41	6.45196E-41
BRCA	<i>APRT</i>	ENSG00000198931	protein_coding	chr16:88525832-90354753	Deletion	56.96790937	7.101736663	1.23199E-12	8.62392E-12	0.498919172	8.52079E-69	5.53851E-68
BRCA	<i>GALNS</i>	ENSG00000141012	protein_coding	chr16:88525832-90354753	Deletion	3.754837571	6.596170349	4.21914E-11	2.18295E-10	0.623325366	7.4598E-117	1.7456E-115
BRCA	<i>TRAPPC2L</i>	ENSG00000167515	protein_coding	chr16:88525832-90354753	Deletion	7.919829903	5.340493493	9.26939E-08	3.3426E-07	0.515643128	3.86634E-74	2.82726E-73
BRCA	<i>ACSF3</i>	ENSG00000176715	protein_coding	chr16:88525832-90354753	Deletion	2.341781298	5.677066623	1.37024E-08	5.59942E-08	0.465061061	7.60787E-59	4.4506E-58
BRCA	<i>RP11-46C24.7</i>	ENSG00000259877	antisense	chr16:88525832-90354753	Deletion	1.611312005	2.433918896	0.014936341	0.026528726	0.398084088	3.51852E-42	1.47024E-41
BRCA	<i>SPG7</i>	ENSG00000197912	protein_coding	chr16:88525832-90354753	Deletion	5.730002396	9.427134926	4.21443E-21	1.00303E-19	0.544083428	6.12864E-84	5.12179E-83
BRCA	<i>RPL13</i>	ENSG00000167526	protein_coding	chr16:88525832-90354753	Deletion	151.954919	7.752485849	9.01109E-15	8.24861E-14	0.33060672	7.40245E-29	2.42421E-28
BRCA	<i>CHMP1A</i>	ENSG00000131165	protein_coding	chr16:88525832-90354753	Deletion	38.89219858	9.35574438	8.30131E-21	1.64643E-19	0.61569016	2.8746E-113	5.6054E-112
BRCA	<i>SPATA33</i>	ENSG00000167523	protein_coding	chr16:88525832-90354753	Deletion	1.933164554	6.908840571	4.88631E-12	2.90735E-11	0.432916797	2.22426E-50	1.08433E-49
BRCA	<i>CDK10</i>	ENSG00000185324	protein_coding	chr16:88525832-90354753	Deletion	6.359586038	10.24827799	1.20471E-24	1.4336E-22	0.338104681	3.49342E-30	1.23857E-29
BRCA	<i>SPATA2L</i>	ENSG00000158792	protein_coding	chr16:88525832-90354753	Deletion	6.314829763	8.923518819	4.51703E-19	5.97252E-18	0.34507552	1.89293E-31	6.92103E-31
BRCA	<i>VPS9D1</i>	ENSG00000075399	protein_coding	chr16:88525832-90354753	Deletion	5.03332479	9.82329055	8.93775E-23	2.65898E-21	0.480905037	2.30992E-63	1.42243E-62
BRCA	<i>VPS9D1-AS1</i>	ENSG00000261373	antisense	chr16:88525832-90354753	Deletion	4.185458521	4.005374865	6.19192E-05	0.000150375	0.20334725	1.662E-11	4.05113E-11
BRCA	<i>ZNF276</i>	ENSG00000158805	protein_coding	chr16:88525832-90354753	Deletion	2.511320339	8.469160135	2.47169E-17	2.94131E-16	0.432727382	2.47987E-50	1.16058E-49
BRCA	<i>FANCA</i>	ENSG00000187741	protein_coding	chr16:88525832-90354753	Deletion	1.620229105	4.756543888	1.96935E-06	5.32621E-06	0.502404014	6.94147E-70	4.77736E-69
BRCA	<i>SPIRE2</i>	ENSG00000204991	protein_coding	chr16:88525832-90354753	Deletion	1.954862952	7.427672073	1.10526E-13	8.76836E-13	0.299454013	9.88766E-24	2.75442E-23
BRCA	<i>TCF25</i>	ENSG00000141002	protein_coding	chr16:88525832-90354753	Deletion	16.43701672	8.40873748	4.14448E-17	4.48357E-16	0.590283709	5.1273E-102	6.6655E-101
BRCA	<i>TUBB3</i>	ENSG00000258947	protein_coding	chr16:88525832-90354753	Deletion	1.035565309	2.955203436	0.003124629	0.006302219	0.374852885	3.13397E-37	1.22225E-36
BRCA	<i>DEF8</i>	ENSG00000140995	protein_coding	chr16:88525832-90354753	Deletion	6.139670482	7.151703884	8.57073E-13	6.37448E-12	0.614104992	1.5514E-112	2.5931E-111
BRCA	<i>AFG3L1P</i>	ENSG00000223959	unitary_pseudogene	chr16:88525832-90354753	Deletion	1.506763831	6.436802598	1.22016E-10	5.5846E-10	0.441502527	1.49285E-52	7.59409E-52
BRCA	<i>DBNDD1</i>	ENSG00000003249	protein_coding	chr16:88525832-90354753	Deletion	12.39147979	5.545301602	2.93448E-08	1.12646E-07	0.45245801	2.037E-55	1.1349E-54
BRCA	<i>GAS8</i>	ENSG00000141013	protein_coding	chr16:88525832-90354753	Deletion	5.051450699	4.087786927	4.35508E-05	0.00010797	0.400327764	1.11421E-42	4.82824E-42
BRCA	<i>POP4</i>	ENSG00000105171	protein_coding	chr19:30072177-30511416	Amplification	5.785090204	5.983815875	2.1797E-09	9.26371E-09	0.805300781	4.3237E-246	5.0588E-244
BRCA	<i>PLEKHF1</i>	ENSG00000166289	protein_coding	chr19:30072177-30511416	Amplification	4.796813308	6.562266927	5.29958E-11	2.62771E-10	0.448994741	1.68276E-54	8.94923E-54
BRCA	<i>C19orf12</i>	ENSG00000131943	protein_coding	chr19:30072177-30511416	Amplification	5.370199636	5.224610951	1.74522E-07	5.76891E-07	0.694559872	8.9758E-156	2.6254E-154
BRCA	<i>CCNE1</i>	ENSG00000105173	protein_coding	chr19:30072177-30511416	Amplification	2.993048162	6.932394985	4.13775E-12	2.59154E-11	0.56069577	4.10816E-90	3.69734E-89
UCEC	<i>RP11-848P1.5</i>	ENSG00000264107	antisense	chr17:29413917-29708443	Deletion	1.79618157	2.415647668	0.015707252	0.027487691	0.261538601	8.12383E-10	1.90098E-09

**Table S9, related to Figure 6. Difference in mutational signatures between AA and EA patients.**

Cancer Type	Mutational signature*	Effect size	Effect size Std. Error	Adjusted regression p value	FDR corrected q values
<b>BRCA</b>	Signature.age	-0.1	0.1	4.1E-01	9.3E-01
	Signature.APOBEC	0.0	0.1	9.1E-01	9.6E-01
	Signature.3	0.2	0.1	2.5E-02	5.0E-01
	Signature.8	-0.1	0.1	5.5E-01	9.3E-01
<b>COAD</b>	Signature.age	-0.1	0.2	6.3E-01	9.3E-01
	Signature.6	0.1	0.2	7.0E-01	9.3E-01
<b>HNSC</b>	Signature.age	0.3	0.2	7.5E-02	6.1E-01
	Signature.APOBEC	-0.2	0.2	2.3E-01	9.1E-01
	Signature.4	0.1	0.2	6.5E-01	9.3E-01
<b>LUAD</b>	Signature.age	-0.1	0.2	7.4E-01	9.3E-01
	Signature.APOBEC	-0.2	0.2	2.4E-01	9.1E-01
	Signature.4	0.3	0.2	9.2E-02	6.1E-01
<b>LUSC</b>	Signature.age	0.1	0.2	4.6E-01	9.3E-01
	Signature.APOBEC	0.0	0.2	8.3E-01	9.6E-01
	Signature.4	-0.2	0.2	3.5E-01	9.3E-01
<b>PRAD</b>	Signature.age	-0.2	0.2	2.7E-01	9.1E-01
	Signature.6	0.1	0.2	5.0E-01	9.3E-01
<b>UCEC</b>	Signature.age	0.0	0.1	9.2E-01	9.6E-01
	Signature.6	0.0	0.1	7.0E-01	9.3E-01
	Signature.10	0.0	0.1	9.7E-01	9.7E-01

\*Mutational signatures with dominant (more than 95%) or insufficient (less than 5%) contribution to a given cancer type were excluded for analysis. Mutational signatures were designated by COSMIC (<https://cancer.sanger.ac.uk/cosmic/signatures>).



**Table S11, related to Figure 6. List of recurrently mutated genes with significantly different alteration frequencies between AA and EA patients by cancer type-specific analysis.**

Cancer Type	Gene Symbol	Alteration Proportion	Differential Altered		
			z value	p value	FDR
BRCA	<i>CDH1</i>	0.135538954	-2.973072837	0.002948345	0.090809019
BRCA	<i>PIK3CA</i>	0.332977588	-3.598450362	0.000320119	0.024649158
BRCA	<i>TP53</i>	0.300960512	3.732272808	0.00018976	0.024649158
COAD	<i>SOX9</i>	0.136138614	2.655147529	0.007927373	0.203469241
UCEC	<i>PIK3R1</i>	0.294238683	-2.994159226	0.002752023	0.090809019
UCEC	<i>PTEN</i>	0.565843621	-3.240369875	0.001193747	0.061279035