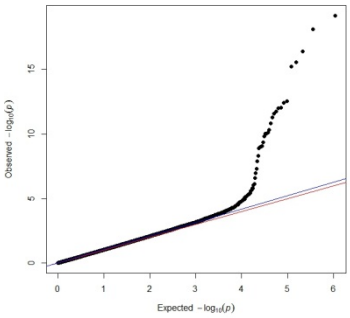
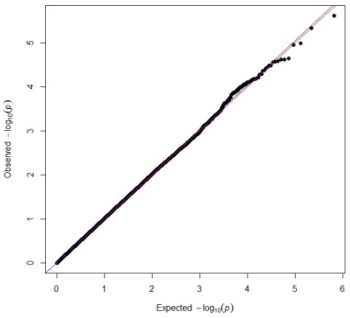
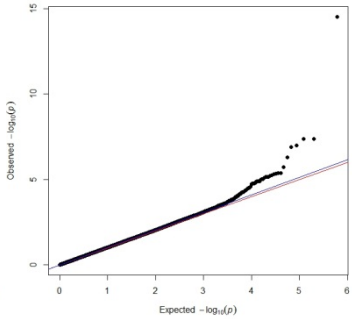


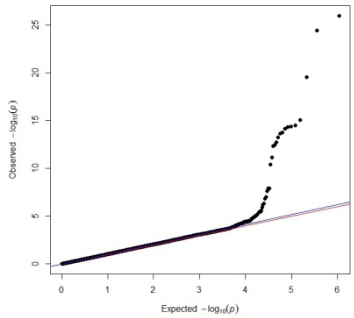
(a) ICR-GWAS (λ=1.05)



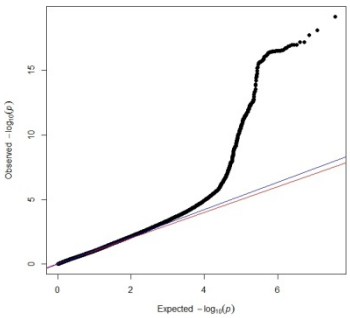
(b) MDACC-GWAS (λ=1.01)



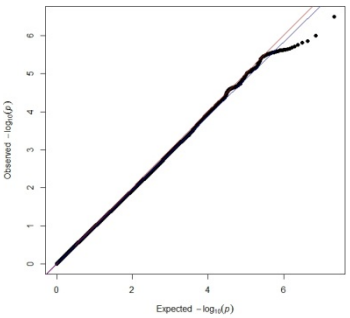
(c) IARC-GWAS (λ=1.02)



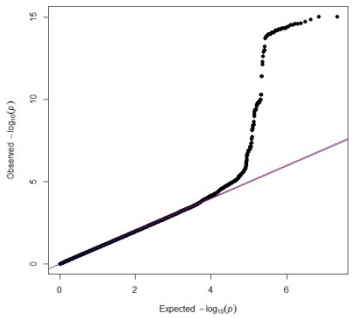
(d) NCI-GWAS (λ=1.04)



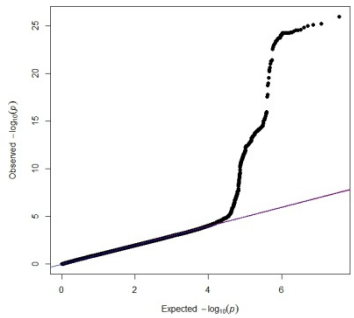
(e) ICR-GWAS post-imputation (λ=1.06)



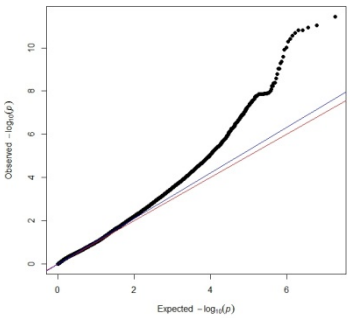
(f) MDACC-GWAS post-imputation (λ=0.99)



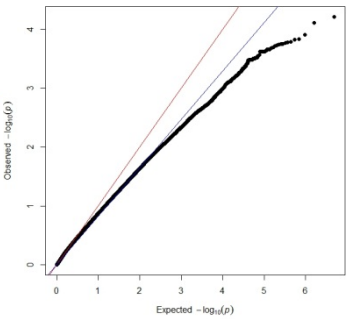
(g) IARC-GWAS post-imputation (λ=1.00)



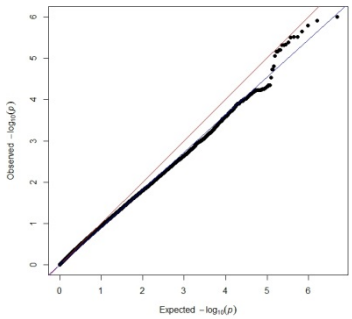
(h) NCI-GWAS post-imputation (λ=0.99)



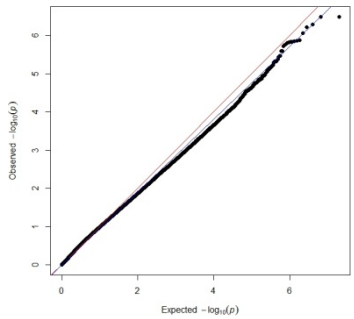
(i) ICR-GWAS post-imputation rare SNPs (λ=1.05)



(j) MDACC-GWAS post-imputation rare SNPs (λ=0.82)



(k) IARC-GWAS post-imputation rare SNPs (λ=0.91)



(l) NCI-GWAS post-imputation rare SNPs (λ=0.95)

Supplementary Figure 1: Q-Q plots of Cochran-Armitage trend test statistics for association based on 11,348 cases and 15,861 controls from discovery phase GWASs pre-imputation (a-d); all SNPs post-imputation (e-h) and rare SNPs post-imputation (i-l).

Nature Genetics: doi:10.1038/ng.3002

Supplementary Table 1: (a) Study participants

Subgroups		Discovery phase								Replication phase												Overall	
		IARC		MDACC		NCI		ICR		deCODE		Harvard		Heidelberg-EPIC		IARC		Toronto		ICR			
		Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
ALL		2,533	3,791	1,150	1,134	5,713	5,736	1,952	5,200	1,319	26,380	984	970	1,362	1,253	3,049	5,737	1,084	966	2,448	2,989	21,594	54,156
Gender	Male	1,900	2,549	655	644	4,464	4,837	1,166	n/a	638	11,511	507	442	1,030	956	2,128	3,302	546	487	1,664	1,469	14,698	26,197
	Female	633	1,242	495	490	1,249	899	786	n/a	681	14,869	477	528	332	297	919	2,415	534	479	784	1,520	6,890	22,739
Smoking	Never	159	1,253	n/a	n/a	350	1,379	109	5,200	55	1,100	92	161	114	474	364	1,861	124	468	129	555	1,496	12,451
	Former	449	891	572	626	1,927	2,001	762	5,200	n/a	n/a	502	555	475	544	482	1,084	495	414	1,724	587	7,388	11,902
	Current	1,814	1,513	578	508	3,415	2,335	1,065	5,200	n/a	n/a	390	254	697	235	1,378	1,678	426	68	537	343	10,300	12,134
	Ever	2,367	2,508	1,150	1,134	5,347	4,349	1,843	5,200	1,018	20,360	892	809	1,172	779	748	n/a	947	492	2,261	930	17,745	36,561
Age	≤ 50	319	1,051	193	137	101	124	n/a	n/a	71	1,934	81	214	25	49	1,209	2,619	77	165	35	473	2,111	6,766
	> 50	2,214	2,737	957	997	5,612	5,612	n/a	n/a	1,245	24,446	903	756	1,337	1,204	1,836	3,098	1,004	801	2,413	2,516	17,521	42,167
Histology	AD	517	2,824	619	1,134	1,841	5,736	465	5,200	547	26,380	488	970	554	1,253	919	5,737	649	966	432	2,989	7,031	53,189
	SQ	911	2,968	306	1,134	1,447	5,736	611	5,200	259	26,380	215	970	345	1,253	1,121	5,737	206	966	1,056	2,989	6,477	53,333
	SCLC	388	2,968	n/a	n/a	706	5,736	530	5,200	193	26,380	n/a	n/a	274	1,253	299	5,737	77	966	506	2,989	2,973	51,229
	LCLC	40	2,968	n/a	n/a	170	5,736	n/a	n/a	42	26,380	81	970	42	1,253	96	5,737	37	966	0	2,989	508	46,999
	NSCLC	n/a	n/a	n/a	n/a	n/a	n/a	1,420	5,200	n/a	n/a	n/a	n/a	1,012	1,253	n/a	n/a	76	966	1,909	2,989	4,417	10,408
	other	677	2,127	224	1,134	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	34	1,253	592	5,737	36	966	33	2,989	1,596	14,206
Stage	I/II[1]	n/a	n/a	378	1,134	1,424	5,736	92	5,200	n/a	n/a	455	970	153	1,253	617	4,665	n/a	n/a	n/a	n/a	3,119	18,958
	III/IV[1]	n/a	n/a	707	1,134	2,064	5,736	320	5,200	n/a	n/a	529	970	452	1,253	706	4,665	n/a	n/a	n/a	n/a	4,778	18,958
	limited stage	n/a	n/a	n/a	n/a	n/a	n/a	56	5,200	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	56	5,200
	extended	n/a	n/a	n/a	n/a	n/a	n/a	104	5,200	n/a	n/a	n/a	n/a	n/a	n/a	0	n/a	n/a	n/a	n/a	n/a	104	5,200
Family history	Yes	130	93	241	167	782	554	285	n/a	n/a	n/a	166	110	n/a	n/a	92	40	194	114	n/a	n/a	1,890	1,078
	No	2,403	3,668	904	964	4,251	4,724	1,667	n/a	n/a	n/a	817	860	n/a	n/a	2,079	2,079	890	852	n/a	n/a	13,011	13,147

Supplementary Table 1: (b) Quality control of GWAS datasets

Quality Control exclusion criteria	ICR-GWAS			MDACC-GWAS		
QC per individual						
	criteria	Cases	Controls	criteria	Cases	Controls
samples attempted		1,978	5,200		1177	1176
failed genotyping		20	0		0	0
Missing rate per person	>5%	0	0	>5%	5	4
X chromosome heterozygosity rate	>20% for men and < 20% for women	6	0	>20% for men and < 20% for women	3	6
Unexpected duplicates and first degree relatives	duplicate ID, genome-wide IBD proportion > 0.185	0	0	duplicate ID, genome-wide IBD proportion > 0.185	4	13
Autosomal heterozygosity rate	>6 SD of the mean	0	0	>6 SD of the mean	4	3
None European ancestry	Non CEU	0	0	Non CEU	5	7
other		0	0		6	9
samples post QC		**1,952**	**5,200**		**1,150**	**1,134**
SNP QC						
genotyping array		Illumina 550k	Illumina 1.2M		Illumina 300K HumanHap v1.1	
SNPs attempted		552,947	888,268			
SNPs attempted in both cases and controls		499,432			317,498	
Test of missingness by case-control status		0			n/a	
Genotyped call rate	<0.95	47		<0.95	1638	
Monomorphic in CEU		0			410	
MAF	<0.01	0		<0.01	584	
HWE	$<1x10^{-4}$	642		$<1x10^{-4}$	639	
Total number of SNPs post QC		**498,744**			**314,227**	

Quality Control exclusion criteria		IARC-GWAS									
		Central Europe		CARET[2]		Estonia		France		Hunt/Tromso[3]	
QC per individual											
	criteria	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
samples attempted		1,841	2,441	397	393	109	875	135	146	403	412
failed genotyping		0	0	0	0	0	0	0	0	0	0
Missing rate per person	>5%	224	278	32	36	2	1	5	0	33	24
X chromosome heterozygosity rate	>10% for men and < 20% for women	2	3	2	1	0	1	1	0	4	3
Unexpected duplicates and first degree relatives	duplicate ID, genome-wide IBD proportion > 0.2	3	7	0	1	1	46	0	2	13	15
Autosomal heterozygosity rate	>6 SD of the mean	3	7	1	0	0	3	1	0	0	0
None European ancestry	Ancestry probability being CEU <80%	3	4	0	0	0	1	0	0	0	3
other		10	15	0	0	0	0	1	0	11	24
samples post QC		**1,596**	**2,127**	**362**	**354**	**106**	**823**	**127**	**144**	**342**	**343**
SNP QC											
genotyping array		Illumina 317k	Illumina 317k	Illumina 317k	Illumina 317k	Illumina 317k	Illumina 370Duo	Illumina 317k	Illumina 317k	Illumina 317k	Illumina 370k
SNPs attempted		318,237	318,237	318,237	318,237	318,237	370,404	318,237	318,237	318,237	318,237
SNPs attempted in both cases and controls		311,129		311,129		311,129		311,129		311,129	
Test of missingness by case-control status		25		0		73		624		1,879	
Genotyped call rate	<0.95	11,986		12,903		11806		19,480		16,556	
Strand issue		21		21		21		21		21	
MAF	<0.01	n/a		n/a		n/a		n/a		n/a	
HWE	$<1x10^{-4}$	342		14		114		503		30	
Total number of SNPs post QC		**298,776**		**289,938**		**299,136**		**290,727**		**292,664**	

Quality Control exclusion criteria	NCI-GWAS		
Quality controls per individual			
	criteria	Cases	Controls
samples attempted[4]		6,045	6,542
failed genotyping		0	0
Missing rate per person	>5%	188	288
X chromosome heterozygosity rate	>0.5% for men and < 20% or >35% for women[5]	32	24
Unexpected duplicates and first degree relatives	genotype concordance > 98%	51	38
Autosomal heterozygosity rate	>34%	2	0
None European ancestry	Estimated CEU admixture <80%	16	2
other		52	456
samples post QC		**5,713**	**5,736**
Quality controls per SNP			
genotyping array		Illumina 550k/610k	Illumina 550k/610k /317k+240S/1M
SNPs attempted		561,466/620,901	561,466/620,901 /561,494/1,192,887
SNPs attempted in both cases and controls		599,947	
Test of missingness by case-control status		0	
Genotyped call rate	<0.95	79,420	
MAF (including monomorphic)	<0.001	19,437	
HWE	$<1x10^{-6}$	15,930	
other		183	
SNPs post QC		**505,770**	

Nature Genetics: doi:10.1038/ng.3002

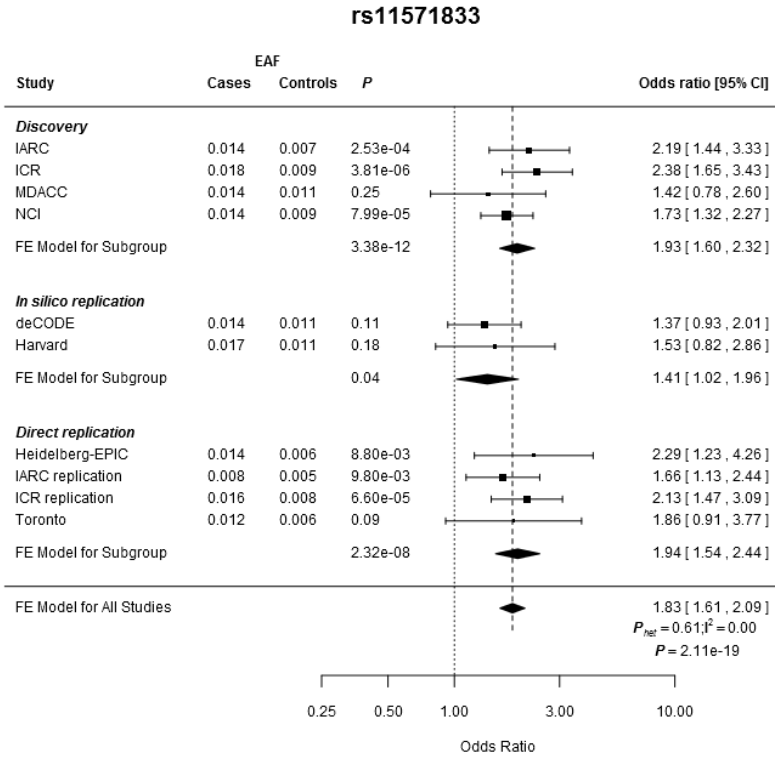
Supplementary Table 1: (c) details of imputation applied to each GWAS dataset.

		IARC	MDACC	NCI	UK
Imputation software		minimac	MaCH1.0	IMPUTE II	IMPUTE II
Reference panel		1000Genome (Phase 1 integrated release version 3)	1000Genome (Phase 1 integrated release version 3)	1000Genome (Phase 1 integrated release version 3)	1000Genome (Phase 1 integrated release version 3)
Program for association analysis		mach2dat	ProbABEL	R (glm function)	SNPTEST
Size of inference panel		289,938 - 299,136¥	314,227	505,770	498,744
Number of SNPs imputed*		10,728,367	10,751,467	18,330,772	18,378,912
	<1%	2,340,682 (21.8%)	2,357,454 (21.9%)	9,665,685 (52.7%)	9,121,221 (49.6%)
	1-5%	2,173,927 (20.3%)	2,165,097 (20.1%)	2,483,010 (13.6%)	2,761,148 (15.0%)
	>5%	6,213,736 (57.9%)	6,228,916 (58.0%)	6,182,077 (33.7%)	6,496,543 (35.4%)

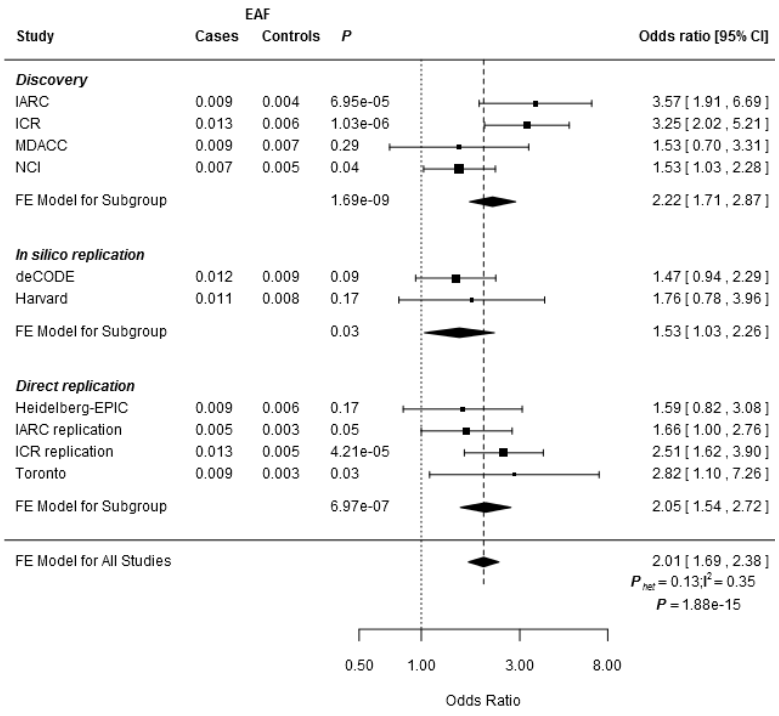
[1]Stage for NSCLC
[2]Carotene and Retinol Efficacy Trial cohort.
[3]North Trondelag Health Study 2 / Tromsø IV
[4]This includes 241 duplicate samples
[5]With exceptions of ATBC study in which the exclusion criterion is >1.2% for men and CPSII controls genotyped on 1M array for which exclusion criteria of either <18% or >26% is used for women
* Numbers presented are those passed the post-imputation quality controls; ¥ Numbers based on 5 series

Supplementary Table 3: Tables of the odds ratios of lung cancer associated with 13q13.1 (rs11571833 and rs56084662), 22q12.1 (rs17879961) and 3q28 (rs13314271) risk loci (a-l). All lung cancer based on 21,594 lung cancer cases and 54,156 controls (a-d), SQ based on 6,477 SQ and 53,333 controls (e-h) and AD based on 7,031 AD and 53,189 controls (i-l). Studies are weighted according to the inverse of the variance of the log of the OR calculated by unconditional logistic regression. *Horizontal lines:* 95% confidence intervals (*95% CI*). *Box:* OR point estimate; its area is proportional to the weight of the study. *Diamond (and broken line):* overall summary estimate, with confidence interval given by its width. *Unbroken vertical line:* at the null value (OR = 1.0).

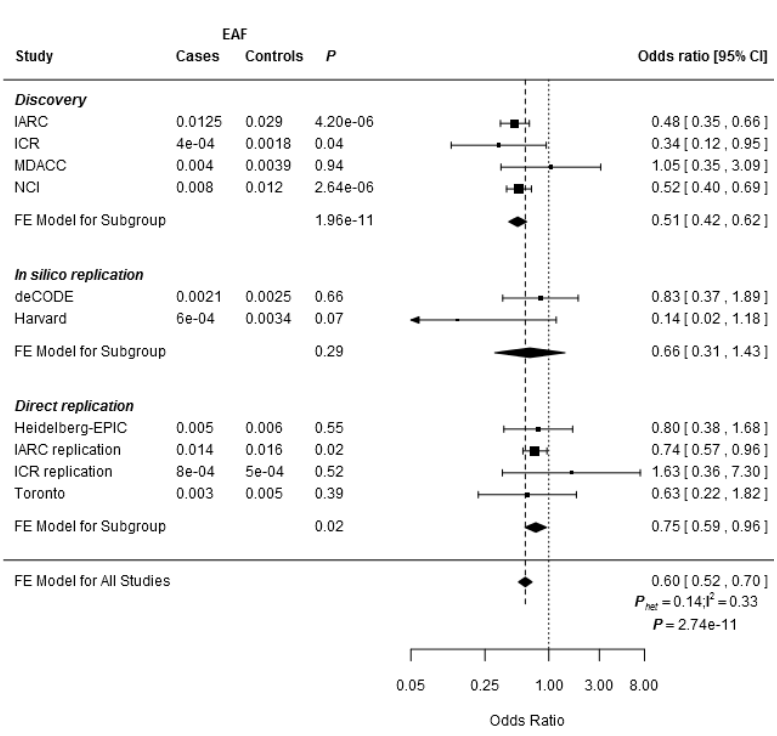
(a)



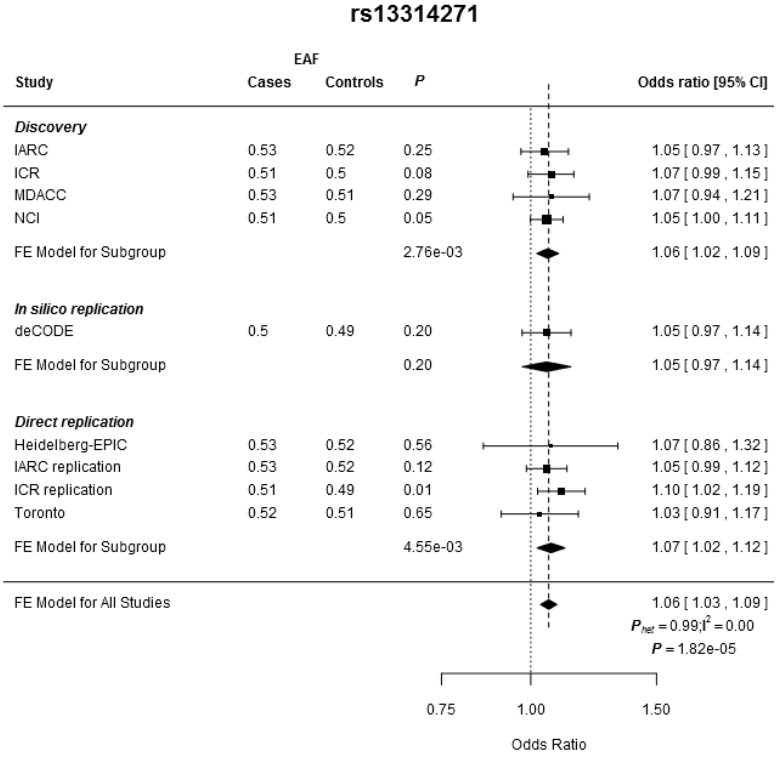
(b)



(c)

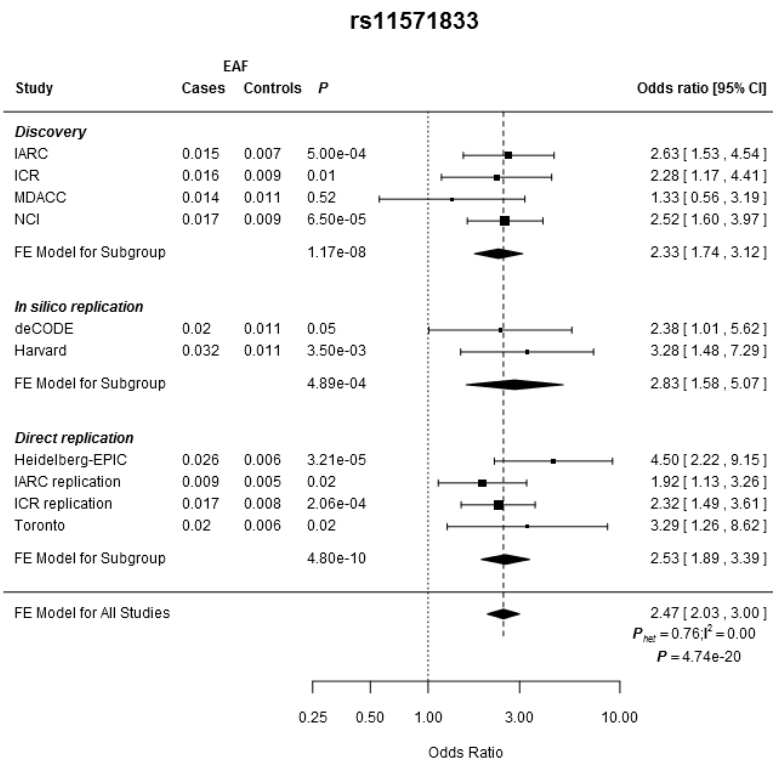


(d)

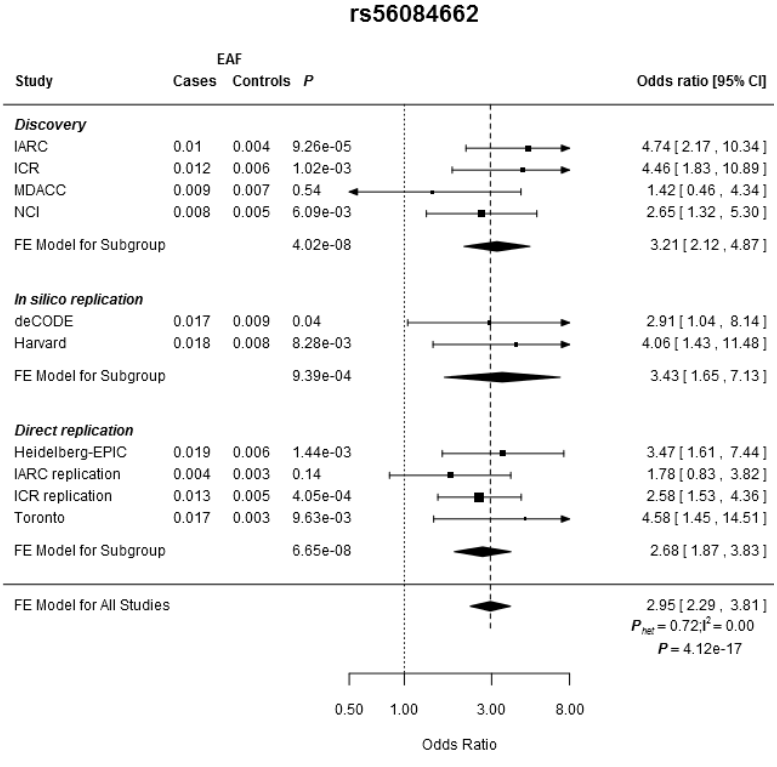


Nature Genetics: doi:10.1038/ng.3002

(e)

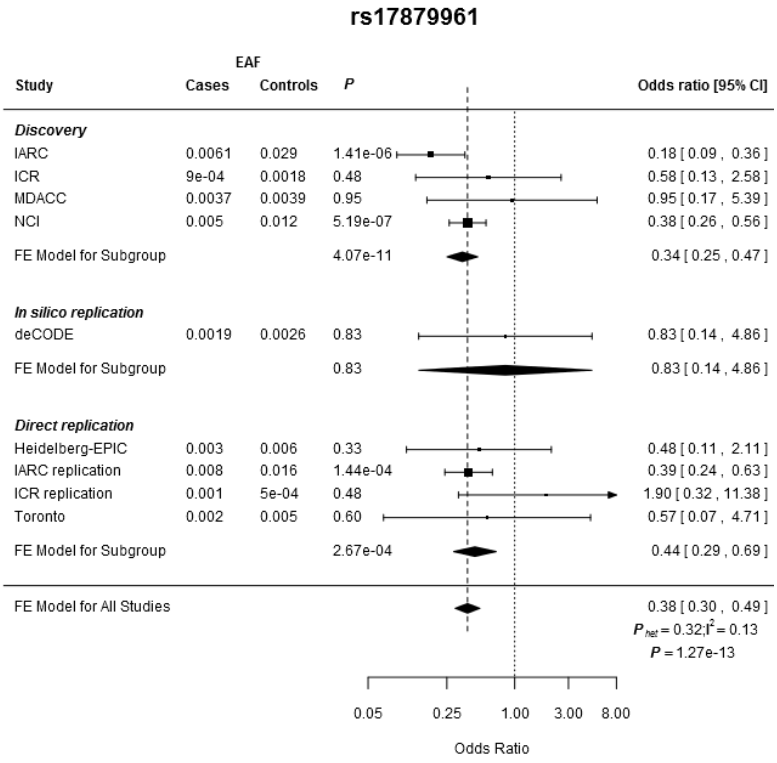


(f)

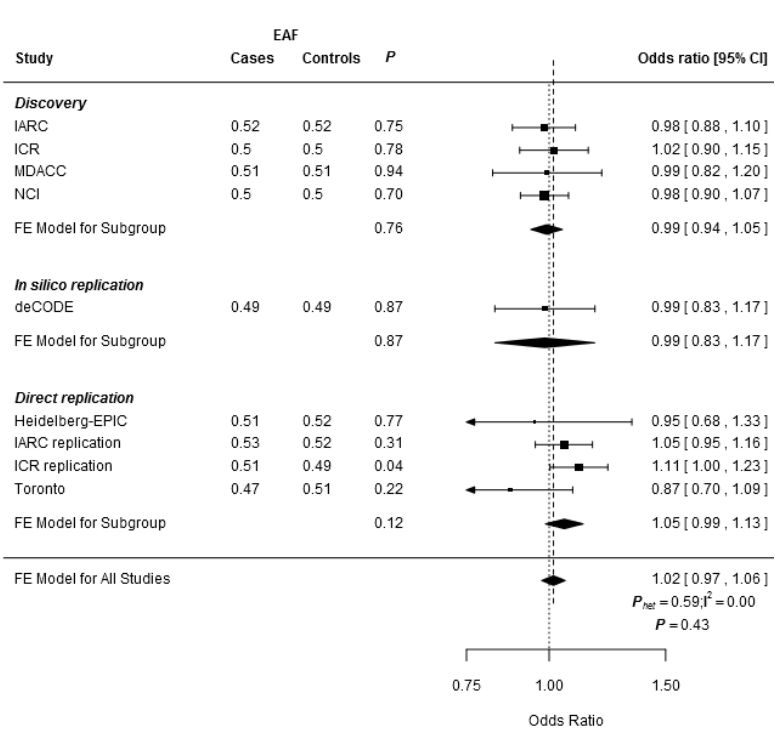


Nature Genetics: doi:10.1038/ng.3002

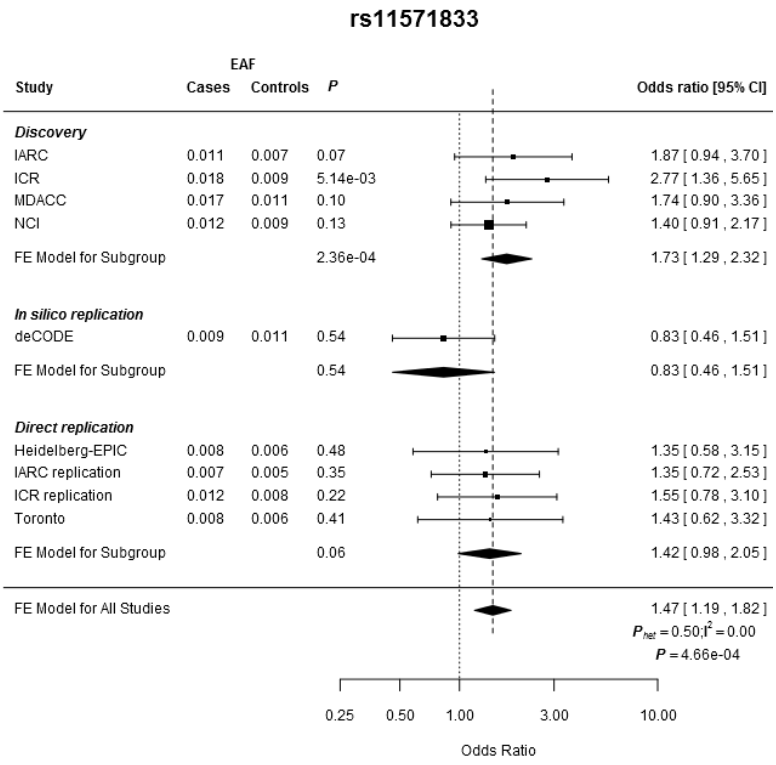
(g)


rs17879961
EAF
Study
Cases
Controls
P
Odds ratio [95% CI]
Discovery
IARC 0.0061 0.029 1.41e-06 0.18 [0.09 , 0.36]
ICR 9e-04 0.0018 0.48 0.58 [0.13 , 2.58]
MDACC 0.0037 0.0039 0.95 0.95 [0.17 , 5.39]
NCI 0.005 0.012 5.19e-07 0.38 [0.26 , 0.56]
FE Model for Subgroup 4.07e-11 0.34 [0.25 , 0.47]
In silico replication
deCODE 0.0019 0.0026 0.83 0.83 [0.14 , 4.86]
FE Model for Subgroup 0.83 0.83 [0.14 , 4.86]
Direct replication
Heidelberg-EPIC 0.003 0.006 0.33 0.48 [0.11 , 2.11]
IARC replication 0.008 0.016 1.44e-04 0.39 [0.24 , 0.63]
ICR replication 0.001 5e-04 0.48 1.90 [0.32 , 11.38]
Toronto 0.002 0.005 0.60 0.57 [0.07 , 4.71]
FE Model for Subgroup 2.67e-04 0.44 [0.29 , 0.69]
FE Model for All Studies 0.38 [0.30 , 0.49]
$P_{het} = 0.32; I^2 = 0.13$
$P = 1.27e-13$
0.05
0.25
1.00
3.00
8.00
Odds Ratio

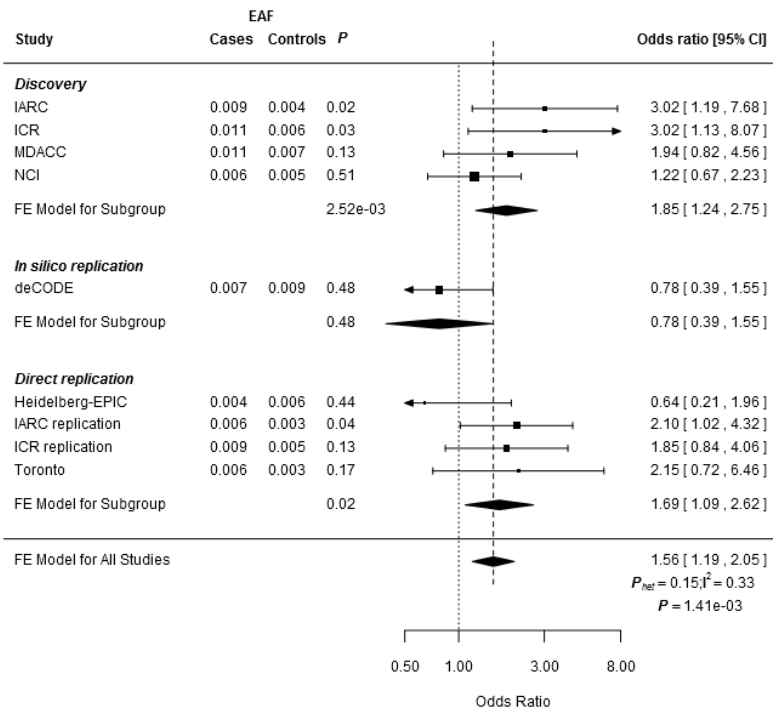
(h)


rs13314271
EAF
Study
Cases
Controls
P
Odds ratio [95% CI]
Discovery
IARC 0.52 0.52 0.75 0.98 [0.88 , 1.10]
ICR 0.5 0.5 0.78 1.02 [0.90 , 1.15]
MDACC 0.51 0.51 0.94 0.99 [0.82 , 1.20]
NCI 0.5 0.5 0.70 0.98 [0.90 , 1.07]
FE Model for Subgroup 0.76 0.99 [0.94 , 1.05]
In silico replication
deCODE 0.49 0.49 0.87 0.99 [0.83 , 1.17]
FE Model for Subgroup 0.87 0.99 [0.83 , 1.17]
Direct replication
Heidelberg-EPIC 0.51 0.52 0.77 0.95 [0.68 , 1.33]
IARC replication 0.53 0.52 0.31 1.05 [0.95 , 1.16]
ICR replication 0.51 0.49 0.04 1.11 [1.00 , 1.23]
Toronto 0.47 0.51 0.22 0.87 [0.70 , 1.09]
FE Model for Subgroup 0.12 1.05 [0.99 , 1.13]
FE Model for All Studies 1.02 [0.97 , 1.06]
$P_{het} = 0.59; I^2 = 0.00$
$P = 0.43$
0.75
1.00
1.50
Odds Ratio

(i)

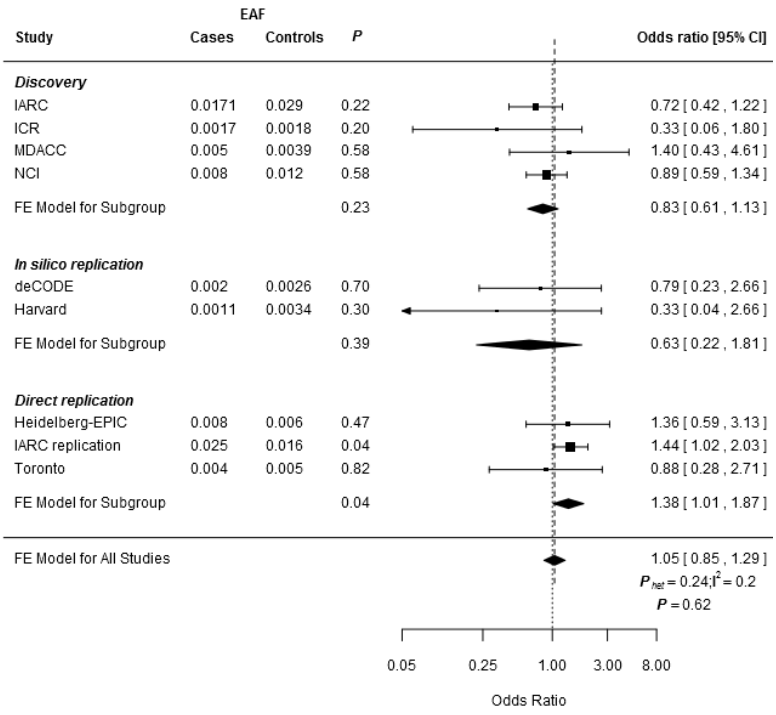

rs11571833
Study
EAF
Cases
Controls
P
Odds ratio [95% CI]
Discovery
IARC 0.011 0.007 0.07 1.87 [0.94 , 3.70]
ICR 0.018 0.009 5.14e-03 2.77 [1.36 , 5.65]
MDACC 0.017 0.011 0.10 1.74 [0.90 , 3.36]
NCI 0.012 0.009 0.13 1.40 [0.91 , 2.17]
FE Model for Subgroup 2.36e-04 1.73 [1.29 , 2.32]
In silico replication
deCODE 0.009 0.011 0.54 0.83 [0.46 , 1.51]
FE Model for Subgroup 0.54 0.83 [0.46 , 1.51]
Direct replication
Heidelberg-EPIC 0.008 0.006 0.48 1.35 [0.58 , 3.15]
IARC replication 0.007 0.005 0.35 1.35 [0.72 , 2.53]
ICR replication 0.012 0.008 0.22 1.55 [0.78 , 3.10]
Toronto 0.008 0.006 0.41 1.43 [0.62 , 3.32]
FE Model for Subgroup 0.06 1.42 [0.98 , 2.05]
FE Model for All Studies 1.47 [1.19 , 1.82]
P_het = 0.50;I^2 = 0.00
P = 4.66e-04
0.25
0.50
1.00
3.00
10.00
Odds Ratio

(j)


rs56084662
Study
EAF
Cases
Controls
P
Odds ratio [95% CI]
Discovery
IARC 0.009 0.004 0.02 3.02 [1.19 , 7.68]
ICR 0.011 0.006 0.03 3.02 [1.13 , 8.07]
MDACC 0.011 0.007 0.13 1.94 [0.82 , 4.56]
NCI 0.006 0.005 0.51 1.22 [0.67 , 2.23]
FE Model for Subgroup 2.52e-03 1.85 [1.24 , 2.75]
In silico replication
deCODE 0.007 0.009 0.48 0.78 [0.39 , 1.55]
FE Model for Subgroup 0.48 0.78 [0.39 , 1.55]
Direct replication
Heidelberg-EPIC 0.004 0.006 0.44 0.64 [0.21 , 1.96]
IARC replication 0.006 0.003 0.04 2.10 [1.02 , 4.32]
ICR replication 0.009 0.005 0.13 1.85 [0.84 , 4.06]
Toronto 0.006 0.003 0.17 2.15 [0.72 , 6.46]
FE Model for Subgroup 0.02 1.69 [1.09 , 2.62]
FE Model for All Studies 1.56 [1.19 , 2.05]
P_het = 0.15;I^2 = 0.33
P = 1.41e-03
0.50
1.00
3.00
8.00
Odds Ratio

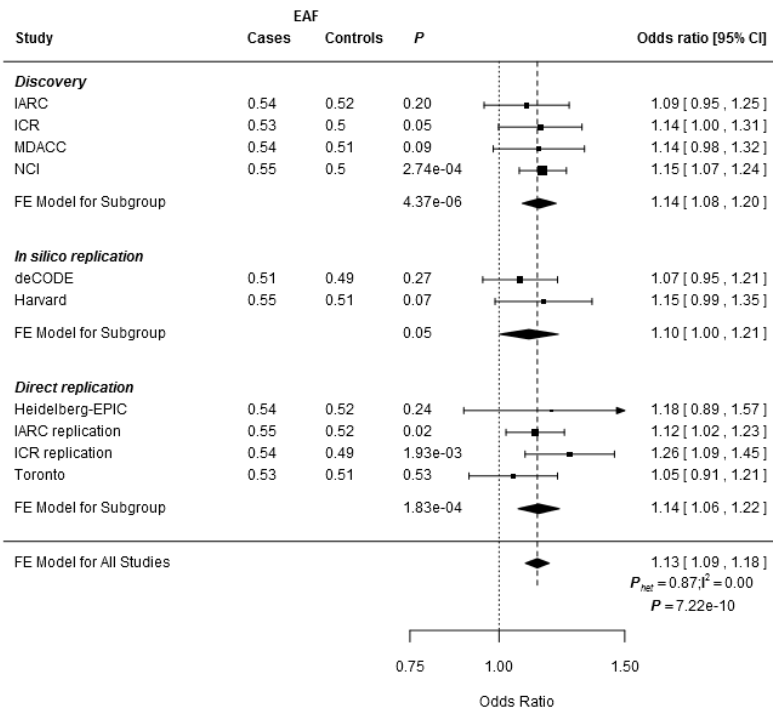
(k)

rs17879961



(l)

rs13314271



Nature Genetics: doi:10.1038/ng.3002

Supplementary Table 4: Association between *BRCA2*-K3326X and *CHEK2*-I157T genotype and lung cancer risk by country of origin.

a) *BRCA2*-K3326X (rs11571833) and risk of all lung cancer

		EAF		Size				
Study	**Country**	**Cases**	**Controls**	**Cases**	**Controls**	**OR**	**95% CI**	***P***
Discovery								
IARC	Romania	0.006	0.009	75	116	0.84	0.04-16.38	0.91
	Hungary	0.011	0.01	223	219	1.08	0.24-4.82	0.92
	Poland	0.01	0.003	543	635	4.18	1.02-17.16	0.05
	Russia	0.017	0.007	318	622	3.34	1.21-9.25	0.02
	Slovakia	0.021	0.023	273	156	0.84	0.31-2.30	0.73
	Czech Republic	0.021	0.004	164	379	8.50	1.67-43.25	9.95×10^{-3}
	France	0.01	0.006	127	144	1.89	0.16-21.68	0.61
	USA	0.017	0.009	362	354	2.20	0.75-6.55	0.15
	Norway	0.012	0.008	342	343	1.65	0.48-5.70	0.43
	Estonia	0.005	0.006	106	823	1.19	0.13-11.35	0.88
NCI	Finland	0.018	0.007	1,732	1,270	2.69	1.48-4.91	1.17×10^{-3}
	Italy	0.015	0.011	1,917	1,978	1.56	0.99-2.46	0.06
	USA	0.01	0.008	2,064	2,488	1.42	0.86-2.35	0.17
UK	UK	0.018	0.009	1,952	5,200	2.38	1.65-3.43	3.81×10^{-6}
USA	MDACC	0.014	0.011	1,150	1,134	1.42	0.77-2.59	0.25
FE Model for Subgroup						1.9	1.57-2.30	5.66×10^{-11} P_{het}=0.38; I^2=6.40%
In silico replication								
deCODE	Iceland	0.014	0.011	1319	26380	1.37	0.93-2.01	0.11
Harvard	USA	0.017	0.011	984	970	1.53	0.82-2.86	0.18
FE Model for Subgroup						1.41	1.02-1.96	0.04 P_{het}=0.76; I^2=0%
Direct replication								
IARC	Denmark	0.005	0.005	279	721	1.11	0.28-4.32	0.88
	France	0.043	0.012	23	42	3.90	0.33-45.59	0.28
	Germany	0.007	0.013	137	300	0.54	0.11-2.58	0.45
	Greece	0.028	0.003	53	159	9.48	0.96-93.19	0.05
	Italy	0	0.006	114	249	N/A	N/A	N/A
	The Netherlands	0.014	0.009	107	225	1.59	0.35-7.25	0.55
	Norway	0.036	0	14	27	N/A	N/A	N/A
	Poland	0.006	0.003	808	1,035	1.65	0.61-4.46	0.32
	Russia	0.009	0.004	1,039	2,046	2.36	1.21-4.61	0.01
	Spain	0.013	0.002	112	235	6.44	0.66-62.62	0.11
	Sweden	0.01	0.009	51	113	1.11	0.10-12.53	0.93
	UK	0.01	0.011	153	330	0.92	0.24-3.62	0.90
Toronto	Canada	0.012	0.006	1,084	966	1.86	0.91-3.77	0.09
UK	UK	0.016	0.008	2,448	2,989	2.13	1.47-3.09	6.60×10^{-5}
Heidelberg-EPIC	Germany	0.014	0.006	1,362	1,253	2.29	1.23-4.26	8.80×10^{-3}
FE Model for Subgroup						2.00	1.58-2.54	1.05×10^{-8} P_{het}=0.73; I^2=0%
FE Model for All Studies						1.84	1.60-2.10	1.84×10^{-18} P_{het}=0.58; I^2=0%

b) *CHEK2*-I157T (rs17879961) and risk of all lung cancer

(Data for directly genotyped samples in the GWAS discovery phase shown in parentheses)

		EAF		Size				
Study	**Country**	**Cases**	**Controls**	**Cases**	**Controls**	**OR**	**95% CI**	***P***
Discovery								
IARC	Romania	0.002	0.006	75	116	0.18	0.00-62.25	0.57
		(0.007)	(0)	(75)	(90)	-	-	(0.45)
	Hungary	0.016	0.037	223	219	0.35	0.13-0.92	0.04
		(0.014)	(0.032)	(220)	(216)	(0.41)	(0.15-1.10)	(0.08)
	Poland	0.019	0.031	543	635	0.52	0.29-0.94	0.03
		(0.014)	(0.030)	(524)	(582)	(0.46)	(0.25-0.86)	(0.01)
	Russia	0.016	0.043	318	622	0.31	0.15-0.64	1.57×10^{-3}
		(0.013)	(0.038)	(309)	(602)	(0.31)	(0.14-0.66)	(2.54×10^{-3})
	Slovakia	0.009	0.004	273	156	4.92	0.35-68.37	0.24
		(0.004)	(0.003)	(265)	(151)	(1.07)	(0.10-11.98)	(0.95)
	Czech Republic	0.013	0.018	164	379	0.66	0.20-2.20	0.50
		(0.007)	(0.014)	(153)	(361)	(0.49)	(0.11-2.29)	(0.37)
	France	0	0.004	127	144	-	-	-
	USA	0.004	0.005	362	354	0.77	0.13-4.62	0.77
	Norway	0.016	0.021	342	343	0.31	0.10-0.93	0.04
	Estonia	0.04	0.046	106	823	0.84	0.40-1.75	0.65
NCI	Finland	0.021	0.038	1,732	1,270	0.52	0.37-0.74	1.66×10^{-4}
	Italy	0.001	0.004	1,917	1,978	0.32	0.11-0.96	0.04
	USA	0.003	0.006	2,064	2,488	0.56	0.27-1.15	0.12
UK	UK	0.0004	0.0018	1,952	5,200	0.34	0.12-0.95	0.04
USA	MDACC	0.004	0.0039	1,150	1,134	1.05	0.35-3.09	0.94
FE Model for Subgroup						0.51	0.41-0.63	2.96×10^{-10} P_{het}=0.59; I^2=0%
In silico replication								
deCODE	Iceland	0.0021	0.0025	1,319	26,380	0.83	0.37-1.89	0.66
Harvard	USA	0.0006	0.0034	984	970	0.14	0.02-1.18	0.07
FE Model for Subgroup						0.66	0.31-1.43	0.29 P_{het}=0.12; I^2=58%
Direct replication								
IARC	Denmark	0.005	0.008	274	709	0.64	0.18-2.30	0.49
	France	0	0	24	43	-	-	-
	Germany	0.004	0.005	138	298	0.72	0.07-6.96	0.78
	Greece	0	0.006	56	165	-	-	-
	Italy	0	0	117	251	-	-	-
	The Netherlands	0	0.002	105	226	-	-	-
	Norway	0	0	15	29	-	-	-
	Poland	0.015	0.026	777	914	0.58	0.35-0.95	0.03
	Russia	0.024	0.027	1,034	2,030	0.86	0.61-1.21	0.39
	Spain	0	0	108	227	-	-	-
	Sweden	0	0	56	118	-	-	-
	UK	0	0	157	330	-	-	-
Toronto	Canada	0.003	0.005	1,084	966	0.63	0.22-1.82	0.39
UK	UK	0.008	0.005	2,448	2,989	1.63	0.36-7.30	0.52
Heidelberg-EPIC	Germany	0.005	0.006	1,365	1,180	0.8	0.38-1.68	0.55
FE Model for Subgroup						0.77	0.60-0.98	0.03 P_{het}=0.83; I^2=0%
FE Model for All Studies						0.61	0.52-0.71	3.86×10^{-10} P_{het}=0.43; I^2=3%

c) *CHEK2*-I157T (rs17879961) and risk of squamous cell lung cancer using directly genotyped samples

Study	Country	EAF Cases	EAF Controls	Size Cases	Size Controls	OR	95% CI	*P**
Discovery								
IARC	Romania	0.031	0	16	90	-	-	0.15
	Hungary	0.011	0.032	88	216	0.35	0.04-1.55	0.26
	Poland	0.003	0.030	186	582	0.09	0.00-0.54	1.10×10^{-3}
	Russia	0.005	0.038	188	600	0.14	0.02-0.54	7.06×10^{-4}
	Slovakia	0.004	0.003	124	151	1.22	0.02-95.83	1.00
	Czech Republic	0	0.014	70	361	-	-	0.38
FE Model for Subgroup						0.17	0.07-0.39	3.57×10^{-5} P_{het}=0.34; I^2=11%
Direct replication								
IARC	Denmark	0.015	0.008	65	709	1.8	0.19-8.23	0.34
	France	-	0	0	43	-	-	-
	Germany	0	0.005	26	298	-	-	-
	Greece	0	0.006	12	165	-	-	-
	Italy	0	0	26	251	-	-	-
	The Netherlands	0	0.002	20	226	-	-	-
	Norway	0	0	1	29	-	-	-
	Poland	0.005	0.026	318	914	0.18	0.04-0.56	4.94×10^{-4}
	Russia	0.013	0.027	496	2,030	0.48	0.25-0.86	0.01
	Spain	0	0	28	227	-	-	-
	Sweden	0	0	16	118	-	-	-
	UK	0	0	55	330	-	-	-
Toronto	Canada	0.002	0.005	205	966	0.52	0.01-3.79	1.00
UK	UK	0.001	5×10^{-4}	1,043	2,989	1.9	0.16-16.57	0.61
Heidelberg-EPIC	Germany	0.003	0.006	345	1,180	0.48	0.05-2.09	0.55
FE Model for Subgroup						0.44	0.29-0.68	2.16×10^{-4} P_{het}=0.13; I^2=41%
FE Model for All Studies						0.34	0.23-0.50	3.52×10^{-8} P_{het}=0.08; I^2=42%

* *P*-values and CIs for individual associations from Fishers exact test (two-sided)

Supplementary Table 5: Relationship between K3326X, 999del5 genotypes and risk of breast, squamous cell head and neck, lung and ovarian cancer in the Icelandic population

Variant	Cancer	OR	GC corrected *P*-value
K3326X	Breast	1.08	0.53
K3326X	Head and neck (squamous)	1.12	0.72
K3326X	Lung	1.55	**2.36×10^{-3}**
K3326X	Ovarian	1.27	0.40
999del5	Breast	13.15	**1.37×10^{-109}**
999del5	Head and neck cancer (squamous)	2.82	**0.01**
999del5	Lung	0.96	0.87
999del5	Ovarian	9.60	**4.47×10^{-19}**

39 of the 2,231 sequenced Icelandic individuals were carriers of the minor allele of K3326X
37 of the 2,231 sequenced Icelandic individuals were carriers of the minor allele of 999del5
None of the 2,231 sequenced Icelandic individuals carried both mutations

Supplementary Table 6: (a) Frequency of somatic *BRCA2* and *CHEK2* mutations in lung cancer (Data derived from TCGA)

		TCGA exome-sequencing data				TCGA Affymetrix SNP 6.0 array data	
Gene variant	**Tumour histology**	**Total number of tumours**	**Number of variant carriers**	**Number of tumours with SNV in respective gene[1]**	**Number of tumours with a somatic INDEL in respective gene**	**Number of tumours showing CNA in respective gene**	**Number of gene carriers with a somatic second hit[2]**
BRCA2-K3326X	SQ (LUSC)	243	13	13	0	25 (25 loss)	2
	AD (LUAD)	338	11	18	0	19 (4 gain, 16 loss)	0
CHEK2-I157T	SQ (LUSC)	243	2	11	0	35 (31 gain, 4 loss)	0
	AD (LUAD)	338	3	16	0	11 (3 gain, 8 loss)	0

SNV=single nucleotide variant; INDEL= insertion/deletion; CNA= copy number alteration

[1]Mutect, exonic mutations, deleterious or missense
[2] Somatic mutation in tumours defined by SNV, INDEL or CNA

Supplementary Table 6: (b) genomic annotation of *TP63* by HaploReg v2 and RegulomeDB. Data are shown for rs13314271 (3q28 *TP63*) and proxy SNPs (r2>0.8 in 1000 Genomes EUR phase 1 data) demonstrating evidence of histone marks, DNAse hypersensitivity sites or transcription factor occupancy in RegulomeDB analysis. Also indicated are GERP scores>2 and RegulomeDB scores for all SNPs.

SNP	Chr	Pos (hg19)	LD[a] (r^2)	LD[a] (D')	EUR freq	GERP score[b]	Enhancer histone marks	DNAse HS	Proteins bound	Motifs changed	GENCODE genes	dbSNP annotation	Regulome DB score
rs36108040	3	189335844	0.95	0.98	0.49					1	13kb 5' of TP63		No Data
rs55779747	3	189354127	0.99	0.99	0.49		H1	WERI-Rb-1		5	TP63	intronic	5
rs4488809	3	189356261	0.98	0.99	0.49		GM12878	GM12864		8	TP63	intronic	5
rs7636839	3	189356941	0.99	0.99	0.49	2.92	GM12878			4	TP63	intronic	6
rs13080835	3	189357199	1	1	0.49		GM12878	GM06990		5	TP63	intronic	5
rs13314271	**3**	**189357602**	**1**	**1**	**0.49**					**4**	**TP63**	**intronic**	**No Data**
rs12696594	3	189357616	1	1	0.49					5	TP63	intronic	6
rs55862124	3	189359836	0.99	0.99	0.49		GM12878, HMEC	4 cell types	ERALPHA_A	4	TP63	intronic	4
rs7619517	3	189360235	0.99	0.99	0.49						TP63	intronic	No Data
rs7629983	3	189360406	0.98	0.99	0.49					6	TP63	intronic	No Data
rs13084874	3	189361400	0.98	0.99	0.49						TP63	intronic	No Data
rs9811174	3	189366595	0.8	-0.98	0.46					4	TP63	intronic	6
rs9811214	3	189366763	0.8	-0.98	0.46					2	TP63	intronic	No Data
rs9821164	3	189368255	0.8	-0.98	0.46					4	TP63	intronic	6
rs4426661	3	189368357	0.8	0.98	0.46					3	TP63	intronic	6
rs1920272	3	189378556	0.96	0.98	0.49		GM12878				TP63	intronic	No Data

Chr,chromosome; DNAse HS, DNAse hypersensitivity; Freq, frequency; GERP, Genomic Evolutionary Rate Profiling; LD, linkage disequilibrium; Pos,position;

[a] LD (r2) is based upon 1000 Genomes EUR data and a threshold of r2>0.8 was imposed to identify correlated SNPs

[b] GERP scores indicative of evolutionary constraint (>2) are listed

[d] RegulomeDB scores: 4, TF binding + DNase peak; 5, TF binding or DNase peak; 6, other binding or DNase peak.

Supplementary Table 7: Relationship between rs11571833 (*BRCA2* K3326X), rs17879961 (*CHEK2* I157T) and rs13314271 (*TP63*) genotype and cigarette consumption. Analysis based on data from 43,693 Icelandic subjects (including 34,850 chip typed individuals).

Variant	Minor allele	Other allele	MAF in Icelandic population	MAF in CEU	Info score	Genomic control corrected *P*-value
BRCA2						
rs11571833	T	A	1.04%	1.05%	1.00	0.62
CHEK2						
rs17879961	G	A	0.66%	0.51%	0.51	0.59
TP63						
rs13314271	T	C	48.95%	48.95%	1.00	0.81

Imputation of variants was performed using 2,231 long range phased whole-genome sequenced Icelandic individuals (Kong A, *et al*., Nature 2009, 462:868; Styrkarsdottir U, *et al*., Nature 2013, 497:517).