Supplementary

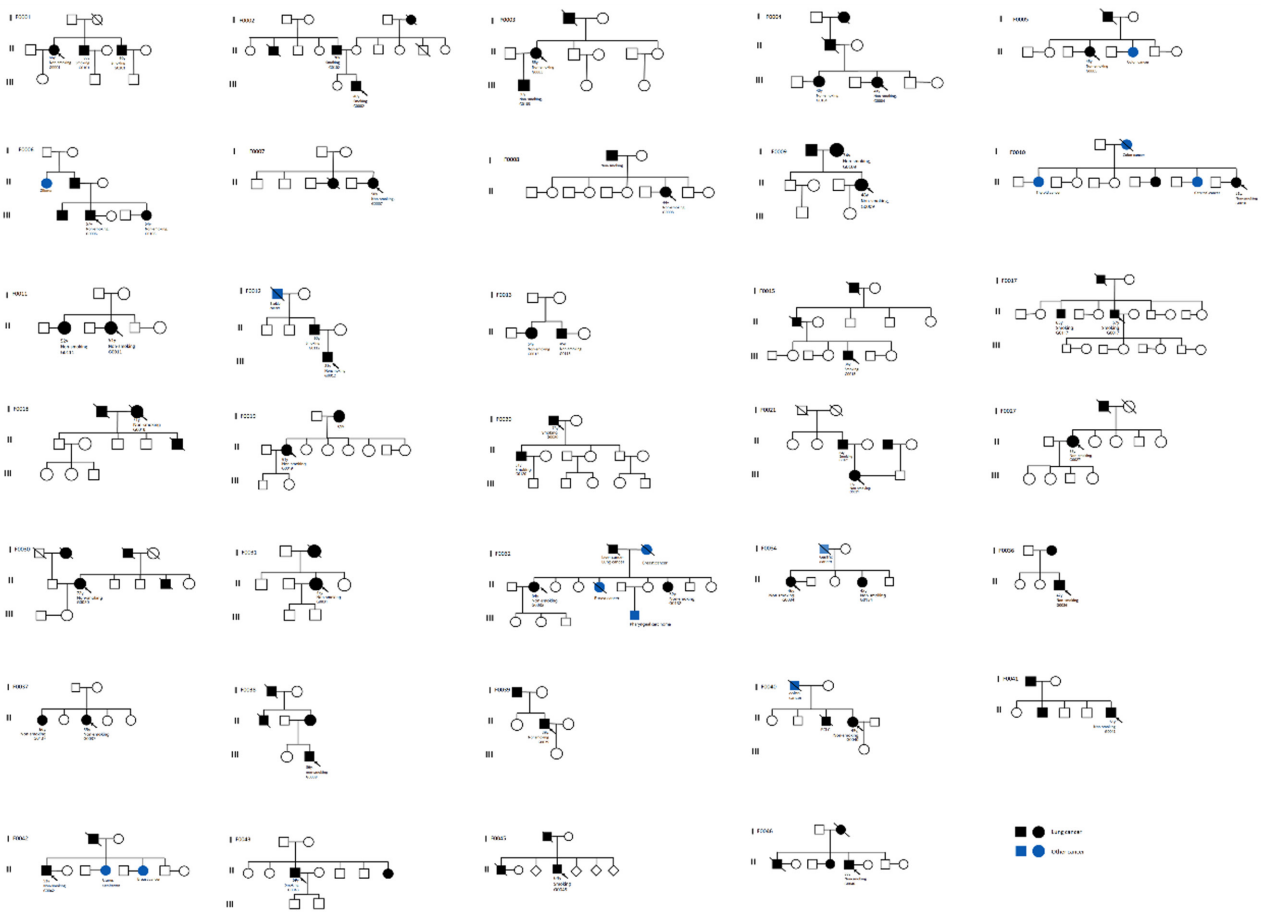

Figure S1 Family pedigree of the 34 GGN families.

https://dx.doi.org/10.21037/tlcr-21-789

Table S1 Clinicopathological characteristics of GGN patients

Characteristic	Patients with GGN (n=50)
Age at diagnosis	
Median	51
Range	30-75
Gender - No. (%)	
Female	27
Male	23
Smoking – No. (%)	
No	37
Yes	13
GGO subtype – No. (%)	
Pure GGO	21
Mixed GGO	23
Others [*1]	6
Numbers of GGO – No. (%)	
1	31
≥2	19
Pathology – No. (%)	
AIS	9
MIA	16
IAC	24
Others [*2]	1
Clinical stage – No. (%)	
0	10
IA1	24
IA2	10
Others [*3]	6

*1 others: solid tumor or solid tumor + pure/mixed GGO; *2 others: mixed mucus adenocarcinoma and non-mucus adenocarcinoma; others *3: IA3 or later stage; GGN, ground-glass nodule; GGO, ground-glass opacity; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma.

https://dx.doi.org/10.21037/tlcr-21-789

Table S2 Clinicopathological characteristics of lung cancer patients without family history of lung cancer

Patient ID	Gender	Cancer family history	Cancer patients in the family	Smoking	Subtypes of lung cancer	Stage	Age at diagnosis	Somatic Actionable mutations
lc001	F	No		No	Adenocarcinoma	I	63	
lc002	F	No		No	Adenocarcinoma	IV	47	EGFR L858R
lc003	F	No		No	Adenocarcinoma	IV	58	EGFR EX19del
lc004	F	No		No	Adenocarcinoma	IV	45	EGFR EX19del
lc005	F	No		No	Adenocarcinoma	IV	54	EGFR L858R
lc006	M	No		Yes	Adenocarcinoma	IV	52	EGFR EX19del
lc007	F	No		No	Adenocarcinoma	IV	49	EGFR L858R
lc008	F	No		No	Adenocarcinoma	III	51	EGFR EX19del
lc009	M	No		Yes	Adenocarcinoma	II	56	
lc010	M	No		No	adenosquamous carcInoma	IV	57	RET fusion
lc011	F	No		No	Adenocarcinoma	II	52	KRAS Q61H
lc012	F	No		No	Adenocarcinoma	II	46	ROS1 fusion
lc013	F	No		No	Adenocarcinoma	II	52	EGFR EX19del
lc014	M	No		No	Adenocarcinoma	I	46	EGFR EX19del
lc015	M	No		No	Adenocarcinoma	II	68	
lc016	F	No		No	Adenocarcinoma	IV	60	
lc017	F	No		No	Adenocarcinoma	IV	22	
lc018	F	No		No	Adenocarcinoma	I	63	EGFR L858R
lc019	M	No		No	Adenocarcinoma	II	70	EGFR L858R
lc020	F	No		No	Adenocarcinoma	I	51	EGFR EX19del
lc021	M	No		Yes	Adenocarcinoma	I	79	EGFR G719S+S768I
lc022	F	No		No	Adenocarcinoma	IV	57	EGFR L858R
lc023	F	No		No	Adenocarcinoma	I	52	EGFR L858R
lc024	M	No		No	Adenocarcinoma	IV	41	EGFR EX19del
lc025	F	No		No	Adenocarcinoma	IV	49	ALK fusion
lc026	F	No		Yes	Adenocarcinoma	IV	62	
lc027	F	No		No	Adenocarcinoma	II	51	EGFR EX19del
lc028	M	Yes	Father, gastric cancer	Yes	Adenocarcinoma	IV	79	EGFR EX19del
lc029	F	No		No	Adenocarcinoma	II	49	BRAF K601E
lc030	M	Yes	Father, CNS cancer	No	Adenocarcinoma	I	64	
lc031	F	No		No	Adenocarcinoma	II	40	ERBB2 EX20Ins
lc032	M	No		Yes	Adenocarcinoma	I	52	
lc033	M	Yes	Father, gastric cancer	No	Adenocarcinoma	I	48	
lc034	F	No		No	Adenocarcinoma	I	50	EGFR L858R
lc035	F	No		No	Adenocarcinoma	I	52	BRAF K601E
lc036	F	No		No	Adenocarcinoma	II	56	EGFR L858R
lc037	F	No		No	Adenocarcinoma	I	50	EGFR L858R
lc038	F	Yes	Mother, breast cancer	No	Adenocarcinoma	IV	45	EGFR EX19del
lc039	F	No		No	Adenocarcinoma	I	46	EGFR EX19del

https://dx.doi.org/10.21037/tlcr-21-789

Table S3 Number of variants during the filtering of the GGN, lung cancer and 678 healthy control cohort

Filters	GGN family cohort	678 healthy people cohort	Patients without lung cancer family history
all loci by WES	435980 SNVs + 119189 Indels	606326 SNVs + 75106 Indels	142387 SNVs+21947 Indels
frameshift + missense + splicing + stop gain	40770 SNVs + 1168 Indels	193556 SNVs + 6118 Indels	39421 SNVs +1044 Indels
MAF<0.01 from ExAC, 1000G,gnomAD and 678 asian WES cohort	13008 SNVs + 440 frameshifts	157129 SNVs + 4643 frameshift	11962 SNVs +396 frameshifts
predicted as damaging and deleterious or framshift	3786 SNVs + 440 frameshifts	46160 SNVs + 4643 frameshift	3456 SNVs +360 frameshifts
presented in all the tested patients of the same family	2325 SNVs + 238 frameshifts	-	-
MAF<0.01 in ALFA database, M-CAP defined as pathogenic or likely pathogenic	1517 SNVs + 238 frameshifts	32329 SNVs + 4643 framefhift	2391 SNVs + 342 frameshifts
Variants in at least 2 families	35 SNVs + 10 frameshifts	-	59 SNVs + 19 frameshifts
Variants in at least 3 families	4 SNVs + 1 frameshift	-	3 SNVS + 8 frameshifts

Table S4 Recurrent rare germline variants identified in the 34 families

SYMBOL	SNV / InDels	Existing_variation	MAF in 678 local healthy people	family	case
ANKRD24	p.Arg1013Gln	rs199779504,COSM3227404, COSM3227405,COSM3227406	0.00737463	8,31	G0008,G0031
ANKRD33	p.Pro429Leu	rs770135401	-	8,37	G0137,G0008,G0037
ARFGAP3	p.Ser317Leu	rs147432132	0.00884956	32,46	G0132,G0023,G0032,G0046
CA5A	p.Pro82Thr	rs377135599	0.00294985	9,17	G0117,G0109,G0017,G0009
CABP1	p.Arg92Trp	rs543428199,COSM5672565	0.00884956	38,45	G0038,G0045
CATSPERG	p.Thr576Asn	rs200132227	0.00147493	5,10,36	G0036,G0010,G0005
CCDC66	p.Thr834AsnfsTer68	rs370165016	-	8,46	G0008,G0046
CKAP2L	p.Pro482Arg	rs1037829641,COSM4084268	0.00147493	36,41	G0036,G0041
CNTROB	p.Arg892Cys	rs151174639	0.00589971	5,15	G0015,G0005
CPPED1	p.Ala192Thr	rs192649616	0.00884956	5,46	G0005,G0046
CRIPAK	p.Cys38SerfsTer369	rs528457959	0.00884956	15,21,46	G0121,G0015,G0021,G0046
CYP2D6	p.Arg441His	rs532668079	-	34,36	G0134,G0034,G0036
DNAH7	p.Val827SerfsTer20	rs752238407	0.00737463	31,45	G0031,G0045
DTHD1	p.Ser754ValfsTer25	rs529758698	0.00737463	7,41	G0007,G0041
FAM178B	p.Gly381Val	rs764132573	0.00147493	2,30	G0102,G0002,G0030
GAS2L2	p.Arg273Cys	rs140842796	0.00294985	15,40	G0015,G0040
GSN	p.His4ArgfsTer86	rs764841269	0.00442478	20,30	G0120,G0030,G0020
KRT73	p.Arg212Cys	rs116282210	0.00147493	3,8,36	G0103,G0008,G0036,G0003
KRTAP5-5	p.Ala53GlyfsTer129	rs762422220	0.00737463	30,41	G0041,G0030
MAPK15	p.Pro140Leu	rs201842849	0.00294985	30,43	G0030,G0043

Table S4 (*continued*)

https://dx.doi.org/10.21037/tlcr-21-789

Table S4 (*continued*)

SYMBOL	SNV / InDels	Existing_variation	MAF in 678 local healthy people	family	case
MCM3AP	p.Leu885Phe	rs201315959	0.00442478	5,36	G0036,G0005
MSH5	p.Ala685Thr	rs561487480	0.00294985	18,32	G0132,G0032,G0018
MTHFD1L	p.Thr619Met	rs143492706	0.00147493	10,42	G0010,G0042
MUC2	p.Thr446Met	rs199865570	0.00737463	15,18	G0015,G0033,G0018
MYO7A	p.Cys1201Ser	rs117966637,CM1212017	0.00442478	4,5	G0104,G0004,G0005
MYOD1	p.His88Arg	rs544592180	0.00442478	4,42	G0104,G0004,G0029,G0042
NUPL2	p.Tyr174Cys	rs199844379	0.00589971	21,40	G0121,G0040,G0023,G0021
ODAM	p.Pro88Ser	rs373877978	0.00147493	36,38	G0036,G0038
PRAMEF1	p.Leu354SerfsTer20	rs531127236	0.00589971	40,41	G0040,G0041
RHBDF2	p.Arg109Cys	rs369829771	0.00147493	11,15	G0111,G0015,G0011
SAMM50	p.Arg267Gln	rs78038328	0.00884956	4,15,18	G0104,G0015,G0004,G0018
SHANK3	p.Ala463GlyfsTer40	-	-	7,27	G0007,G0027
SLC22A12	p.Arg90His	rs121907896,CM042474	0.00737463	31,43	G0031,G0043
SLC2A8	p.Asp119ThrfsTer21	rs749728472	0.00147493	18,34	G0134,G0034,G0018,G0016
SLC6A20	p.Val53Met	rs371916242	-	10,18	G0010,G0018
STON1-GTF2A1L	p.Val617Asp	rs747845774	-	17,27	G0117,G0027,G0017
SYCE1L	p.Arg54Trp	rs368565145	0.00294985	3,43	G0103,G0003,G0043
TAS1R3	p.Ala403Val	rs548456115	-	19,46	G0019,G0046
TCP10	p.Pro269HisfsTer7	rs778595860,COSM216006	0.00442478	8,12	G0112,G0008,G0012
TDRD12	p.Phe897Ser	rs1350637914	-	30,31,45	G0031,G0030,G0045
TINAG	p.Ser324Asn	rs147494351	0.00442478	7,9	G0109,G0007,G0009
TLDC2	p.Arg184Cys	rs148426788	0.00442478	39,41	G0039,G0041
TRIO	p.Pro67Ser	rs146453151	0.00294985	10,36	G0036,G0010
TRPM1	p.Asp842His	rs771110434	0.00147493	10,40	G0040,G0010
VPS33B	p.Ile383Thr	rs149121639,COSM4400756	-	6,38	G0106,G0006,G0038

https://dx.doi.org/10.21037/tlcr-21-789

Figure S2 Chromosomal position, gene structure, protein domain(s), sequencing reads and evolutional conservation analysis of the candidate mutations of *CYP2D6*.

https://dx.doi.org/10.21037/tlcr-21-789

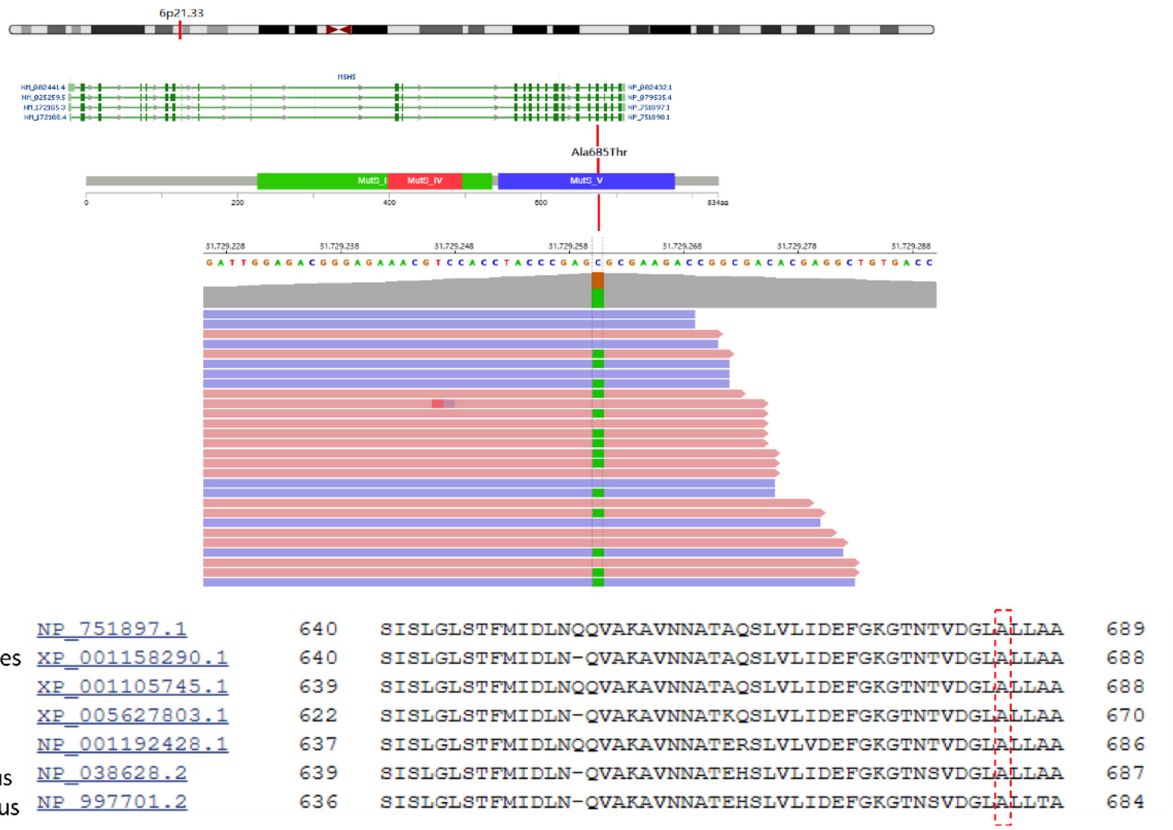

Figure S3 Chromosomal position, gene structure, protein domain(s), sequencing reads and evolutional conservation analysis of the candidate mutations of *MSH5*.

https://dx.doi.org/10.21037/tlcr-21-789

Table S5 Gene-based burden testing of the DNA repair genes with rare mutations

SYMBOL	Existing_variation	Family ID	Case ID	p value	somatic DDR variants (case ID)
APEX1	rs1413851946,rs780293860	27	G0027,G0027	0.1367	
BRCA2	rs200598289	46	G0046	>0.9999	
BRIP1	rs201869624	10	G0010	>0.9999	
FANCD2	rs767860064	46	G0046	0.2176	
FANCI	rs200186938	40	G0040	>0.9999	TP53 p.E28Kfs*14 (G0040)
FANCM	rs202171930,rs148304968	15,31	G0015,G0031	0.2105	TOPBP1 p.G274V (G0031)
GTF2H1	rs191761375	40	G0040	0.3578	TP53 p.E28Kfs*14 (G0040)
MNAT1	rs118051600	11	G0111,G0011	0.3578	
MSH4	rs116141807,rs780475342	21,42	G0121,G0021,G0042	0.1083	
MSH5	rs561487480,rs746903566	18,32,38	G0018,G0132,G0032,G0038	0.0095	REV3L p.T2275S (G0038)
PNKP	rs756933064	17	G0117,G0017	>0.9999	PRKDC p.H1464L (G0117)
POLG	rs796052895	27	G0027	>0.9999	
POLQ	rs759231797	19	G0019	>0.9999	
RAD52	-	15	G0015	0.4736	
RAD54L	rs186059216,-	9,27	G0109,G0009,G0027	0.1407	REV3L p.R2499* (G0109), TP53 p.Y163H (G0109)
SLX4	rs377440877	19	G0019	>0.9999	
XAB2	rs200271935	42	G0042	0.6101	SLX4 p.R237Q (G0042)
XPA	-	10	G0010	0.0933	
XRCC1	rs199748521	45	G0045	0.5236	

https://dx.doi.org/10.21037/tlcr-21-789