Supplemental Information

Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer

Kai Wang[1,7], Junsuo Kan[2,7], Siu Tsan Yuen[2], Stephanie T Shi[3], Kent Man Chu[4], Simon Law[4], Tsun Leung Chan[2], Zhengyan Kan[1], Annie SY Chan[2], Wai Yin Tsui[2], Siu Po Lee[2], Siu Lun Ho[2], Anthony KW Chan[2], Grace HW Cheng[2], Peter C Roberts[5], Paul A Rejto[1], Neil W Gibson[1,6], David J Pocalyko[1], Mao Mao[1], Jiangchun Xu[1], Suet Yi Leung[2]

[1]Oncology Research Unit, [3]External Research Solutions, [5]Research Embedded Business Technology, Pfizer Worldwide Research and Development, La Jolla, California, USA.

Department of [2]Pathology and [4]Surgery, The University of Hong Kong, Queen Mary Hospital, Pokfulam, Hong Kong.

[6]Present address: Regulus Therapeutics, San Diego, California, USA

[7]These authors contributed equally to this work.

Correspondence should be addressed to J.X. (jiangchun.xu@pfizer.com) or S.Y.L. (suetyi@hku.hk)

Index of all supplementary figures, tables and notes

Supplementary Figures

Supplementary Tables

Supplementary tables 4, 6, 7, 9, 10 and 13 are attached individually in Excel

Supplementary Table 3. Sequenom validation results for whole exome sequencing predicted somatic SNVs and somatic indels

Supplementary Table 4. List of all somatic mutations identified in the SureSelect All Exon kit baited regions in 22 gastric cancers

Supplementary Table 5. Comparison of somatic point mutation spectra across different subtypes of gastric cancers

Supplementary Table 6. Comparison of list of gastric cancer genes with protein altering mutations detected in exome study versus kinome study

Supplementary Table 7. List of genes with protein-altering somatic mutations in at least two patients in the gastric cancer cohort and their driver gene scores

Supplementary Table 8. Core signaling pathways and processes frequently altered in gastric cancers.

Supplementary Table 9. Pathways enriched in somatically mutated genes in gastric cancer

Supplementary Table 10. List of all protein-altering somatic mutations involving chromatin modification genes in 22 gastric cancers

Supplementary Table 11. Clinico-pathological summary of 109 gastric cancer patients sequenced for ARID1A

Supplementary Table 12. Summary of relationship of ARID1A mutation with protein expression and molecular subtypes of gastric cancer

Supplementary Table 13. Complete list of ARID1A and TP53 somatic mutations in 109 gastric cancers and their associated molecular parameters

Supplementary Table 14. Univariate and multivariate cox regression analysis of variables related to recurrence-free survival in 62 gastric cancer patients undergoing curative resection and analyzed for ARID1A mutation

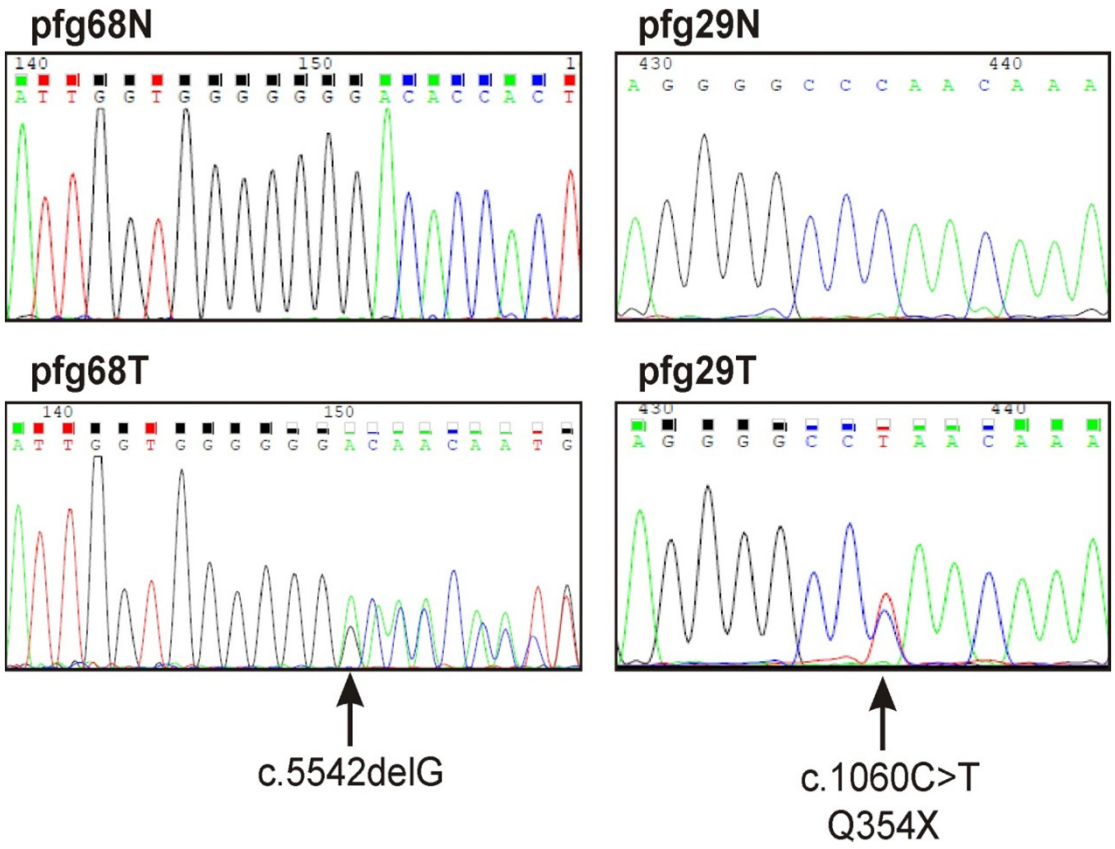
Supplementary Table 15. Comparison of mutation rate of mononucleotide repeats in ARID1A gene versus background mutation rate in MSI gastric cancers

Supplementary Table 16. Primers used for ARID1A and TP53 Sanger sequencing

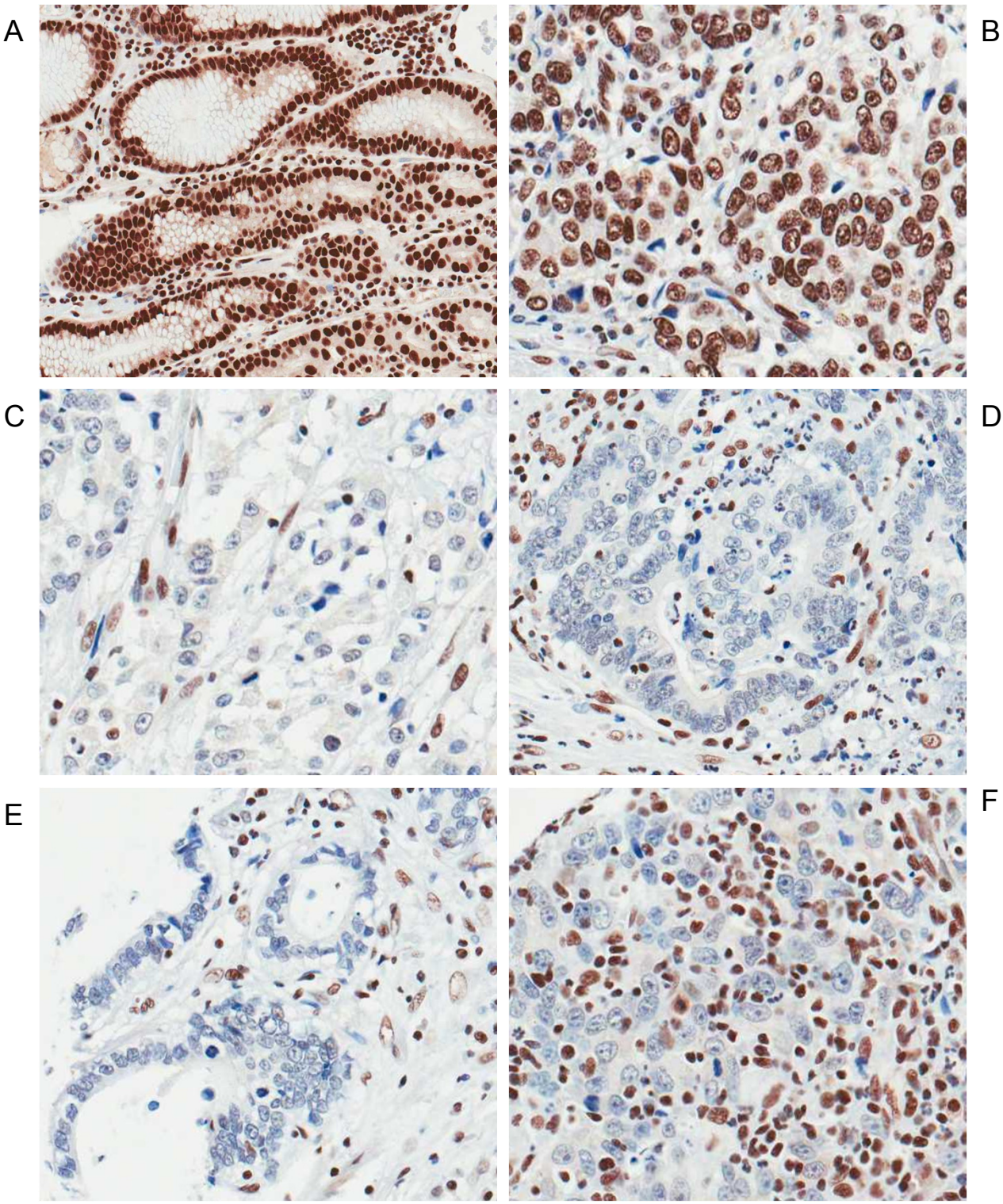
Supplementary Table 17. Background mutation rate estimates for somatic SNVs and somatic indels in the studied gastric cancer cohort

Supplementary Note

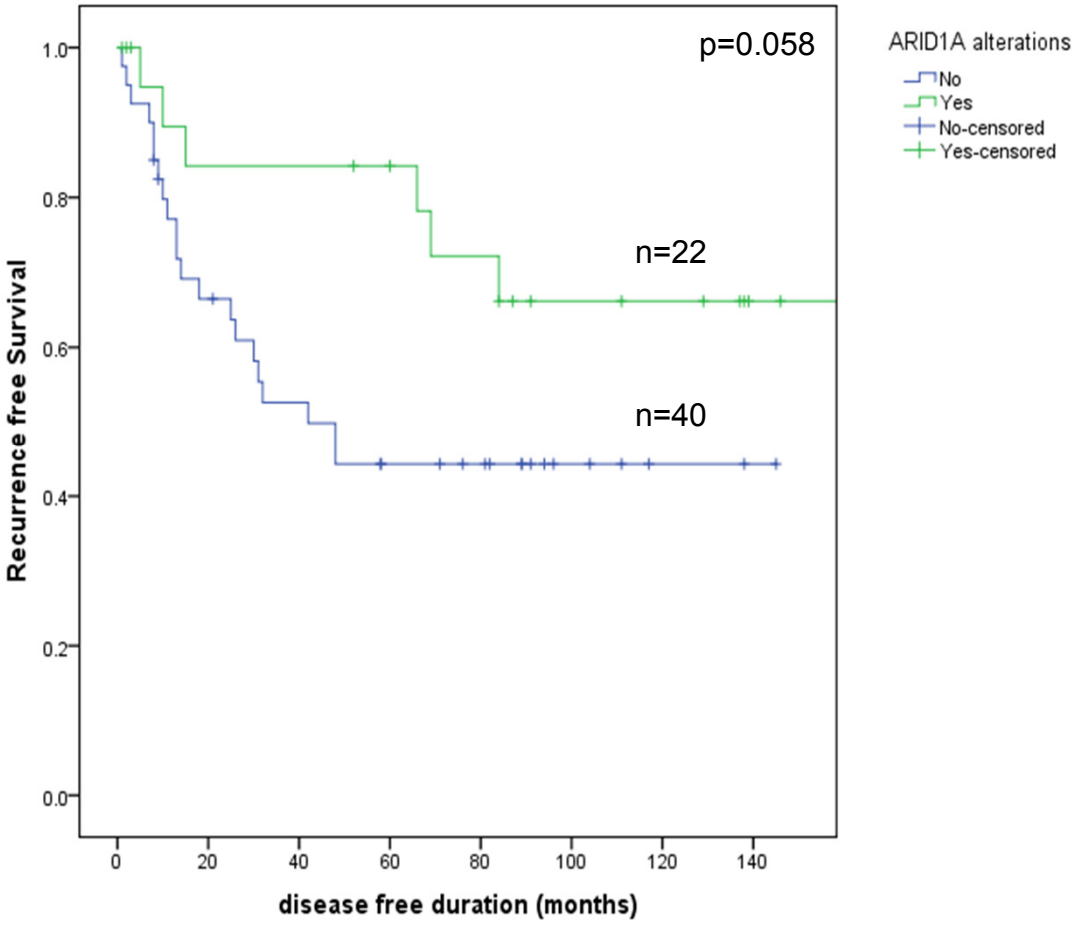
Algorithms for exome-sequencing data analysis, somatic mutation callers, mutation annotation and functional impact prediction, driver gene prediction and pathway analysis



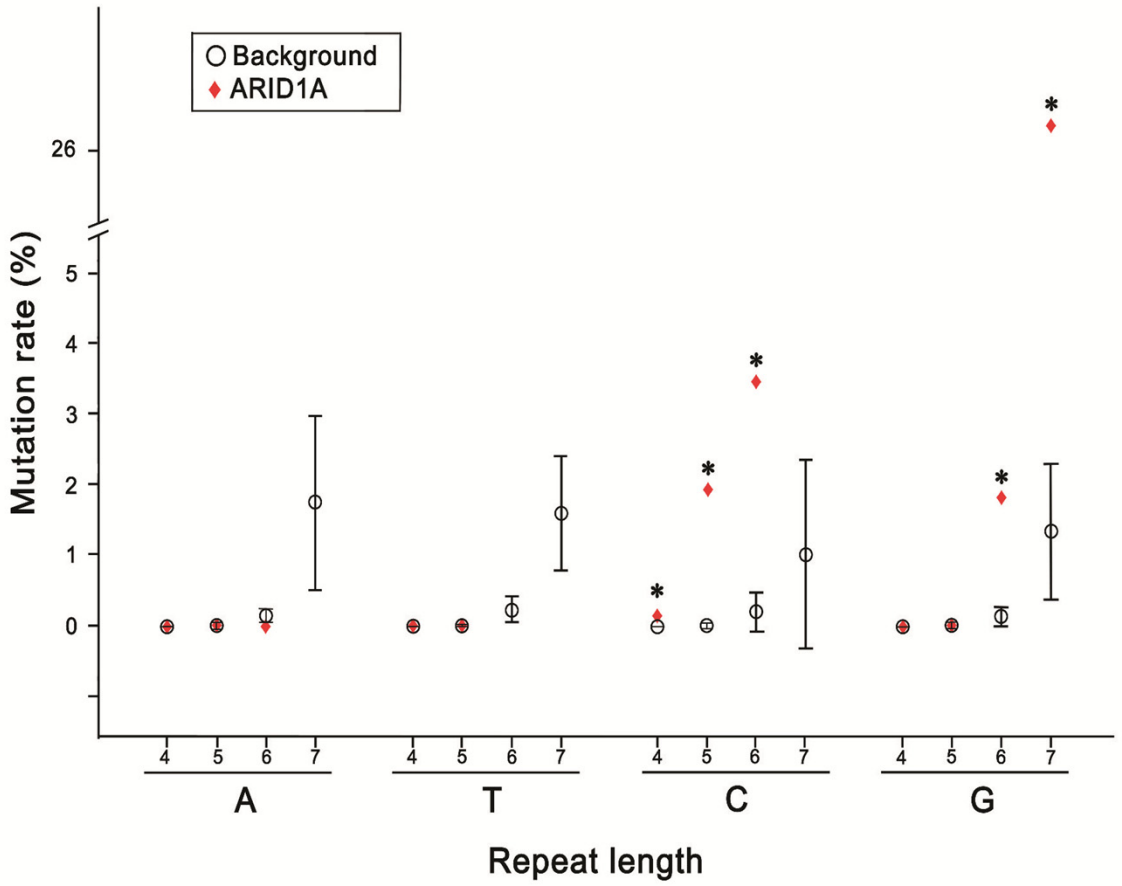
Supplementary Figure 1 Representative DNA sequencing chromatogram of gastric cancers with ARID1A mutation



Supplementary Figure 2 Representative immunostaining results for ARID1A in gastric cancers. (A) Normal gastric mucosa shows uniform nuclear staining of epithelial and stromal cells. (B) A gastric cancer with normal nuclear expression of ARID1A (case pfg014). (C-F) Gastric cancers with loss of nuclear expression of ARID1A. Note the strong positive staining of stromal fibroblasts and lymphocytes. (C) Case pfg018, EBV GC with 1 ARID1A mutation. (D) Case SX186, MSI GC with 2 ARID1A mutation. (E) Case pfg029, MSS non-EBV GC with 1 ARID1A mutation. (F) Case SX165, EBV GC with protein loss despite absence of ARID1A mutation.



Supplementary Figure 3 Kaplan-meier plot of ARID1A alterations (mutation and/or reduced protein level) in relation to recurrence-free survival in 62 gastric cancer patients who underwent curative resection



Supplementary Figure 4 Somatic indel mutation rate of mononucleotide repeat tracts in ARID1A relative to background mutation rate in MSI gastric cancers. The X-axis denotes the mononucleotide repeat length of the 4 nucleotides A, T, C and G. Black circles denote mean background somatic indel mutation rate of specific repeats in 4 MSI GCs derived from the interrogation of all coding exons of 16,586 genes in SureSelect capture baits with adequate sequence coverage. The error bar indicates ±2 standard deviations. Red diamonds indicates somatic indel mutation rate of the specific repeat in ARID1A coding exons in 23 MSI GCs. Some repeats are not present in ARID1A gene. Asterisks denote statistically significant difference at $p \leq 0.000025$ (see Supplementary Table 15 for detailed statistics).

Supplementary Table 1 Clinico-pathological data of 109 gastric cancer patients in the discovery and validation cohort

Sample code	Study cohort	Tumor content (%)	SEX	AGE	EBV status[a]	Microsatellite instability status[b]	Helicobacter pylori infection[c]	Tumor site	Tumor differentiation	Lauren's type	UICC tumor stage[d]	T stage	N stage	M stage	Operation type	Recurrence status	Disease free survival (month)	Survival status	Cancer specific survival (month)
pfg001T	Exome sequencing	70	F	73	negative	MSS	positive	cardia	poor	mixed	IV	T4	2	0	palliative		4	Died of disease	4
pfg002T	Exome sequencing	90	M	39	negative	MSS	negative	antrum	poor	intestinal	IV	T4	0	1	palliative		3	Died of disease	3
pfg003T	Exome sequencing	70	F	65	negative	MSS	positive	body	poor	mixed	IIIA	T3	1	0	curative	No recurrence	146	Alive or censored	146
pfg005T	Exome sequencing	75	F	71	negative	MSS	positive	antrum	poor	diffuse (signet ring cell)	IV	T2a	3	0	curative	Recur	26	Died of disease	38
pfg006T	Exome sequencing	95	M	83	negative	MSS	negative	antrum	moderate	intestinal	II	T3	0	0	curative	No recurrence	76	Alive or censored	76
pfg007T	Exome sequencing	95	M	87	negative	MSS	positive	cardia	moderate	intestinal	IV	T4	1	1	palliative		3	Died of disease	3
pfg008T	Exome sequencing	80	F	68	negative	MSI	positive	antrum	moderate	intestinal	II	T2b	1	0	curative	Recur	84	Died of disease	87
pfg009T	Exome sequencing	90	M	67	positive	MSS	positive	body	poor	intestinal	II	T2b	1	0	curative	Recur	31	Died of disease	31
pfg010T	Exome sequencing	95	M	79	negative	MSS	positive	antrum	well	intestinal	IB	T2b	0	0	curative	Recur	48	Died of disease	72
pfg011T	Exome sequencing	90	M	60	negative	MSS	negative	cardia	poor	intestinal	II	T2b	1	0	curative	No recurrence	96	Alive or censored	96
pfg014T	Exome sequencing	85	M	74	negative	MSS	positive	cardia	poor	intestinal	IV	T3	3	0	palliative		3	Died of disease	5
pfg015T	Exome sequencing	80	F	77	negative	MSS	negative	antrum	poor	diffuse (signet ring cell)	IV	T3	1	1	palliative		2	Died of disease	2
pfg016T	Exome sequencing	90	M	55	negative	MSI	negative	antrum	moderate	intestinal	IV	T4	1	0	curative	No recurrence	178	Alive or censored	178
pfg017T	Exome sequencing	80	M	63	negative	MSI	negative	antrum	moderate	intestinal	II	T2b	1	0	curative	No recurrence	184	Alive or censored	184
pfg018T	Exome sequencing	80	M	37	positive	MSS	positive	body	moderate	intestinal	IIIB	T3	2	0	palliative		1	Alive or censored	1
pfg019T	Exome sequencing	90	M	57	negative	MSI	positive	body	moderate	intestinal	IB	T2b	0	0	curative	Recur	69	Died of disease	72
pfg020T	Exome sequencing	90	M	60	negative	MSS	negative	cardia	poor	intestinal	IV	T4	3	0	palliative		21	Died of disease	28
pfg021T	Exome sequencing	80	M	83	negative	MSS	positive	body	moderate	intestinal	IIIA	T3	1	0	curative	Recur	25	Died of disease	29
pfg022T	Exome sequencing	90	M	61	positive	MSS	negative	body	poor	intestinal	IIIB	T3	2	0	palliative		11	Died of disease	11
pfg024T	Exome sequencing	85	F	77	negative	MSS	negative	cardia	moderate	intestinal	IV	T2a	1	1	palliative		17	Died of disease	17
pfg025T	Exome sequencing	95	M	64	negative	MSS	negative	antrum	poor	intestinal	IIIA	T3	1	0	palliative		9	Died of disease	16
pfg029T	Exome sequencing	80	M	64	negative	MSS	positive	antrum	moderate	intestinal	II	T2b	1	0	curative	No recurrence	129	Alive or censored	129
pfg036T	Validation	85	M	81	positive	MSS	positive	body	poor	intestinal	IIIB	T3	2	0	curative	Recur	5	Died of disease	5
pfg062T	Validation	90	F	75	positive	MSS	positive	body	poor	intestinal	IV	T4	2	0	curative	No recurrence	3	Alive or censored	3
pfg088T	Validation	85	M	56	positive	MSS	positive	cardia	poor	intestinal	IV	T4	2	0	curative	Recur	10	Died of disease	16
pfg043T	Validation	80	M	76	positive	MSS	negative	cardia	poor	intestinal	IIIA	T3	1	0	palliative		120	Alive or censored	120
SX125T	Validation	90	M	71	positive	MSS	positive	cardia	moderate	intestinal	IV	T4	1	0	palliative		1	Alive or censored	1
SX318T	Validation	90	M	57	positive	MSS	positive	cardia	poor	intestinal	IV	T3	3	1	palliative		8	Died of disease	15
pfg037T	Validation	50	M	77	negative	MSS	negative	cardia	moderate	intestinal	IV	T4	1	1	palliative		4	Died of disease	8
pfg046T	Validation	60	F	77	negative	MSS	negative	body	poor	intestinal	IV	T4	1	0	palliative		10	Died of disease	12
pfg078T	Validation	95	M	71	negative	MSS	negative	cardia	moderate	intestinal	IB	T2b	0	0	curative	No recurrence	111	Alive or censored	111
pfg040T	Validation	90	F	81	negative	MSI	negative	antrum	moderate	intestinal	II	T2b	1	0	palliative		37	Alive or censored	37
pfg051T	Validation	90	F	75	negative	MSI	negative	antrum	moderate	intestinal	II	T3	0	0	palliative		8	Died of disease	8
pfg067T	Validation	90	F	72	negative	MSI	negative	body	moderate	intestinal	II	T2a	1	0	curative	No recurrence	138	Alive or censored	138
pfg068T	Validation	95	F	74	negative	MSI	negative	antrum	poor	intestinal	IIIA	T2b	2	0	palliative		7	Died of disease	13
SX186T	Validation	90	M	72	negative	MSI	positive	antrum	moderate	intestinal	II	T2b	1	0	curative	No recurrence	137	Alive or censored	137
SX256T	Validation	80	F	81	negative	MSI	negative	antrum	poor	intestinal	IB	T2b	0	0	curative	No recurrence	1	Alive or censored	1
SX259T	Validation	90	M	82	negative	MSI	positive	antrum	moderate	intestinal	IIIA	T4	0	0	curative	No recurrence	87	Alive or censored	87
pfg012T	Validation	90	F	73	negative	MSI	positive	antrum	poor	indeterminate	IIIA	T3	1	0	curative	Recur	15	Died of disease	20
pfg048T	Validation	90	F	72	negative	MSI	positive	antrum	poor	intestinal	IIIA	T3	1	0	palliative		10	Died of disease	16
pfg057T	Validation	90	F	41	negative	MSI	negative	cardia	poor	intestinal	IV	T2b	2	1	palliative		2	Died of disease	3
pfg070T	Validation	75	M	74	negative	MSI	positive	antrum	well	intestinal	IIIA	T3	1	0	palliative		6	Died of disease	6
pfg083T	Validation	80	F	62	negative	MSI	positive	antrum	poor	mixed	II	T3	0	0	curative	No recurrence	91	Alive or censored	91
SX067T	Validation	80	M	79	negative	MSI	negative	body	poor	intestinal	II	T2a	1	0	curative	No recurrence	52	Alive or censored	52

SX100T	Validation	95	F	79	negative	MSI	positive	antrum	moderate	intestinal	II	T2a	1	0	curative	No recurrence	2	Alive or censored	2
pfg060T	Validation	90	F	83	negative	MSS	positive	cardia	moderate	intestinal	IIIA	T2b	2	0	palliative		10	Died of disease	14
pfg090T	Validation	90	M	72	positive	MSS	negative	antrum	poor	intestinal	IB	T2a	0	0	curative	No recurrence	84	Alive or censored	84
SX115T	Validation	90	M	68	positive	MSS	positive	body	poor	intestinal	IV	T3	3	1	palliative		6	Died of disease	10
SX165T	Validation	95	M	46	positive	MSS	positive	body	moderate	intestinal	IV	T3	3	0	curative	Recur	66	Died of disease	77
SX388T	Validation	90	F	77	positive	MSS	positive	body	poor	intestinal	IB	T2a	0	0	curative	No recurrence	60	Alive or censored	105
pfg061T	Validation	70	M	77	negative	MSI	positive	antrum	moderate	intestinal	IB	T2b	0	0	curative	No recurrence	139	Alive or censored	139
pfg049T	Validation	90	F	80	negative	MSI	negative	antrum	moderate	intestinal	II	T2b	1	0	curative	No recurrence	58	Alive or censored	58
pfg071T	Validation	95	F	75	negative	MSI	negative	cardia	poor	intestinal	II	T2b	1	0	curative	No recurrence	21	Alive or censored	21
pfg077T	Validation	70	M	76	negative	MSI	positive	cardia	poor	mixed	IIIA	T3	1	0	palliative		10	Died of disease	10
SX041T	Validation	50	F	79	negative	MSI	positive	antrum	poor	mixed	IB	T2a	0	0	curative	No recurrence	58	Alive or censored	58
SX190T	Validation	60	M	51	positive	MSS	negative	body	poor	mixed	IV	T4	2	0	curative	No recurrence	8	Alive or censored	8
SX394T	Validation	70	M	70	positive	MSS	negative	body	poor	intestinal	II	T2a	1	0	curative	No recurrence	104	Alive or censored	104
pfg013T	Validation	80	M	60	negative	MSS	positive	cardia	moderate	intestinal	IV	T4	3	0	palliative		9	Died of disease	20
pfg023T	Validation	95	F	65	negative	MSS	negative	body	poor	mixed	IIIB	T3	2	0	palliative		7	Died of disease	7
pfg026T	Validation	60	M	58	negative	MSS	positive	body	moderate	intestinal	IIIA	T3	1	0	curative	Recur	32	Died of disease	37
pfg027T	Validation	80	M	62	negative	MSS	positive	antrum	moderate	intestinal	IIIA	T4	0	0	curative	Recur	30	Died of disease	41
pfg028T	Validation	85	M	75	negative	MSS	negative	cardia	poor	intestinal	IIIA	T3	1	0	curative	No recurrence	145	Alive or censored	145
pfg030T	Validation	80	F	74	negative	MSS	negative	antrum	moderate	intestinal	IIIA	T2b	2	0	palliative		8	Died of disease	12
pfg032T	Validation	90	M	84	negative	MSS	positive	antrum	poor	intestinal	IV	T3	3	0	palliative		23	Died of disease	23
pfg033T	Validation	80	M	70	negative	MSS	positive	antrum	poor	diffuse (signet ring cell)	II	T3	0	0	curative	No recurrence	71	Alive or censored	71
pfg034T	Validation	85	M	63	negative	MSS	negative	antrum	poor	mixed	IV	T2a	3	0	palliative		18	Died of disease	21
pfg035T	Validation	70	M	74	negative	MSS	positive	cardia	moderate	intestinal	IV	T4	1	0	palliative		13	Died of disease	20
pfg038T	Validation	90	M	67	negative	MSS	negative	antrum	moderate	intestinal	IV	T4	2	0	palliative		6	Died of disease	6
pfg041T	Validation	70	M	76	negative	MSS	positive	cardia	poor	mixed	IIIB	T3	2	0	palliative		7	Died of disease	7
pfg042T	Validation	80	F	74	negative	MSS	positive	antrum	moderate	intestinal	IV	T3	1	1	palliative		44	Died of disease	44
pfg044T	Validation	80	F	69	negative	MSS	negative	cardia	moderate	intestinal	IIIA	T3	1	0	palliative		4	Died of disease	5
pfg045T	Validation	50	F	74	negative	MSS	negative	antrum	poor	diffuse	IIIA	T3	1	0	curative	Recur	3	Died of disease	4
pfg047T	Validation	95	M	81	negative	MSS	positive	cardia	poor	intestinal	IV	T3	1	1	palliative		6	Died of disease	6
pfg050T	Validation	80	M	61	negative	MSS	negative	antrum	poor	diffuse	IIIB	T3	2	0	palliative		4	Died of disease	5
pfg052T	Validation	80	M	70	negative	MSS	positive	body	moderate	intestinal	IIIA	T3	1	0	curative	Recur	11	Died of disease	14
pfg053T	Validation	90	M	80	negative	MSS	negative	cardia	poor	intestinal	IB	T2b	0	0	curative	Recur	42	Died of disease	42
pfg054T	Validation	90	M	72	negative	MSS	positive	antrum	moderate	intestinal	IV	T4	2	0	palliative		32	Died of disease	42
pfg055T	Validation	70	M	73	negative	MSS	positive	antrum	moderate	intestinal	IIIB	T3	2	0	curative	Recur	18	Died of disease	29
pfg056T	Validation	85	M	73	negative	MSS	positive	cardia	moderate	intestinal	IB	T2b	0	0	curative	No recurrence	111	Alive or censored	111
pfg058T	Validation	80	M	59	negative	MSS	positive	antrum	moderate	intestinal	IIIB	T3	2	0	palliative		5	Died of disease	10
pfg063T	Validation	80	M	61	negative	MSS	positive	antrum	moderate	intestinal	IIIA	T3	1	0	curative	Recur	13	Died of disease	31
pfg064T	Validation	70	F	84	negative	MSS	negative	body	poor	diffuse	IV	T3	3	0	curative	No recurrence	9	Alive or censored	9
pfg065T	Validation	60	M	84	negative	MSS	positive	cardia	moderate	intestinal	IIIA	T3	1	0	curative	Recur	7	Died of disease	11
pfg066T	Validation	70	M	51	negative	MSS	positive	antrum	moderate	intestinal	II	T2b	1	0	curative	No recurrence	138	Alive or censored	138
pfg069T	Validation	80	F	72	negative	MSS	negative	cardia	poor	intestinal	IIIA	T3	1	0	palliative		3	Died of disease	4
pfg072T	Validation	90	M	78	negative	MSS	positive	antrum	moderate	intestinal	IV	T3	2	1	palliative		19	Died of disease	26
pfg073T	Validation	60	F	63	negative	MSS	negative	antrum	poor	diffuse	IIIA	T2b	2	0	curative	Recur	48	Died of disease	61
pfg075T	Validation	90	M	60	negative	MSS	negative	cardia	moderate	intestinal	IB	T1	1	0	curative	No recurrence	117	Alive or censored	117
pfg076T	Validation	80	F	88	negative	MSS	positive	antrum	moderate	intestinal	II	T3	0	0	curative	No recurrence	89	Alive or censored	89
pfg079T	Validation	60	F	43	negative	MSS	negative	body	poor	diffuse	IIIA	T3	1	0	curative	Recur	9	Died of disease	17
pfg080T	Validation	80	M	78	negative	MSS	negative	body	poor	intestinal	IV	T3	3	1	palliative		3	Alive or censored	3
pfg081T	Validation	95	M	72	negative	MSS	positive	antrum	poor	intestinal	IV	T4	3	0	curative	Recur	2	Died of disease	9
pfg082T	Validation	90	M	72	negative	MSS	negative	cardia	poor	intestinal	IV	T3	2	1	palliative		2	Died of disease	4
pfg084T	Validation	90	M	53	negative	MSS	positive	cardia	moderate	intestinal	IB	T2b	0	0	curative	No recurrence	89	Alive or censored	89
pfg085T	Validation	90	M	85	negative	MSS	negative	body	moderate	intestinal	IB	T2b	0	0	curative	No recurrence	81	Alive or censored	81
pfg086T	Validation	80	M	74	negative	MSS	negative	cardia	poor	intestinal	II	T2b	1	0	palliative		8	Died of disease	20
pfg087T	Validation	90	M	35	negative	MSS	negative	cardia	poor	intestinal	IB	T2b	0	0	curative	No recurrence	91	Alive or censored	91
pfg089T	Validation	60	M	66	negative	MSS	positive	body	poor	mixed	IV	T3	2	1	palliative		57	Died of disease	65
pfg091T	Validation	70	F	67	negative	MSS	positive	diffuse	poor	diffuse (signet ring cell)	IV	T3	3	0	curative	Recur	8	Died of disease	14
pfg092T	Validation	95	M	72	negative	MSS	negative	cardia	moderate	intestinal	IA	T1	0	0	curative	No recurrence	82	Alive or censored	82
pfg093T	Validation	95	M	23	negative	MSS	negative	cardia	poor	intestinal	II	T2a	1	0	curative	No recurrence	94	Alive or censored	94

pfg094T	Validation	65	F	52	negative	MSS	negative	diffuse	poor	diffuse	IIIB	T3	2	0	curative	Recur	10	Died of disease	14
pfg095T	Validation	60	M	70	negative	MSS	negative	body	poor	intestinal	IV	T4	1	0	palliative		4	Alive or censored	4
pfg096T	Validation	70	M	78	negative	MSS	negative	antrum	moderate	intestinal	IIIA	T2b	2	0	curative	Recur	13	Died of disease	14
pfg097T	Validation	50	F	74	negative	MSS	negative	body	poor	diffuse	IV	T4	3	0	curative	Recur	1	Died of disease	14
pfg098T	Validation	65	F	39	negative	MSS	negative	antrum	poor	intestinal	IIIB	T3	2	0	curative	Recur	14	Died of disease	29
pfg099T	Validation	98	M	62	negative	MSS	positive	antrum	poor	intestinal	IV	T3	3	1	palliative		9	Died of disease	16
pfg100T	Validation	80	M	82	negative	MSS	negative	cardia	poor	intestinal	IV	T3	1	1	palliative		2	Died of disease	3
pfg101T	Validation	70	M	70	negative	MSS	positive	cardia	poor	diffuse	IIIB	T3	2	0	curative	Recur	8	Died of disease	9

[a]EBV status is defined by in-situ-hybridisation showing tumor cell specific expression of EBER

[b]MSI denotes high-level microsatellite instabiliy; MSS denotes either microsatellite stable or low-level microsatellite instability

[c]Helicobacter pylori status is defined by its presence in gastrectomy specimen or endoscopic gastric biopsy preceding surgery

[d]Staging is according to UICC TNM classifition 6th Edition 2002

Supplementary Table 2 Summary of whole exome sequencing statistics for the discovery cohort. Genotyping concordance is calculated on all markers on the array that has a GenCall score ≥0.15 and are covered in the sequencing data with more than the equivalent of three Q30 bases.

Sample	Phase	Tissue	Read length	Total raw sequencing yield (bp)	Total PF reads	% duplicate read pairs	Total uniquely mapped reads	% reads in targeted regions +/-100bp (min fraction: 0.5)	Mean coverage	Median coverage	>=1X	>=10X	>=30X	>=50X	>=100X	Genotyping concordance with Omni1-Quad BeadArray
pfg001N	1	N	75	8,770,644,000	115,403,280	4.3%	99,347,657	60.7%	80.9	67	93.2%	84.7%	71.2%	59.2%	33.8%	99.6%
pfg001T	1	T	75	7,959,616,000	104,731,680	5.0%	89,134,121	61.7%	75.6	60	98.1%	90.3%	72.9%	57.2%	26.1%	99.0%
pfg002N	1	N	75	7,672,940,000	100,959,840	4.9%	85,842,578	60.2%	68.7	57	96.2%	87.7%	69.9%	54.8%	25.3%	99.6%
pfg002T	1	T	75	7,783,392,000	102,413,040	4.4%	88,127,719	57.9%	68.3	55	98.3%	89.8%	70.3%	53.4%	23.3%	98.6%
pfg003N	1	N	75	8,074,218,000	106,239,840	4.0%	91,459,025	58.7%	71.8	60	94.6%	85.8%	69.8%	56.1%	28.0%	99.6%
pfg003T	1	T	75	9,671,160,000	127,252,080	4.7%	109,332,982	57.7%	84.3	67	97.9%	89.1%	72.2%	59.3%	33.9%	99.2%
pfg005N	1	N	75	8,329,064,000	109,593,120	4.4%	93,853,623	60.4%	76.0	65	94.1%	86.3%	72.5%	59.0%	30.2%	99.6%
pfg005T	1	T	75	8,149,656,000	107,232,480	4.7%	91,264,936	62.2%	76.6	63	98.1%	90.7%	73.9%	58.5%	28.4%	99.6%
pfg006N	1	N	75	8,196,594,000	107,850,000	4.5%	91,856,656	54.6%	65.9	54	97.6%	88.8%	69.5%	52.8%	22.3%	99.6%
pfg006T	1	T	75	8,441,936,000	95,930,880	4.6%	81,867,485	66.1%	72.2	56	98.0%	88.3%	69.5%	53.9%	25.6%	99.4%
pfg007N	1	N	75	9,089,964,000	119,604,720	6.0%	100,773,009	59.5%	80.8	64	97.1%	88.6%	71.6%	58.1%	32.2%	99.5%
pfg007T	1	T	75	7,841,860,000	96,562,320	5.5%	80,600,296	63.8%	68.3	52	97.9%	87.7%	67.5%	50.8%	22.7%	99.3%
pfg008N	1	N	75	8,768,710,000	115,377,840	7.6%	95,226,295	58.3%	72.7	57	96.0%	86.8%	68.7%	54.1%	26.7%	99.5%
pfg008T	1	T	75	7,956,486,000	104,690,640	7.6%	86,427,141	63.3%	73.3	59	97.1%	88.4%	70.6%	55.7%	27.0%	99.5%
pfg009N	1	N	75	8,660,716,000	113,956,800	4.7%	96,719,769	58.9%	76.0	61	97.9%	89.4%	71.5%	56.7%	28.6%	99.6%
pfg009T	1	T	75	8,185,490,000	107,703,840	8.4%	88,331,375	66.8%	78.7	60	97.8%	88.0%	69.7%	55.6%	29.6%	99.5%
pfg010N	1	N	75	8,632,244,000	113,582,400	3.9%	97,518,432	60.1%	79.1	67	96.7%	89.5%	73.9%	60.1%	31.7%	99.6%
pfg010T	1	T	75	9,658,924,000	127,091,040	5.3%	107,287,194	65.6%	94.7	76	96.1%	89.2%	76.0%	64.1%	38.0%	99.2%
pfg011N	1	N	75	15,970,628,000	210,140,160	4.2%	177,388,028	58.1%	138.3	110	98.5%	94.2%	83.5%	73.9%	53.5%	99.6%
pfg011T	1	T	75	17,426,842,000	229,300,560	6.5%	189,867,761	62.0%	158.1	125	98.4%	93.7%	84.0%	75.4%	57.5%	98.8%
pfg014G	2	N	75	13,539,154,000	178,146,752	11.1%	143,427,621	64.1%	157.5	114	98.3%	93.2%	82.1%	72.6%	54.0%	99.6%
pfg014T	2	T	75	13,908,096,000	183,001,216	7.5%	152,340,189	51.5%	124.6	93	98.5%	93.8%	81.7%	70.2%	46.8%	99.2%
pfg015G	2	N	75	11,843,810,000	155,839,552	10.1%	127,118,454	67.8%	144.8	109	98.2%	92.9%	81.4%	71.7%	52.8%	99.6%
pfg015T	2	T	75	15,171,146,000	199,620,288	4.8%	171,905,345	48.7%	123.6	93	98.2%	91.8%	78.2%	67.4%	47.4%	99.6%
pfg016G	2	N	75	14,220,888,000	187,116,928	6.3%	160,055,671	63.7%	158.6	114	98.5%	93.0%	81.2%	71.5%	53.7%	99.6%
pfg016T	2	T	75	13,004,148,000	171,107,200	4.3%	148,077,823	49.1%	113.0	85	98.4%	92.1%	78.0%	66.2%	44.2%	99.6%
pfg017G	2	N	75	12,055,940,000	158,630,784	10.7%	128,740,246	69.2%	147.3	108	98.3%	92.7%	80.8%	70.8%	52.2%	99.6%
pfg017T	2	T	75	16,696,310,000	219,688,256	8.6%	179,266,022	51.3%	141.4	112	98.5%	93.6%	82.5%	73.0%	53.8%	99.6%
pfg018G	2	N	75	14,584,252,000	191,898,048	5.8%	164,963,622	65.7%	169.7	120	98.4%	93.0%	81.7%	72.3%	55.1%	99.6%
pfg018T	2	T	75	13,718,686,000	180,508,992	5.4%	154,497,400	49.3%	117.5	93	98.6%	93.8%	81.7%	70.3%	47.0%	99.5%
pfg019G	2	N	75	14,501,480,000	190,808,896	10.7%	155,311,626	72.7%	160.2	118	98.5%	93.6%	82.4%	72.9%	55.0%	99.7%
pfg019T	2	T	75	16,055,844,000	211,261,120	9.1%	171,184,417	51.2%	135.4	109	98.6%	94.4%	84.0%	74.2%	53.1%	99.5%
pfg020G	2	N	75	15,094,106,000	198,606,656	3.7%	173,664,649	57.5%	149.6	107	98.6%	93.5%	81.8%	71.5%	51.9%	99.6%
pfg020T	2	T	75	15,469,006,000	203,539,520	6.2%	173,104,487	53.7%	134.7	107	98.7%	95.0%	85.4%	75.2%	52.4%	98.8%
pfg021G	2	N	75	14,424,806,000	189,800,064	4.6%	165,047,708	59.0%	152.5	105	98.2%	91.0%	77.6%	67.8%	51.1%	99.6%
pfg021T	2	T	75	13,476,904,000	177,327,680	6.1%	151,235,241	53.4%	126.5	96	98.5%	93.6%	81.7%	70.5%	48.1%	99.1%
pfg022G	2	N	75	16,002,760,000	210,562,560	9.1%	173,797,199	66.6%	189.3	139	98.6%	94.9%	85.8%	77.4%	60.2%	99.6%
pfg022T	2	T	75	17,033,108,000	224,119,808	10.8%	176,966,759	50.0%	146.6	109	98.4%	93.6%	82.8%	72.8%	52.8%	99.3%
pfg024G	2	N	75	13,088,068,000	172,211,392	7.8%	144,589,918	70.4%	155.7	121	98.6%	94.9%	85.3%	76.0%	56.5%	99.7%
pfg024T	2	T	75	13,412,996,000	176,486,784	5.3%	151,433,688	50.7%	111.3	85	98.4%	92.6%	79.1%	67.1%	43.8%	99.5%
pfg025G	2	N	75	15,706,254,000	206,661,248	3.8%	182,058,799	58.9%	162.9	115	98.6%	93.6%	81.8%	72.0%	54.0%	99.6%
pfg025T	2	T	75	12,932,678,000	170,166,784	5.9%	145,548,949	49.5%	115.8	88	98.4%	92.5%	79.1%	67.4%	45.0%	99.5%
pfg029G	2	N	75	15,524,396,000	204,268,352	5.7%	176,476,707	64.1%	167.6	119	98.5%	93.8%	82.6%	72.9%	55.1%	99.6%
pfg029T	2	T	75	15,039,294,000	197,885,440	6.4%	162,649,521	52.6%	130.0	102	98.8%	96.1%	87.5%	76.7%	50.8%	99.1%
Average				**12,085,118,500**	**158,520,020**	**6.3%**	**133,538,367**	**59.3%**	**115.8**	**89**	**97.8%**	**91.4%**	**77.6%**	**65.7%**	**42.3%**	**99.5%**

Supplementary Table 3 Sequenom validation results for whole exome sequencing predicted somatic SNVs and somatic indels. Sequenom MassARRAY genotyping assays were performed on 472 predicted somatic SNVs and 189 somatic indels in the corrsponding tumor and match normal samples. Due to the single base extension chemistry, assays for indels only confirm the presence or absense of the predicted indel, but cannot define the sequence within the indel (the actual indel sequences in the table under "Normal call" and "Tumor call" were added based on whole-exome sequencing results for clarity purpose). Position": 1-indexed position of the somatic SNV event, or 1-indexed start position of the indel event. "Mutation type": S - single nucleotide variants; D - small deletions; I - small insertions. "Normal match": "Y" if the Sequenom call agrees with the exome-seq prediction in the normal sample; "N" otherwise; "Tumor match": "Y" if the Sequenom call agrees with the exome-seq prediction in the tumor sample; "N" otherwise; "Somatic match": "Y" if Sequenom calls in both tumor and the matched normal agree with the exome-seq prediction; "N" otherwise. Please see Supplementary Methods for details on the the Sequenom assays and genotyping call method.

Whole-exome sequencing							Sequenom confirmation				
Sample	Chromosome	Posistion	Ref/Mut (exome-seq)	Mutatnt zygosity	Mutation type	Mutation length	Normal call (sequenom)	Tumor call (sequenom)	Normal match	Tumor match	Somatic match
pfg001T	15	26054233	CCCCAACTTCCCATCTTCTGCCTCA/-	het	D	25	CCCCAACTTCCCATCTTCTGCCTCA	CCCCAACTTCCCATCTTCTGCCTCA/-	Y	Y	Y
pfg002T	18	46838582	TT/-	het	D	2	TT	TT/-	Y	Y	Y
pfg003T	14	52411670	A/-	het	D	1	A	A/-	Y	Y	Y
pfg006T	2	227868309	ACATTTTCAAGGAGAGAAA/-	het	D	19	ACATTTTCAAGGAGAGAAA	ACATTTTCAAGGAGAGAAA/-	Y	Y	Y
pfg006T	5	112203650	C/-	het	D	1	C	C/-	Y	Y	Y
pfg006T	6	149316689	T/-	het	D	1	T	T/-	Y	Y	Y
pfg006T	7	32994706	A/-	het	D	1	A	A/-	Y	Y	Y
pfg006T	7	136958780	ACCAGG/-	het	D	6	ACCAGG	ACCAGG/-	Y	Y	Y
pfg006T	18	21869113	A/-	het	D	1	A	A/-	Y	Y	Y
pfg007T	4	124038270	CA/-	het	D	2	CA	CA/-	Y	Y	Y
pfg007T	13	35308281	GAAAGAATCATTAATAC/-	het	D	17	GAAAGAATCATTAATAC	GAAAGAATCATTAATAC/-	Y	Y	Y
pfg008T	1	7996842	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	1	25539617	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	1	26965338	G/-	het	D	1	G	G/-	Y	Y	Y
pfg008T	1	62863524	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	1	86013964	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	1	116923874	C/-	het	D	1	C	C/-	Y	Y	Y
pfg008T	1	117840984	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	1	152562404	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	1	198860665	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	1	199093643	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	2	42525179	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	2	43900583	C/-	het	D	1	C	C/-	Y	Y	Y
pfg008T	2	95178762	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	2	118295255	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	2	170514775	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	2	201145249	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	2	202327632	G/-	het	D	1	G	G/-	Y	Y	Y
pfg008T	2	203469082	CCTCTACAGCAGTGTAGGG/-	het	D	19	CCTCTACAGCAGTGTAGGG	CCTCTACAGCAGTGTAGGG/-	Y	Y	Y
pfg008T	2	205874543	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	2	210595925	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	2	230867269	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	3	38924445	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	3	56642137	TCT/-	het	D	3	TCT	TCT/-	Y	Y	Y
pfg008T	3	113782546	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	3	130638238	T/-	het	D	1	T/-	T/-	N	Y	N
pfg008T	3	134788171	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	3	143513333	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	3	180952846	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	3	182148922	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	3	197683519	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	4	22055831	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	4	70655434	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	4	126617972	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	4	128784160	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	4	155939830	G/-	het	D	1	G	G/-	Y	Y	Y
pfg008T	4	160487596	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	5	72235301	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	5	80006671	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	5	94843971	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	5	98234309	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	5	140482671	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	5	176640436	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	10665228	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	38924418	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	39688971	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	74408311	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	74551833	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	6	131522969	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	139139023	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	146912403	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	6	162126924	G/-	het	D	1	G	G/-	Y	Y	Y

pfg008T	7	77261396	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	7	86816416	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	7	91708362	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	7	104910046	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	7	115381893	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	7	136351148	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	7	142322616	C/-	het	D	1	C	C/-	Y	Y	Y
pfg008T	8	28725769	GT/-	het	D	2	GT	GT/-	Y	Y	Y
pfg008T	8	38306379	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	8	49046532	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	8	68254679	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	8	103358525	T/-	het	D	1	T/-	T/-	N	Y	N
pfg008T	8	113554049	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	8	124734062	C/-	het	D	1	C	C/-	Y	Y	Y
pfg008T	9	111210462	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	9	116706227	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	9	130059211	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	10	7860860	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	10	13736458	G/-	het	D	1	G	G/-	Y	Y	Y
pfg008T	10	25926983	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	10	91488215	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	10	97436867	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	10	97437020	C/-	het	D	1	C	C/-	Y	Y	Y
pfg008T	10	112323484	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	11	31434487	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	11	32594096	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	11	93126562	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	11	93134358	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	11	101442485	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	11	110931201	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	11	124476392	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	12	19405785	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	12	54045459	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	12	57570840	CCAGCCCTAGA/-	het	D	11	CCAGCCCTAGA	CCAGCCCTAGA/-	Y	Y	Y
pfg008T	12	63920957	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	13	44431635	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	13	45441576	G/-	het	D	1	G	G/-	Y	Y	Y
pfg008T	13	50837973	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	14	49311030	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	14	49676411	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	14	54888308	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	14	57884204	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	14	63554140	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	14	63558457	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	14	91985193	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	15	19870388	G/-	het	D	1	G	G/-	Y	Y	Y
pfg008T	15	40530249	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	15	48010720	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	15	57351797	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	15	89111212	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	16	2753494	TCAC/-	het	D	4	TCAC	TCAC/-	Y	Y	Y
pfg008T	16	75913812	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	17	32654606	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	17	34819192	TACTGAAATAA/-	het	D	11	TACTGAAATAA	TACTGAAATAA/-	Y	Y	Y
pfg008T	17	52383015	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	18	17408748	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	18	37883560	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	18	55164174	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	19	7908975	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	19	62338199	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	19	62339269	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	21	18588619	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	21	36524696	C/-	het	D	1	C	C/-	Y	Y	Y
pfg008T	X	32376567	T/-	het	D	1	T	T/-	Y	Y	Y
pfg008T	X	83014703	A/-	het	D	1	A	A/-	Y	Y	Y
pfg008T	X	105952069	T/-	het	D	1	T	T/-	Y	Y	Y
pfg009T	1	118375862	TTC/-	het	D	3	TTC	TTC/-	Y	Y	Y
pfg009T	2	157890221	TTCCTCCTGAC/-	het	D	11	TTCCTCCTGAC	TTCCTCCTGAC/-	Y	Y	Y
pfg009T	2	201565901	CCAATACTATTCGA/-	het	D	14	CCAATACTATTCGA	CCAATACTATTCGA/-	Y	Y	Y
pfg009T	7	6475916	AATAAAAGAAAAAACACCAG/-	het	D	20	AATAAAAGAAAAAACACCAG	AATAAAAGAAAAAACACCAG/-	Y	Y	Y
pfg009T	7	24685403	TA/-	het	D	2	TA	TA/-	Y	Y	Y
pfg009T	14	49311035	AATC/-	het	D	4	AATC	AATC/-	Y	Y	Y
pfg009T	19	39402288	AG/-	het	D	2	AG	AG/-	Y	Y	Y
pfg009T	21	37444339	AAGT/-	het	D	4	AAGT	AAGT/-	Y	Y	Y
pfg010T	1	154163099	G/-	het	D	1	G	G/-	Y	Y	Y
pfg010T	3	2919711	A/-	het	D	1	A	A	Y	N	N
pfg010T	10	29796660	TG/-	het	D	2	TG	TG/-	Y	Y	Y
pfg010T	14	102872850	CATGAGGCGAATAAGCTG/-	het	D	18	CATGAGGCGAATAAGCTG	CATGAGGCGAATAAGCTG	Y	N	N
pfg010T	X	50394015	TCC/-	het	D	3	TCC	TCC/-	Y	Y	Y
pfg011T	11	125274521	AGA/-	het	D	3	AGA	AGA/-	Y	Y	Y
pfg011T	15	33061643	AT/-	het	D	2	AT	AT/-	Y	Y	Y
pfg001T	1	245835691	-/T	het	I	1	-	-/T	Y	Y	Y

pfg001T	14	49187735	-/A	het	I	1	-	-/A	Y	Y	Y
pfg003T	8	38110166	-/A	het	I	1	-	-/A	Y	Y	Y
pfg005T	12	6053020	-/A	het	I	1	-	-/A	Y	Y	Y
pfg007T	6	12229214	-/A	het	I	1	-	-/A	Y	Y	Y
pfg007T	7	50099857	-/A	het	I	1	-	-/A	Y	Y	Y
pfg007T	10	123799948	-/T	het	I	1	-	-/T	Y	Y	Y
pfg007T	14	72534787	-/AT	hom	I	2	-	-/AT	Y	Y	Y
pfg007T	X	76831035	-/G	het	I	1	-	-/G	Y	Y	Y
pfg008T	1	89043123	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	1	113438872	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	1	153222980	-/C	het	I	1	-	-/C	Y	Y	Y
pfg008T	1	240031085	-/C	het	I	1	-	-/C	Y	Y	Y
pfg008T	2	48455596	-/A	het	I	1	-	-/A	Y	Y	Y
pfg008T	2	131954695	-/G	het	I	1	-	-/G	Y	Y	Y
pfg008T	2	169429652	-/A	het	I	1	-	-/A	Y	Y	Y
pfg008T	4	56874260	-/A	het	I	1	-	-/A	Y	Y	Y
pfg008T	4	83567649	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	6	99962736	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	6	111321458	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	6	142441690	-/C	het	I	1	-	-/C	Y	Y	Y
pfg008T	8	134328061	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	10	50360764	-/C	het	I	1	-	-/C	Y	Y	Y
pfg008T	10	89707750	-/A	het	I	1	-	-/A	Y	Y	Y
pfg008T	10	102738565	-/GTCTT	het	I	5	-	-/GTCTT	Y	Y	Y
pfg008T	11	59331817	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	11	118559088	-/G	het	I	1	-	-/G	Y	Y	Y
pfg008T	12	50120823	-/A	het	I	1	-	-/A	Y	Y	Y
pfg008T	12	63103250	-/A	het	I	1	-	-/A	Y	Y	Y
pfg008T	16	14253801	-/C	het	I	1	-	-/C	Y	Y	Y
pfg008T	19	42336274	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	19	55554118	-/C	het	I	1	-	-/C	Y	Y	Y
pfg008T	21	31996526	-/T	het	I	1	-	-/T	Y	Y	Y
pfg008T	X	131988079	-/T	het	I	1	-	-/T	Y	Y	Y
pfg009T	10	70718465	-/G	het	I	1	-	-/G	Y	Y	Y
pfg009T	12	94783997	-/T	het	I	1	-	-/T	Y	Y	Y
pfg010T	1	115017252	-/T	het	I	1	-	-/T	Y	Y	Y
pfg010T	5	162842208	-/G	het	I	1	-	-/G	Y	Y	Y
pfg010T	7	82957434	-/GCC	het	I	3	-	-/GCC	Y	Y	Y
pfg010T	16	24280245	-/A	het	I	1	-	-/A	Y	Y	Y
pfg011T	4	170273976	-/T	het	I	1	-	-/T	Y	Y	Y
pfg011T	5	111604317	-/A	het	I	1	-	-/A	Y	Y	Y
pfg011T	6	111759659	-/GCGT	het	I	4	-	-/GCGT	Y	Y	Y
pfg011T	6	133177863	-/C	het	I	1	-	-/C	Y	Y	Y
pfg011T	10	22902417	-/A	het	I	1	-	-/A	Y	Y	Y
pfg011T	18	30649871	-/A	het	I	1	-	-/A	Y	Y	Y
pfg001T	1	24866005	T/A	het	S	1	T	T	Y	N	N
pfg001T	1	68669602	G/C	het	S	1	G	GC	Y	Y	Y
pfg001T	1	86813842	A/C	het	S	1	A	CA	Y	Y	Y
pfg001T	1	115028497	T/G	het	S	1	T	GT	Y	Y	Y
pfg001T	1	246268718	T/A	het	S	1	T	AT	Y	Y	Y
pfg001T	2	73858982	C/G	het	S	1	C	CG	Y	Y	Y
pfg001T	2	130629248	A/G	het	S	1	GA	GA	N	Y	N
pfg001T	2	192572086	T/G	het	S	1	T	GT	Y	Y	Y
pfg001T	2	203819826	C/G	het	S	1	C	GC	Y	Y	Y
pfg001T	3	103663834	A/T	het	S	1	A	TA	Y	Y	Y
pfg001T	3	152633244	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	4	5718145	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	4	89602480	A/G	het	S	1	A	GA	Y	Y	Y
pfg001T	5	54676116	G/C	het	S	1	G	CG	Y	Y	Y
pfg001T	5	128984304	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	5	140553133	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	5	140994570	G/A	het	S	1	G	AG	Y	Y	Y
pfg001T	5	176495526	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	6	50919016	A/G	het	S	1	A	AG	Y	Y	Y
pfg001T	7	92808816	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	8	26557429	G/T	het	S	1	G	GT	Y	Y	Y
pfg001T	8	113392467	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	9	16426138	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	9	74721841	C/A	het	S	1	C	C	Y	N	N
pfg001T	9	74757693	T/G	het	S	1	T	T	Y	N	N
pfg001T	9	132952773	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	10	61501858	T/A	het	S	1	T	AT	Y	Y	Y
pfg001T	10	84108522	C/T	het	S	1	C	TC	Y	Y	Y
pfg001T	10	127412194	C/T	het	S	1	C	TC	Y	Y	Y
pfg001T	11	46706282	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	11	55078681	T/A	het	S	1	T	TA	Y	Y	Y
pfg001T	11	65948880	G/C	het	S	1	G	GC	Y	Y	Y
pfg001T	11	73313316	A/T	het	S	1	A	AT	Y	Y	Y
pfg001T	11	123945308	T/G	het	S	1	T	GT	Y	Y	Y
pfg001T	12	6505869	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	12	12993782	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	13	77372663	A/C	het	S	1	A	CA	Y	Y	Y

pfg001T	13	100593470	G/A	het	S	1	G	G	Y	N	N
pfg001T	14	44044386	C/A	het	S	1	C	CA	Y	Y	Y
pfg001T	14	93826180	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	14	94034196	T/C	het	S	1	T	CT	Y	Y	Y
pfg001T	15	32871719	A/C	het	S	1	A	CA	Y	Y	Y
pfg001T	15	32872918	A/C	het	S	1	A	A	Y	N	N
pfg001T	15	33617864	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	15	40529403	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	15	48331928	G/A	het	S	1	G	AG	Y	Y	Y
pfg001T	15	54510898	C/G	het	S	1	C	GC	Y	Y	Y
pfg001T	15	60766249	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	15	88147933	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	17	36229326	T/C	het	S	1	T	TC	Y	Y	Y
pfg001T	17	36273542	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	18	7878344	C/T	het	S	1	C	CT	Y	Y	Y
pfg001T	18	27308247	G/A	het	S	1	G	GA	Y	Y	Y
pfg001T	18	59805274	T/G	het	S	1	T	GT	Y	Y	Y
pfg001T	18	68356992	A/G	het	S	1	A	A	Y	N	N
pfg001T	20	4653396	G/A	het	S	1	G	AG	Y	Y	Y
pfg001T	20	19625476	C/T	het	S	1	C	TC	Y	Y	Y
pfg001T	20	51304500	G/T	het	S	1	G	GT	Y	Y	Y
pfg002T	1	218901836	G/A	het	S	1	G	AG	Y	Y	Y
pfg002T	1	235116231	C/A	het	S	1	C	CA	Y	Y	Y
pfg002T	2	165891626	G/C	het	S	1	G	GC	Y	Y	Y
pfg002T	2	212697780	A/T	het	S	1	A	AT	Y	Y	Y
pfg002T	3	173648352	G/A	het	S	1	G	AG	Y	Y	Y
pfg002T	4	110968636	C/T	het	S	1	C	CT	Y	Y	Y
pfg002T	5	3652422	C/T	het	S	1	C	CT	Y	Y	Y
pfg002T	5	101862432	C/T	het	S	1	C	TC	Y	Y	Y
pfg002T	6	43596096	G/A	het	S	1	G	AG	Y	Y	Y
pfg002T	7	134544962	C/T	het	S	1	C	CT	Y	Y	Y
pfg002T	8	19407271	C/T	het	S	1	C	CT	Y	Y	Y
pfg002T	9	703252	G/A	het	S	1	G	GA	Y	Y	Y
pfg002T	10	44806460	A/G	het	S	1	A	GA	Y	Y	Y
pfg002T	10	98172451	C/T	het	S	1	C	CT	Y	Y	Y
pfg002T	11	20342482	G/C	het	S	1	G	CG	Y	Y	Y
pfg002T	11	129250455	G/A	het	S	1	G	AG	Y	Y	Y
pfg002T	12	25289551	C/T	het	S	1	C	CT	Y	Y	Y
pfg002T	14	71124923	C/A	het	S	1	C	CA	Y	Y	Y
pfg002T	19	36461425	C/T	het	S	1	C	TC	Y	Y	Y
pfg002T	19	49628346	G/A	het	S	1	G	AG	Y	Y	Y
pfg002T	19	54384132	G/A	het	S	1	G	AG	Y	Y	Y
pfg002T	22	40853788	G/A	het	S	1	G	AG	Y	Y	Y
pfg002T	22	49248600	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	1	26967035	T/A	het	S	1	T	TA	Y	Y	Y
pfg003T	1	75457603	G/A	het	S	1	G	GA	Y	Y	Y
pfg003T	2	170136683	T/G	het	S	1	T	GT	Y	Y	Y
pfg003T	3	49703637	G/A	het	S	1	G	GA	Y	Y	Y
pfg003T	3	140560531	G/A	het	S	1	G	GA	Y	Y	Y
pfg003T	5	137763520	A/G	het	S	1	A	GA	Y	Y	Y
pfg003T	5	179664387	C/T	het	S	1	C	TC	Y	Y	Y
pfg003T	6	46324548	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	6	168210307	G/A	het	S	1	G	GA	Y	Y	Y
pfg003T	8	133896275	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	9	136826792	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	12	6497095	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	12	39617738	G/A	het	S	1	G	GA	Y	Y	Y
pfg003T	12	100076160	T/C	het	S	1	T	CT	Y	Y	Y
pfg003T	12	121780190	T/G	het	S	1	T	GT	Y	Y	Y
pfg003T	16	54076809	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	16	70717373	C/G	het	S	1	C	GC	Y	Y	Y
pfg003T	19	12876462	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	19	46546089	C/T	het	S	1	C	CT	Y	Y	Y
pfg003T	19	62794448	A/G	het	S	1	A	GA	Y	Y	Y
pfg005T	1	173621829	G/T	het	S	1	G	GT	Y	Y	Y
pfg005T	3	44573763	C/T	het	S	1	C	CT	Y	Y	Y
pfg005T	14	88139080	T/G	het	S	1	T	GT	Y	Y	Y
pfg005T	18	347401	T/C	het	S	1	T	T	Y	N	N
pfg006T	1	47284779	G/A	het	S	1	G	GA	Y	Y	Y
pfg006T	1	62836105	G/C	het	S	1	G	GC	Y	Y	Y
pfg006T	1	154548559	T/G	het	S	1	T	GT	Y	Y	Y
pfg006T	1	154981418	G/T	het	S	1	G	GT	Y	Y	Y
pfg006T	1	199284854	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	1	212236906	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	1	213326403	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	1	224192028	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	2	26852035	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	2	27444807	T/C	het	S	1	T	CT	Y	Y	Y
pfg006T	2	86134784	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	2	108452908	C/G	het	S	1	C	CG	Y	Y	Y
pfg006T	2	141175872	G/A	het	S	1	G	GA	Y	Y	Y
pfg006T	2	202278430	C/T	het	S	1	C	CT	Y	Y	Y

pfg006T	3	3091541	T/G	het	S	1	T	GT	Y	Y	Y
pfg006T	3	10392120	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	3	58055719	G/T	het	S	1	G	GT	Y	Y	Y
pfg006T	3	102660586	A/T	het	S	1	A	AT	Y	Y	Y
pfg006T	3	141377505	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	3	152645406	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	5	49759814	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	5	115378965	T/A	het	S	1	T	T	Y	N	N
pfg006T	5	115378966	C/T	het	S	1	C	C	Y	N	N
pfg006T	6	26358734	G/A	het	S	1	G	GA	Y	Y	Y
pfg006T	6	62815836	A/G	het	S	1	A	GA	Y	Y	Y
pfg006T	6	127838595	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	6	129755965	C/T	het	S	1	C	TC	Y	Y	Y
pfg006T	7	3969003	G/A	het	S	1	G	GA	Y	Y	Y
pfg006T	7	63804459	C/T	het	S	1	C	TC	Y	Y	Y
pfg006T	7	121542412	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	7	136350321	G/A	het	S	1	G	AG	Y	Y	Y
pfg006T	7	146723713	G/A	het	S	1	G	GA	Y	Y	Y
pfg006T	8	56867860	G/T	het	S	1	G	GT	Y	Y	Y
pfg006T	8	79790027	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	8	113668638	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	8	133251794	A/C	het	S	1	A	CA	Y	Y	Y
pfg006T	9	125173679	G/A	het	S	1	G	AG	Y	Y	Y
pfg006T	10	35535908	G/A	het	S	1	G	GA	Y	Y	Y
pfg006T	10	50354333	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	10	89682878	A/T	het	S	1	A	TA	Y	Y	Y
pfg006T	11	4551283	C/T	het	S	1	C	TC	Y	Y	Y
pfg006T	11	6773085	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	11	117848591	C/T	het	S	1	C	TC	Y	Y	Y
pfg006T	12	7186774	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	12	62775176	C/A	het	S	1	C	CA	Y	Y	Y
pfg006T	13	110085830	G/C	het	S	1	G	CG	Y	Y	Y
pfg006T	14	22953150	G/A	het	S	1	G	AG	Y	Y	Y
pfg006T	14	38718317	A/G	het	S	1	A	GA	Y	Y	Y
pfg006T	15	56254512	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	16	65871678	A/G	het	S	1	A	GA	Y	Y	Y
pfg006T	17	3161260	C/T	het	S	1	C	TC	Y	Y	Y
pfg006T	17	27672133	G/A	het	S	1	G	AG	Y	Y	Y
pfg006T	17	44938763	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	17	75628456	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	19	14737199	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	20	59942553	C/T	het	S	1	C	CT	Y	Y	Y
pfg006T	X	34872165	G/A	het	S	1	G	AG	Y	Y	Y
pfg006T	X	65340030	G/T	het	S	1	G	GT	Y	Y	Y
pfg006T	X	83015083	T/G	het	S	1	T	GT	Y	Y	Y
pfg006T	X	123606738	G/A	het	S	1	G	AG	Y	Y	Y
pfg007T	1	11777770	G/A	het	S	1	G	AG	Y	Y	Y
pfg007T	1	28473214	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	1	110268185	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	1	111664527	C/A	het	S	1	C	CA	Y	Y	Y
pfg007T	1	112800109	G/T	het	S	1	G	GT	Y	Y	Y
pfg007T	1	115630311	G/A	het	S	1	G	AG	Y	Y	Y
pfg007T	1	150550811	T/C	het	S	1	T	CT	Y	Y	Y
pfg007T	1	156565367	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	1	167776472	A/G	het	S	1	A	GA	Y	Y	Y
pfg007T	1	184373613	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	1	226649173	G/C	het	S	1	G	CG	Y	Y	Y
pfg007T	1	244851468	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	2	99003288	T/G	het	S	1	T	GT	Y	Y	Y
pfg007T	2	102323697	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	2	120823163	G/A	het	S	1	G	GA	Y	Y	Y
pfg007T	2	148942850	G/T	het	S	1	G	GT	Y	Y	Y
pfg007T	2	215573889	A/G	het	S	1	A	GA	Y	Y	Y
pfg007T	3	27191024	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	3	95262589	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	4	6749541	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	4	64957517	G/C	het	S	1	G	GC	Y	Y	Y
pfg007T	5	75633013	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	5	179128612	C/G	het	S	1	C	CG	Y	Y	Y
pfg007T	6	43263159	T/C	het	S	1	T	CT	Y	Y	Y
pfg007T	6	43279286	C/T	het	S	1	C	TC	Y	Y	Y
pfg007T	6	152776245	T/C	het	S	1	T	CT	Y	Y	Y
pfg007T	7	2241406	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	7	4057945	C/T	het	S	1	C	TC	Y	Y	Y
pfg007T	7	27153702	G/A	het	S	1	G	GA	Y	Y	Y
pfg007T	7	36901033	G/A	het	S	1	G	AG	Y	Y	Y
pfg007T	8	113387427	T/A	het	S	1	T	TA	Y	Y	Y
pfg007T	8	139230098	T/A	het	S	1	T	AT	Y	Y	Y
pfg007T	8	144883181	G/A	het	S	1	G	GA	Y	Y	Y
pfg007T	9	34986541	G/A	het	S	1	G	AG	Y	Y	Y
pfg007T	10	20493451	G/T	het	S	1	G	GT	Y	Y	Y
pfg007T	11	120521563	T/A	het	S	1	T	AT	Y	Y	Y

pfg007T	12	13417608	C/T	het	S	1	C	TC	Y	Y	Y
pfg007T	12	20657673	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	14	19833406	G/C	het	S	1	G	G	Y	N	N
pfg007T	14	23793764	C/T	het	S	1	C	TC	Y	Y	Y
pfg007T	14	25987156	G/C	het	S	1	G	CG	Y	Y	Y
pfg007T	14	73893266	G/T	het	S	1	G	GT	Y	Y	Y
pfg007T	15	66276876	C/T	het	S	1	C	TC	Y	Y	Y
pfg007T	16	180565	C/G	het	S	1	C	GC	Y	Y	Y
pfg007T	17	71820651	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	18	29180289	C/T	het	S	1	C	TC	Y	Y	Y
pfg007T	19	8507417	C/T	het	S	1	C	CT	Y	Y	Y
pfg007T	19	63760252	C/G	het	S	1	C	CG	Y	Y	Y
pfg007T	21	31517614	C/T	het	S	1	C	TC	Y	Y	Y
pfg007T	22	22416573	T/A	het	S	1	T	AT	Y	Y	Y
pfg007T	X	105755165	A/G	het	S	1	A	GA	Y	Y	Y
pfg007T	X	144717017	G/T	het	S	1	G	GT	Y	Y	Y
pfg008T	1	75449800	A/T	het	S	1	A	TA	Y	Y	Y
pfg008T	1	156850708	A/G	het	S	1	A	AG	Y	Y	Y
pfg008T	1	156870994	A/G	het	S	1	A	GA	Y	Y	Y
pfg008T	1	212236722	T/C	het	S	1	T	TC	Y	Y	Y
pfg008T	2	71763226	C/A	het	S	1	C	CA	Y	Y	Y
pfg008T	3	58107762	T/C	het	S	1	T	CT	Y	Y	Y
pfg008T	3	151860639	A/G	het	S	1	A	AG	Y	Y	Y
pfg008T	4	57035263	A/G	het	S	1	A	GA	Y	Y	Y
pfg008T	5	115364721	G/A	het	S	1	G	GA	Y	Y	Y
pfg008T	5	129067937	A/G	het	S	1	A	AG	Y	Y	Y
pfg008T	7	121520425	G/T	het	S	1	G	GT	Y	Y	Y
pfg008T	7	121731189	G/A	het	S	1	G	GA	Y	Y	Y
pfg008T	8	18706539	A/G	het	S	1	A	GA	Y	Y	Y
pfg008T	8	133898876	A/G	het	S	1	A	GA	Y	Y	Y
pfg008T	9	34986664	G/A	het	S	1	G	GA	Y	Y	Y
pfg008T	9	118037497	A/G	het	S	1	A	GA	Y	Y	Y
pfg008T	10	127416545	A/G	het	S	1	A	GA	Y	Y	Y
pfg008T	10	127416552	A/G	het	S	1	A	AG	Y	Y	Y
pfg008T	11	66373992	G/A	het	S	1	G	AG	Y	Y	Y
pfg008T	11	111287465	A/C	het	S	1	A	CA	Y	Y	Y
pfg008T	12	76925278	G/A	het	S	1	G	AG	Y	Y	Y
pfg008T	14	63722891	G/A	het	S	1	G	AG	Y	Y	Y
pfg008T	15	47079368	A/G	het	S	1	A	AG	Y	Y	Y
pfg008T	16	20898196	C/T	het	S	1	C	CT	Y	Y	Y
pfg008T	18	39104616	A/G	het	S	1	A	AG	Y	Y	Y
pfg008T	18	39108280	A/C	het	S	1	A	CA	Y	Y	Y
pfg008T	19	14736412	T/C	het	S	1	T	CT	Y	Y	Y
pfg008T	19	62793820	A/C	het	S	1	A	CA	Y	Y	Y
pfg008T	20	20217454	A/T	het	S	1	A	AT	Y	Y	Y
pfg008T	20	59882332	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	1	59846090	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	1	149327390	C/G	het	S	1	C	GC	Y	Y	Y
pfg009T	1	154981491	C/G	het	S	1	C	CG	Y	Y	Y
pfg009T	1	178120470	A/T	het	S	1	A	AT	Y	Y	Y
pfg009T	1	180882550	G/C	het	S	1	G	CG	Y	Y	Y
pfg009T	1	233660732	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	1	245085669	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	2	8876304	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	2	53898594	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	2	71503605	T/A	het	S	1	T	TA	Y	Y	Y
pfg009T	2	96883014	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	2	150147000	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	2	170079721	A/C	het	S	1	A	CA	Y	Y	Y
pfg009T	2	196573812	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	2	228590627	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	3	1390376	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	3	41242304	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	3	43097497	G/A	het	S	1	G	AG	Y	Y	Y
pfg009T	3	47430697	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	3	52531903	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	3	52533177	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	3	151860601	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	3	170968107	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	3	175479822	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	3	185379484	A/G	het	S	1	A	AG	Y	Y	Y
pfg009T	4	35799320	T/C	het	S	1	T	TC	Y	Y	Y
pfg009T	4	46431934	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	4	77524521	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	5	21878110	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	5	140481864	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	5	141314954	C/T	het	S	1	C	TC	Y	Y	Y
pfg009T	5	147698347	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	5	176234061	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	6	43700266	A/C	het	S	1	A	CA	Y	Y	Y
pfg009T	6	56589017	T/A	het	S	1	T	TA	Y	Y	Y
pfg009T	6	73961181	A/G	het	S	1	A	GA	Y	Y	Y

pfg009T	6	74129218	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	6	82984483	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	7	56122787	A/T	het	S	1	A	TA	Y	Y	Y
pfg009T	7	94719086	G/A	het	S	1	G	AG	Y	Y	Y
pfg009T	7	112194684	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	7	127763328	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	7	139438108	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	8	90995622	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	9	115897413	G/C	het	S	1	G	GC	Y	Y	Y
pfg009T	9	116838216	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	9	118154827	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	10	16982777	G/A	het	S	1	G	AG	Y	Y	Y
pfg009T	10	23288007	G/C	het	S	1	G	CG	Y	Y	Y
pfg009T	10	50336918	A/G	het	S	1	A	GA	Y	Y	Y
pfg009T	10	61625917	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	10	75237374	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	10	115345472	A/T	het	S	1	A	AT	Y	Y	Y
pfg009T	11	40093449	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	11	46965433	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	11	55189393	T/A	het	S	1	T	AT	Y	Y	Y
pfg009T	11	56014624	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	11	66374328	T/G	het	S	1	T	GT	Y	Y	Y
pfg009T	11	118711383	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	11	123129506	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	11	123500213	G/C	het	S	1	G	GC	Y	Y	Y
pfg009T	12	4790986	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	12	11796691	C/G	het	S	1	C	CG	Y	Y	Y
pfg009T	12	51795576	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	12	53255178	T/A	het	S	1	T	AT	Y	Y	Y
pfg009T	12	122367745	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	12	132144535	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	13	45436024	A/G	het	S	1	A	GA	Y	Y	Y
pfg009T	13	101165938	A/C	het	S	1	A	CA	Y	Y	Y
pfg009T	14	50518500	T/C	het	S	1	T	CT	Y	Y	Y
pfg009T	14	63556509	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	15	38700527	A/T	het	S	1	A	AT	Y	Y	Y
pfg009T	15	56627861	G/C	het	S	1	G	CG	Y	Y	Y
pfg009T	15	60006794	T/G	het	S	1	T	GT	Y	Y	Y
pfg009T	15	73479533	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	15	90448712	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	15	98409419	T/A	het	S	1	T	AT	Y	Y	Y
pfg009T	16	666834	G/T	het	S	1	G	GT	Y	Y	Y
pfg009T	16	29913484	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	16	65340611	C/G	het	S	1	C	CG	Y	Y	Y
pfg009T	16	66419273	A/C	het	S	1	A	CA	Y	Y	Y
pfg009T	17	37373951	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	17	47065876	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	18	32039235	C/T	het	S	1	C	TC	Y	Y	Y
pfg009T	19	5555800	T/A	het	S	1	T	AT	Y	Y	Y
pfg009T	19	9932230	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	19	59536671	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	20	29428743	C/A	het	S	1	C	CA	Y	Y	Y
pfg009T	20	29517039	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	20	49834591	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	21	31517729	C/T	het	S	1	C	CT	Y	Y	Y
pfg009T	22	20365587	G/A	het	S	1	G	GA	Y	Y	Y
pfg009T	22	41610413	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	1	36069651	A/G	het	S	1	A	AG	Y	Y	Y
pfg010T	1	66599940	A/C	het	S	1	A	CA	Y	Y	Y
pfg010T	1	150222306	T/C	het	S	1	T	CT	Y	Y	Y
pfg010T	1	151903237	G/T	het	S	1	G	GT	Y	Y	Y
pfg010T	1	167963661	C/T	het	S	1	C	CT	Y	Y	Y
pfg010T	1	195150384	G/T	het	S	1	G	GT	Y	Y	Y
pfg010T	1	209600711	C/G	het	S	1	C	CG	Y	Y	Y
pfg010T	2	69203662	C/G	het	S	1	C	GC	Y	Y	Y
pfg010T	2	178947007	A/G	het	S	1	A	GA	Y	Y	Y
pfg010T	2	201185624	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	3	161577970	A/C	het	S	1	A	CA	Y	Y	Y
pfg010T	4	5746959	T/G	het	S	1	T	GT	Y	Y	Y
pfg010T	4	85969334	A/T	het	S	1	A	AT	Y	Y	Y
pfg010T	4	140407591	T/A	het	S	1	T	AT	Y	Y	Y
pfg010T	5	33685474	C/T	het	S	1	C	CT	Y	Y	Y
pfg010T	5	114576106	A/C	het	S	1	A	CA	Y	Y	Y
pfg010T	5	170273787	T/G	het	S	1	T	GT	Y	Y	Y
pfg010T	5	178438486	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	6	83867246	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	6	147629932	C/G	het	S	1	C	GC	Y	Y	Y
pfg010T	6	147629940	G/T	het	S	1	G	G	Y	N	N
pfg010T	7	20751484	G/C	het	S	1	G	GC	Y	Y	Y
pfg010T	7	56093883	T/C	het	S	1	T	CT	Y	Y	Y
pfg010T	7	121731574	G/T	het	S	1	G	GT	Y	Y	Y
pfg010T	7	128285801	C/G	het	S	1	C	CG	Y	Y	Y

pfg010T	8	10503553	C/A	het	S	1	C	CA	Y	Y	Y
pfg010T	8	30820449	C/T	het	S	1	C	CT	Y	Y	Y
pfg010T	8	100513070	T/C	het	S	1	T	CT	Y	Y	Y
pfg010T	8	114255190	G/C	het	S	1	G	GC	Y	Y	Y
pfg010T	9	5223845	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	9	121115286	A/T	het	S	1	A	AT	Y	Y	Y
pfg010T	9	130808768	T/C	het	S	1	T	CT	Y	Y	Y
pfg010T	9	133375548	C/T	het	S	1	C	CT	Y	Y	Y
pfg010T	10	94659215	C/T	het	S	1	C	CT	Y	Y	Y
pfg010T	10	117030947	C/G	het	S	1	C	GC	Y	Y	Y
pfg010T	11	128345096	A/G	het	S	1	A	GA	Y	Y	Y
pfg010T	12	412651	A/G	het	S	1	A	GA	Y	Y	Y
pfg010T	12	7252512	C/T	het	S	1	C	CT	Y	Y	Y
pfg010T	12	7545147	A/G	het	S	1	A	AG	Y	Y	Y
pfg010T	12	77055139	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	12	109113197	T/A	het	S	1	T	TA	Y	Y	Y
pfg010T	14	20094918	A/T	het	S	1	A	AT	Y	Y	Y
pfg010T	14	49190481	T/C	het	S	1	T	CT	Y	Y	Y
pfg010T	15	41854167	C/G	het	S	1	C	CG	Y	Y	Y
pfg010T	15	47079383	C/T	het	S	1	C	CT	Y	Y	Y
pfg010T	15	51676732	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	16	55089204	T/C	het	S	1	T	CT	Y	Y	Y
pfg010T	16	63542280	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	16	82828210	G/A	het	S	1	GA	GA	N	Y	N
pfg010T	17	31967780	C/G	het	S	1	C	CG	Y	Y	Y
pfg010T	18	7013155	G/A	het	S	1	G	GA	Y	Y	Y
pfg010T	18	39108265	A/T	het	S	1	A	AT	Y	Y	Y
pfg010T	18	45817276	C/T	het	S	1	C	TC	Y	Y	Y
pfg010T	20	9508913	G/C	het	S	1	G	CG	Y	Y	Y
pfg010T	21	21578496	A/G	het	S	1	A	AG	Y	Y	Y
pfg011T	1	16134611	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	1	31921918	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	1	32518365	G/T	het	S	1	G	GT	Y	Y	Y
pfg011T	1	79156167	T/G	het	S	1	T	T	Y	N	N
pfg011T	1	86678530	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	1	90952700	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	1	148502352	A/T	het	S	1	A	TA	Y	Y	Y
pfg011T	1	149188269	A/G	het	S	1	A	AG	Y	Y	Y
pfg011T	1	150999908	G/A	het	S	1	G	AG	Y	Y	Y
pfg011T	1	156899166	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	1	158366583	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	1	212623501	C/T	het	S	1	C	TC	Y	Y	Y
pfg011T	1	246045343	A/G	het	S	1	A	GA	Y	Y	Y
pfg011T	2	71679300	G/A	het	S	1	G	AG	Y	Y	Y
pfg011T	2	160405585	C/T	het	S	1	C	TC	Y	Y	Y
pfg011T	2	169721102	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	2	171395867	T/A	het	S	1	T	AT	Y	Y	Y
pfg011T	2	189570320	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	3	45852127	C/G	het	S	1	C	GC	Y	Y	Y
pfg011T	3	64507473	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	3	174177488	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	4	100747028	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	4	114498425	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	4	118225130	T/G	het	S	1	T	TG	Y	Y	Y
pfg011T	4	122177623	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	4	126556216	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	4	130083599	A/G	het	S	1	A	GA	Y	Y	Y
pfg011T	5	102337985	T/A	het	S	1	T	AT	Y	Y	Y
pfg011T	5	126806646	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	5	147758836	T/C	het	S	1	T	CT	Y	Y	Y
pfg011T	6	7515684	C/T	het	S	1	C	TC	Y	Y	Y
pfg011T	6	47757318	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	7	37746610	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	7	43499235	T/C	het	S	1	T	TC	Y	Y	Y
pfg011T	8	32737373	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	8	97866687	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	8	107760626	A/G	het	S	1	A	AG	Y	Y	Y
pfg011T	8	107851480	C/A	het	S	1	C	CA	Y	Y	Y
pfg011T	9	23691404	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	9	108729663	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	9	124695083	T/C	het	S	1	T	CT	Y	Y	Y
pfg011T	9	129576483	C/T	het	S	1	C	TC	Y	Y	Y
pfg011T	10	123836530	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	11	55491795	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	11	82861472	A/G	het	S	1	A	GA	Y	Y	Y
pfg011T	11	116543430	G/A	het	S	1	G	AG	Y	Y	Y
pfg011T	11	129844395	C/G	het	S	1	C	CG	Y	Y	Y
pfg011T	12	1760422	A/G	het	S	1	A	GA	Y	Y	Y
pfg011T	12	4892179	A/G	het	S	1	A	A	Y	N	N
pfg011T	12	106569597	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	13	66699756	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	13	66700441	T/G	het	S	1	T	GT	Y	Y	Y

pfg011T	13	76493903	G/A	het	S	1	G	AG	Y	Y	Y
pfg011T	14	41426564	C/G	het	S	1	C	CG	Y	Y	Y
pfg011T	14	44044924	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	14	51264476	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	15	23522228	G/T	het	S	1	G	GT	Y	Y	Y
pfg011T	15	40854866	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	15	72490964	G/A	het	S	1	G	AG	Y	Y	Y
pfg011T	16	12706014	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	16	20988359	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	16	24280346	A/C	het	S	1	A	CA	Y	Y	Y
pfg011T	17	39693642	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	17	45154921	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	17	64425148	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	17	71516038	C/G	het	S	1	C	CG	Y	Y	Y
pfg011T	18	336549	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	18	53294791	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	19	35732082	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	20	20116675	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	20	57002668	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	20	59263293	T/G	het	S	1	T	GT	Y	Y	Y
pfg011T	21	32277741	A/T	het	S	1	A	AT	Y	Y	Y
pfg011T	22	45031788	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	22	45238870	C/T	het	S	1	C	CT	Y	Y	Y
pfg011T	X	71276472	G/A	het	S	1	G	GA	Y	Y	Y
pfg011T	X	152786702	G/A	het	S	1	G	AG	Y	Y	Y

Supplementary Table 5 Comparison of somatic point mutation spectra across different subtypes of gastric cancers

	MSI		MSS	
Base pair change	%	number	%	number
T>G/A>C	4%	121	8%	142
T>C/A>G	30%	876	13%	220
T>A/A>T	4%	108	8%	137
C>T/G>A*	53%	1562	48%	831
C>G/G>C	1%	31	9%	150
C>A/G>T	9%	271	15%	261

*68% of T>C/A>G mutation occur in CpG dinucleotide setting

Supplementary Table 8 Core signaling pathways and processes frequently altered in gastric cancers. Pathways altered in more than one-third of gastric cancers and involving biological processes highly relevant to carcinogenesis are shown and arranged according to most statistically significant enrichment (See Supplementary Table 9 for the complete list and statistics)

Core cancer process	Number of genes mutated	% tumors with mutation	Representative mutated genes
Chromatin modification*	20 (46)	50% (59%)	ARID1A,CARM1,CREBBP,H1FNT,HDAC6,MAP3K12,MYST1,MYST3,MYST4,NAP1L4,NSD1,PBRM1,PPARGC1A,RPS6KA5,RSF1,SIRT1,SMARCC1,SYCP3,TLK2,WHSC1L1 (ARID1B,ASH1L,AURKB,BPTF,BRD8,CBX6,CHD1,CHD3,CHD4,DNMT1,DNMT3B,EHMT1,EP400,EZH1,INO80,KAT5,MBD2,MLL,MLL5,MORF4L1,PRDM2,SETD2,SETDB1,SMARCA1,SRCAP,UBE2E1)*
Cell junction organization	21	59%	CADM3,CDH1,CDH11,CDH12,CDH18,CDH2,CDH24,CDH4,CDH7,CTNNA1,CTNNB1,DST,FLNC,INADL,ITGB4,LAMA3,PARD3,PARD6B,PRKCI,TESK1
ALK pathway (WNT/BMP/TGFB)	12	55%	APC,ATF2,BMP4,BMPR1A,BMPR2,CHRD,CTNNB1,MAP3K7,SMAD4,TGFB1,TGFB3,TGFBR3
Axon guidance#	32	59%	CACNB2,COL6A3,DPYSL5,ERBB2,MYH11,ROCK1,ROCK2,SEMA3E,SEMA6D,SEMA7A,SOS2,SPTA1
BMP signaling	9	36%	ACVR2A,ACVR2B,BMPR1A,BMPR2,CER1,CHRDL1,FSTL1,GREM2,SMAD4
TGFB pathway	8	41%	APC,CDH1,CREBBP,MAP2K1,MAP3K7,SMAD4,TGFB1,TGFB3
MAPK signaling	44	82%	ATF2,CACNA2D2,CACNB2,CACNG3,EGF,FGF23,FGF5,FLNB,FLNC,HSPA6,IL1R1,KRAS,MAP2K1,MAP3K1,MAP3K12,MAP3K3,MAP3K7,MAP4K1,MAP4K3,MAPK10,MAPK14,MAPK8IP3,MECOM,MOS,NFKB1,NGF,NRAS,NTF3,PDGFRB,PLA2G12B,PPP3CA,PRKCA,PTPN5,PTPRR,RAPGEF2,RASA2,RELA,RPS6KA3,RPS6KA5,SOS2,TAOK2,TGFB1,TGFB3,TP53
DNA replication	21	41%	ACHE,CDC6,EGF,EXO1,GLI1,GTPBP4,KCTD13,KIN,MCM3,MCM3AP,MCM7,MRE11A,MSH3,MSH6,MUTYH,NAE1,NUP98,ORC4L,RAD17,REV3L,TP73
Focal adhesion#	36	64%	COL11A1,COL5A3,COL6A3,CTNNB1,ERBB2,FLNB,FLNC,ITGA2,LAMA1,LAMA2,LAMA3,PTEN,ROCK1,ROCK2,SOS2,TNR
ERBB signaling	19	45%	ABL2,EGF,ERBB2,ERBB3,ERBB4,KRAS,MAP2K1,MAPK10,NRAS,NRG1,NRG3,PAK6,PAK7,PIK3R3,PLCG1,PRKCA,PTK2,SHC4,SOS2
Cell cycle regulation#	50	77%	APC,BRCA2,FANCG,MRE11A,PLK1,PTEN,SMC4,TP53,TTK,ZBTB17
ATR/BRCA pathway	7	64%	ATM,BRCA1,BRCA2,FANCG,MRE11A,RAD17,TP53
RB signaling	5	50%	ATM,MAPK14,MYT1,TP53,WEE1
Fas signaling	13	55%	ALG2,CAD,DEDD,FADD,MAP3K1,MAPK10,MAPK8IP3,MET,NFKB1,NFKBIE,ROCK1,TP53,TUFM
Rho GTPase cycle#	22	36%	ARAP2,ARHGAP20,ARHGAP22,ARHGAP28,FAM13B,MYO9A,SOS2,TIAM1
Ubiquitin mediated proteolysis#	24	46%	BRCA1,HERC2,TRIP12,UBE3C,UBR5

* Number, % and Genes in bracket are calculated with inclusion of manually curated chromatin modification genes not included in the GO biological process pathway list (see Supplementary Table 10)

Only genes mutated in 2 or more samples are shown in these pathways because of the large number, see Supplementary table 9 for detail list

Supplementary Table 11 Summary of clinico-pathological data of 109 gastric cancer patients sequenced for ARID1A

Variable	Number
Sex	
M	72
F	37
Age	
mean=68.8 (range 23-88)	
Lauren's tumor type	
Intestinal	86
Diffuse	12
Mixed	10
Indeterminate	1
Tumor differentiation	
Well/moderate	48
Poor	61
Tumor site	
Cardia	35
Body	26
Antrum	46
Diffuse	2
Tumor stage (UICC 2002)	
Stage I	15
Stage II	21
Stage III	35
Stage IV	38
EBV status	
Present	15
Absent	94
MSI status	
MSI	23
MSS	86
Helicobacter pylori infection	
Present	56
Absent	53
Operation type	
Curative resection	62
Palliative resection	47
Recurrence status for patients who have undergone curative resection	
Recur	27
No recurrence	35
Median follow-up period for patients with no recurrence = 89 months	
Median time to recurrence = 14 months	
Follow up status	
Alive or censored	41
Died of disease	68
Median follow-up period for patients alive = 89 months	
Median time to death = 14 months	

Supplementary Table 12 Summary of relationship of ARID1A mutation with protein expression and molecular subtypes of gastric cancer. Results are derived from Sanger sequencing of ARID1A and TP53 in 109 GCs. There is no significant association of ARID1A alteration with sex, age, tumor location, Lauren's tumor type, differentiation, Helicobacter pylori infection, and stage.

	Total no.	No. with ARID1A mutation (%)	No. with reduced ARID1A protein level (%)[a]	Total no. with ARID1A alteration (mutation and/or reduced protein) (%)	p-value (χ^2 test)[b]
Molecular subtype					
MSI	23	18 (78%)	15 (65%)	19 (83%)	<0.001[c]
MSS EBV	15	7 (47%)	10 (67%)	11 (73%)	0.002[d]
MSS non-EBV	71	7 (10%)	5 (7%)	8 (11%)	
TP53 mutation					
mutated	45	7 (16%)	5 (11%)	8 (18%)	0.008[e]
wildtype	64	25 (39%)	25 (39%)	30 (47%)	
ARID1A immunostaining					
normal	79	8 (10%)			<0.001
weak	8	7 (88%)			
loss	22	17 (77%)			

[a] Reduced protein level corresponds to either loss or weak expression.

[b] p-values of ARID1A mutation in relation to various molecular parameters are shown. Similar significance levels. are observed for ARID1A protein or combined mutation and/or protein alterations.

[c] compared with MSS non-EBV.

[d] compared with MSS non-EBV (Fisher exact test).

[e] p=0.002 for ARID1A alteration in relation to TP53 mutation.

Supplementary Table 14 Univariate and multivariate cox regression analysis of variables related to recurrence-free survival in 62 gastric cancer patients undergoing curative resection and analyzed for ARID1A alteration

	Univariate analysis				Multivariate analysis			
Variable	**n**	**HR**	**95% CI**	**p**	**n**	**adjusted HR**	**95% CI**	**p**
Stage								
I	15	1				1		
II	18	0.55	0.09-3.31	0.517		0.59	0.1-3.55	0.567
III	19	8.86	2.54-30.93	0.001		9.25	2.62-32.62	0.001
IV	10	8.63	2.13-35.00	0.003		13.8	3.15-60.50	0.001
ARID1A alteration								
Mutated or reduced protein level	22	1			22	1		
No alteration	40	2.35	0.94-5.84	0.067	40	3.09	1.13-8.50	0.029
MSI status								
MSI	16	1						
MSS	46	3.46	1.04-11.52	0.043				
Lauren's tumor type								
Intestinal/mixed/indeterminate	52	1						
diffuse	10	4.85	2.06-11.40	<0.001				

Supplementary Table 15 Comparison of somatic indel mutation rate of mononucleotide repeats in ARID1A gene versus background mutation rate in MSI gastric cancers

		ARID1A mutation rate			Background mutation rate				
Type of Mononucleotide repeat	Number of specific repeat in coding region of ARID1A	Total no of specific repeat in coding region of ARID1A in 23 MSI GCs	No of indel mutation in specific repeat in coding region of ARID1A in 23 MSI GCs	Indel mutation rate of specific repeat in ARID1A for 23 MSI GCs (%)	Total no of specific repeat in coding region of all genes adequately covered by exome sequencing in 4 MSI GCs[a]	No of indel mutation in specific repeat in coding region of all genes in 4 MSI GCs[b]	Indel mutation rate of specific repeat in coding region of all genes for 4 MSI GCs (%)	Fold change of indel mutation rate (ARID1A coding repeats/global coding repeats)	p-value (chi-square test)
G7	1	23	6[c]	26.087%	1520	20	1.316%	20	0.000000
G6	5	115	2	1.739%	12170	18	0.148%	12	0.000025
G5	9	207	0	0.000%	65488	16	0.024%	0	NS
G4	23	529	0	0.000%	245287	5	0.002%	0	NS
C6	9	207	7	3.382%	13721	28	0.204%	17	0.000000
C5	17	391	7	1.790%	68308	20	0.029%	61	0.000000
C4	52	1196	1	0.084%	257183	9	0.003%	24	0.000009
A6	1	23	0	0.000%	22514	46	0.204%	0	NS
A5	2	46	0	0.000%	86007	33	0.038%	0	NS
A4	8	184	0	0.000%	286739	19	0.007%	0	NS
T5	1	23	0	0.000%	85092	18	0.021%	0	NS
T4	3	69	0	0.000%	287378	15	0.005%	0	NS

[a]counting all mononucleotide repeats in coding region of the 16,586 genes targeted by the SureSelect All Exon kit and with adequate sequence coverage (≥10X in tumor and paired normal) in 4 MSI gastric cancers

[b]Counting total number of indels involving these specific repeat sequences in coding regions as specified in column F

[c]One of the MSI GC (sample SX186) has two indels (one insertion and one deletion) involving the same G7 repeat and were counted as one in the statistical analysis

Supplementary Table 16 Primers used for Sanger sequencing of ARID1A and TP53

Gene Name	Genome_Region	Primer Name (forward)	Sequence	Primer Name (reverse)	Sequence	Coding Exon number
ARID1A	Chr1:26895428-26896189	Ex01F0*	GGGGAGAAGACGAAGACAGG	Ex01R2	ACCTCTCGGGGAGCTCAG	1
ARID1A	Chr1:26896075-26896759	Ex01F3	CAGCAGAACTCTCACGACCA	Ex01R3	GAGAAGAGCCAGACAATGGC	1
ARID1A	Chr1:26928590-26929031	Ex02F	AGGTTGGTCTCATTGCTCTTTC	Ex02R	TTGGAAGCCAAGGATACATTC	2
ARID1A	Chr1:26930094-26930785	Ex03F1	ACCCTGGGCCTCCTAAGTATG	Ex03R2	TGCACGTTAGAGAACCACTCTG	3
ARID1A	Chr1:26931669-26932043	Ex04F	CAGTCCCATAACCCTTTCACAG	Ex04R	CTGGGCAGGGAGACAGAAC	4
ARID1A	Chr1:26959808-26960291	Ex05F	AAAGAACGTGTGTGATGTATTTGC	Ex05R	GAAACTATGCAGGCATGAGCC	5
ARID1A	Chr1:26960325-26960643	Ex06F	TTGGCTGGATCTCTTTGTGTG	Ex06R	TTCATGGTCAAACAGCTCTCC	6
ARID1A	Chr1:26961155-26961503	Ex07F	TCCCAGGATAAGGATGGAGAG	Ex07R	GGACAGCCCTTCTCTCACAAG	7
ARID1A	Chr1:26961960-26962445	Ex08F	TTGAATGACATTGTTTGGTGTTC	Ex08R	GGTCCAGAAGCATCTCAATAATC	8
ARID1A	Chr1:26965217-26965737	Ex09F	CACAGCACTATTTGGCTCCAG	Ex10R	GGCTGGGATCTTGTCACTCTC	9,10
ARID1A	Chr1:26966793-26967161	Ex11F	CAAGAGACTTCTGAGACCCTTAGC	Ex11R	CATGGTACCACATGAAGCCAG	11
ARID1A	Chr1:26970107-26970509	Ex12F	GAATACCTTACAGCCTGATGGG	Ex12R	ATCCTTGGCATATCCTGTTGG	12
ARID1A	Chr1:26971469-26972155	Ex13F	GGCCTTAGGAAGAACTTTCCC	Ex14R	CAAGAACCCTGAGCCATTCTC	13,14
ARID1A	Chr1:26972337-26973137	Ex15F	GAACTCTGAAGAGGGCCTGG	Ex17R	ATTGAGGACGTGGCTCTTCAG	15,16,17
ARID1A	Chr1:26973227-26974866	Ex18F1	GGAAGAAAGAGTGGTGGTTGC	Ex19R	TGGCTAAAGATGAGACATTCCC	18,19
ARID1A	Chr1:26977966-26978947	Ex20F1	GGAGAACCTTTGGGAAAGGAG	Ex20R2	GGCTTCGAATGGTATTGGACA	20
ARID1A	Chr1:26978755-26979672	Ex20F3	GCTAAGAGTTCAGAGGCCATCA	Ex20R4	CCGCATCATGTCCACACTA	20
ARID1A	Chr1:26979439-26981359	Ex20F5	CTCAGTGACCGAAAGAACCC	Ex20R5	GAGGTGGAAGGAGGAGAGAGA	20
TP53	Chr17:7518795-7519311	TP53_5F	CTTTCAACTCTGTCTCCTTCCTC	TP53_6R_new	GGGAGGTCAAATAAGCAGCA	5,6
TP53	Chr17:7517448-7518362	TP53_7F	CCTCATCTTGGGCCTGTGTT	TP53_9R_new	TGTCTTTGAGGCATCACTGC	7,8,9

*all primers for ARID1A were according to a previous study (Jones, S. et al. Science 330, 228-31 (2010)) except the one labeled with "asterisk" which was newly designed.

ARID1A: Ensembl Gene ID ENSG00000117713, Ensembl transcript ID ENST00000324856

TP53: Ensembl Gene ID ENSG00000141510, Ensembl transcript ID ENST00000269305

Supplementary Table 17 Background mutation rate estimates for somatic SNVs and somatic indels in the studied gastric cancer cohort. Background mutation rates were separately estimated in microsatellite unstable gastric tumors (MSI) and microsatellite stable gastric tumors (MSS) for both somatic SNVs and somatic indels. For somatic SNVs, background mutation rates are calcuated for different nulceotide bases.

	Somatic SNVs				Somatic indels
	A	C	G	T	
MSI	1.48E-05	2.38E-05	3.91E-05	1.41E-05	3.78E-05
MSS	9.66E-07	3.68E-06	4.74E-06	1.04E-06	3.24E-07

Supplementary Note

Algorithms for exome-sequencing data analysis, somatic mutation callers, mutation annotation and functional impact prediction, driver gene prediction and pathway analysis

Contents

Primary sequence analysis

Alignment of sequencing reads to the human reference genome (hg18, NCBI36) was done using Illumina's Eland-v2 short read aligner. Single nucleotide variants (SNVs) were called using Samtools[1]. SNV calls from the MAQ caller[2] were further filtered using the Samtools "varFilter" to retain only those with a Phred-scale SNP quality ≥40, root mean square mapping quality ≥25 and read depth ≥3X. In addition, SNVs within ±10bp of an indel are filtered out if the indel has a score ≥25; and in case of more than 2 SNVs called within a 10bp window, all of them were removed. Small insertions and deletions (indels) were predicted using CASAVA's GROUPER indel caller[3] with default parameter settings except that the "indelsCovCutoff" is set to -1 to disable the local coverage filter, which is recommended for targeted re-sequencing.

Identification of somatic SNVs and somatic indels

We developed a new procedure for calling somatic SNVs and somatic indels from short read sequencing data, called SMART (Somatic Mutation Assessment Research Tools).

Algorithm for somatic SNV identification

For identification of somatic SNVs, SMART takes as inputs the two sets of SNVs called by Samtools in pileup format[1] in tumor and its matched normal sample. It then subtracts the set of SNVs called in the normal sample from those called in the corresponding tumor. The remaining candidates are subsequently filtered out with the following criteria:

1. The Phred scale SNP quality score in tumor must be ≥45
2. The position must be covered by at least 10 reads in the matched normal data
3. The root-mean-square mapping quality of the reads covering the position in matched normal data must be ≥25
4. Predicted mutant allele with minimal base quality of Q20 can present in at most 5% of total reads in the matched normal sample
5. Validated SNPs in dbSNP130 were removed

In addition to controlling the observed mutant allele frequency in the match normal sample, we also required that the mutant allele cannot appear in more than 5% of reads in the normal data of at least 95% of all patients in our discovery cohort. Our study has shown that this helps to further reduce potential false positive somatic calls due to the undercalling of germline variants in the matched normal sample, especially at positions with relatively low coverage.

Criteria (1) and (4) control, respectively, the stringency of SNV calling in tumor and how lenient the presence of tumor allele is tolerated in matched normal data. Their values were selected to achieve optimal performance, in terms of both sensitivity and specificity, on the 10 tumor and matched normal pairs from the first phase of our study. The value for the parameter in (1) was assessed by 34 known somatic SNVs in a subset of the 10 patients by Sanger sequencing characterized in a previous study[4]. On the other hand, we assess the parameter value in (4) using false discovery rate (FDR) computed from replicate sample experiments. More specifically, we have generated data on two technical replicate sample pairs, one pair for a gastric tumor and another for the normal gastric sample from the same patient. Both replicate pairs started from the same DNA extract but underwent independent library preparation and sequencing. SMART was then applied to both replicate pairs treating one sample in the pair as "tumor" and the other as "matched normal", thus any "somatic" call made on such replicate pair should be considered false discovery. FDR was then computed as the average number of false discoveries divided by the average number of discoveries made on real tumor and matched normal comparisons. (Note that there are four possible combinations of real tumor and matched normal pairs.) The optimal parameter combination was then determined based on a 2-dimentional grid search. The final parameter set achieved a sensitivity of 85.3% at FDR $\leq$5% in this training process.

Somatic SNV validation

SMART predicted a total of 1,848 somatic SNVs among the 10 tumor and matched normal pairs sequenced in the first phase. We performed independent Sequenom MassARRAY genotyping assays on a subset of 472 predicted somatic SNVs in protein coding regions (after removing the known ones

used in the training process and those that failed primer/assay design) (Supplementary Table 3). Each position was assayed in both tumor and matched normal samples from the original patient. 457 (96.8%) of the 472 positions were confirmed as somatic. Of the 15 that failed, 13 were due to false positive SNV calls made in tumor samples and 2 were due to false negative detection of the mutant allele in the matched normal samples (Supplementary Table 3). We removed these 15 invalidated somatic SNVs from our final results. Although we have confirmed only a subset of predicted somatic SNVs from the first 10 cases, the high confirmation rate demonstrated here indicates that the vast majority of the unconfirmed somatic SNVs in our final results on all 22 cases are likely to be true.

Algorithm for somatic indel identification

SMART uses a similar process for calling somatic indels. The indel calls from tumors and their matched normal sample come from the GROUPER indel caller in CASAVA[3]. Due to ambiguity of indel mapping[5], subtraction of indels called in matched normal data from those in tumors is done requiring no overlap between the called indel position $\pm$5bp on both sides[6]. Remaining candidates are subsequently filtered by the following steps:

1. Average depth of coverage in matched normal data for the indel$\pm$5bp window must be at least 10X
2. All positions in the indel$\pm$5bp window must be covered
3. Number of reads supporting the indel in tumor must be $\geq$10
4. The total number of gaps and substitutions in the aligned reads within the same window must be $\leq$3% of the total bases mapped to this window in the matched normal data

We have further required that criterion (4) to be satisfied in the normal data of at least 95% of all cases in our discovery cohort, in order to minimize false negative detections in the matched normal sample due to insufficient sampling, especially at relatively low coverage regions.

Again, criteria 3 and 4 specify how strong the evidence for the indel is in tumor and how "clean" (i.e. absence of gaps and substitutions) the corresponding region is in the normal data. Their values were

fine tuned to achieve optimal performance for both sensitivity and specificity on the first 10 tumor and matched normal pairs sequenced. Similar to the training procedures for somatic SNVs, sensitivity was assessed using data on 8 known somatic indels in a subset of these 10 patients, and FDR was estimated using replicate sample pairs. The final parameter set selected achieved a sensitivity of 100% at FDR≤3% in our training procedure, and led to the identification of 597 somatic indels among the first 10 tumor and matched normal pairs sequenced in the first phase.

Somatic indel validation

We performed independent Sequenom MassARRAY genotyping assays on a subset of 189 predicted somatic indels. This subset was selected to include all coding somatic indels plus non-coding somatic indels from case "pfg008" (which is a microsatellite instable case), that are shorter than 40bp, after removing the known ones used in the training process and those failed primer/assay design (Supplementary Table 3). Each indel was assayed in both the tumor and the matched normal samples from the original patient. 185 (97.9%) of the 189 cases were confirmed as somatic. Of the 4 that failed, two were due to false positive indel calls made in tumor samples and 2 were due to false negative detection of germ-line indels in the matched normal samples (Supplementary Table 3). We removed these 4 invalidated somatic indels from our final results. Although we have confirmed only a subset of the predicted somatic indels in the first 10 cases, the high confirmation rate we have demonstrated here indicates that vast majority of the unconfirmed somatic indels in our final results on all 22 cases are likely to be true.

Mutation annotation and functional impact prediction

Predicted somatic mutations and somatic indels were annotated using the Ensembl database version 54 (May 2009, http://may2009.archive.ensembl.org/index.html) via the associated Perl API. A list of predicted variation consequences relative to the transcript structure can be found at http://uswest.ensembl.org/info/docs/variation/index.html. When multiple transcripts were found for annotating a given variant, the longest transcript(s) was used.

Potential functional impact of the somatic mutations were predicted using an in-house installation of SIFT (version 4.0.3, http://sift.jcvi.org/) with default parameters provided by the tool.

Known germ-line variations in our predictions were searched using dbSNP build 130 downloaded from the UCSC genome browser (http://genome.ucsc.edu). We only included SNPs that have a validation status, including those 'by-cluster', 'by-frequency', 'by-submitter', 'by-2hit-2allele', 'by-hapmap' and 'by-1000genomes'.

Driver gene prediction

Somatic mutations that are passenger events are not under selection pressure thus may simply reflect background mutations in tumors. We therefore devised a "driver gene score" that tests whether the observed mutation frequency of a gene is statistical significantly higher than expected by background mutation alone, after correcting for molecular subtypes of gastric cancer and different types of somatic mutations. The score consists of 4 components, each assessing the probability that the observed number of 1) somatic SNVs in microsatellite stable (MSS) gastric tumors, 2) somatic SNVs in microsatellite unstable (MSI) gastric tumors, 3) somatic indels in MSS tumors and 4) somatic indels in MSI tumors, can be expected simply by the background mutation frequency. The 4 MSI cases were separated from the 18 MSS cases due to their drastically different background mutation rates as a result of DNA mismatch repair deficiency. Somatic SNVs and somatic indels were also considered different categories as their background mutation rates need to be separately estimated.

Estimating the background mutation rate for somatic SNVs

For somatic SNVs, it has been shown that the background mutation rate can be different at different nucleotide bases[7,8]. We started by counting the total number of exon bases in each nucleotide category, N_i, $i \in \{A, C, G, T\}$, that are sufficiently covered in individual tumors (≥3X) and their matched normal samples (≥10X). The total number of protein-altering background mutations in each nucleotide category was estimated by multiplying the number of observed synonymous somatic SNVs, s_i, by the nonsynonymous-to-synonymous SNV ratio, r_i. The latter was computed by a

simulation where we randomly substituted each position in the DNA codons and counted whether it will lead to a change in the encoded amino acid. The background mutation rate for protein-altering somatic SNVs can then be calculated as

$$f_i^s = \frac{s_i \cdot r_i}{N_i}$$

For a gene, g, with sufficiently covered coding bases in each nucleotide category, n_{gi}, the expected number of protein-altering mutations is:

$$\lambda_g^s = \sum_i f_i^s \cdot n_{gi}$$

Then given the observed number of protein-altering somatic SNVs in this gene, x_g^s, the probability that it can be expected by background mutation rate was computed using the Poisson distribution as

$$p_g^s = \sum_{k \geq x_g^s} \frac{\left(\lambda_g^s\right)^k e^{-\lambda_g^s}}{k!}$$

This process was separately applied to both MSS cases and MSI cases. For each group we aggregated the counts across all samples to obtain one set of background mutation rates and a single p-value for each gene. The background mutation rates for somatic SNVs in both MSS and MSI cases are summarized in Supplementary Table 17.

Estimating the background mutation rate for somatic indels

Background mutation rate for somatic indels is more difficult to estimate accurately than that for somatic SNVs, as all indels in the coding regions are protein-altering thus may be under general negative selection pressure. We decided to treat all somatic indels not overlapping with any coding exon or essential splice site in the SureSelect baits as background indel mutations, since they are unlikely to have functional consequences. In addition, since these noncoding regions mainly consist of sequences flanking the coding exons, such as those from nearby introns and UTRs, recent evidence has suggested that there is no significant differences in the overall mutation rate between

such regions and coding exons[9]. One possible explanation is that because these regions are also transcribed they should be at least under the same negative selection pressure by transcription-coupled mismatch repair, as previously suggested[3,10,11]. Even if additional negative selection pressure exists for coding exons compared to these noncoding regions, it will cause our background mutation rate to be slightly over-estimated, which won't inflate the statistical significance of our findings. Of the 37.6Mb bait sequences in the Agilent SureSelect exome capture kit, 10.3Mb (or 27.5%) belong to flanking non-coding regions. Due to the fact that SureSelect capture baits are designed to be centered around coding exons[12], the depth of coverage in coding regions is slightly higher than that in the flanking non-coding regions (mean coverage of 122X and 98X, respectively). However, at such high depth, the influence of coverage depth on variant detection is likely to have reached saturation[13,14], which justifies a direct comparison of the somatic indel mutation rate between these regions.

We counted the total number of non-coding bases in the SureSelect bait sequences that were sufficiently covered in individual tumors (≥3X) and their matched normal samples (≥10X), N_{nc}, and the number of detected somatic indels, d_{nc}, in these non-coding regions. The background somatic indel mutation rate was then calculated as:

$$f^d = \frac{d_{nc}}{N_{nc}}$$

Then for any gene, g, with sufficiently covered coding bases, n_g, the expected number of background somatic indels is

$$\lambda_g^d = f^d \cdot n_g$$

Given the observed number of somatic indels overlapping with the coding regions or essential splice sites of this gene, x_g^d, we can again compute the *p*-value based on the Poisson distribution as:

$$p_g^d = \sum_{k \geq x_g^d} \frac{\left(\lambda_g^d\right)^k e^{-\lambda_g^d}}{k!}$$

Similar to somatic SNVs, background mutation rates for somatic indels were estimated separately for the MSS cases and the MSI cases. For each group, counts from all samples were aggregated to obtain a single background mutation rate and *p*-value for each gene. The background mutation rates for somatic indels in both MSS and MSI cases are summarized in Supplementary Table 17.

Calculating driver gene score

We then combined the *p*-values from each of the 4 categories (i.e. somatic SNVs in MSS cases, somatic SNVs in MSI cases, somatic indels in MSS cases and somatic indels in MSI cases) using the truncated product method[15], which takes the product of only those *p*-values less than a specified cut-off value (0.05 in our study) and evaluates the probability of such a product, or a smaller value, under the overall hypothesis that the observed somatic mutations in all 4 categories are due to the background.

The combined *p*-value was further adjusted for multiple hypotheses testing of all genes carrying at least one protein-altering somatic mutation in at least two patients in our dataset using the *q*-value method[16]. Lastly, the resulting *q*-score, q_g, was converted to the "driver gene score", DS_g, by

$$DS_g = -log_{10}(q_g)$$

Genes with $DS_g \geq 0.7$ (i.e. $q_g \leq 0.2$) were selected to be top drivers of gastric cancer.

This process led to the selection of 21 putative driver genes. For mutations in these genes that have not been experimentally validated, they were all manually confirmed by examining the alignments in Integrative Genomic Viewer (http://www.broadinstitute.org/software/igv), except for two 99bp deletions (chr14:20749306-20749405) in gene HNRNPC. These two mutations were selected because they overlap with an essential splice site. However, our manual examination revealed that the deletion simply remove the entire intron between the first and second exon of HNRNPC, which

should have no effect on the resulting protein. In addition, there is no evidence of alternative splicing between the first two exons of this gene. Based on this information, we removed HNRNPC from our final putative driver gene list.

Pathway analysis

MSigDB[17] pathway databases (v3.0) were downloaded from the Broad Institute (http://www.broadinstitute.org/gsea/msigdb/index.jsp). We searched for over-represented molecular pathways among genes with protein-altering somatic mutations. In order to retain only the driver mutations while removing those passenger ones as much as possible, we used SIFT functional impact predictions (http://sift.jcvi.org/) as a crude filter and removed all nonsynonymous somatic SNVs not predicted as "damaging". Genes with somatic mutations in at least one patient in the discovery cohort, including SNVs predicted to be "damaging" by SIFT, those leading to stop-codon gains and losses or at essential splice site, and somatic indels overlapping coding regions, were searched for enrichment among gene sets in GO biological processes, KEGG pathways and Biocarta pathways. Gene sets with less than 10 members or more than 500 members were excluded. Enrichment was tested using Fisher's exact test. False discovery rate (FDR) was computed based on permutation, where 1000 random gene sets of same size as the input mutated gene list were queried for enrichment in the pathway database, keeping intact the intrinsic correlation structure among the pathway gene sets. Then for each gene set in the database with enrichment p-value of p with respect to the mutated gene list, the number of other gene sets with enrichment p-value $\leq p$ was counted and denoted N_p. The same number was also counted with respect to each of the 1000 random input gene lists, and the average, N_0, was taken as an empirical estimate of number of false discoveries at the p-value cutoff p. FDR was then computed as N_0/N_p.

Reference

1. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
2. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851-8 (2008).

3. Pleasance, E.D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-6 (2010).
4. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-8 (2007).
5. Krawitz, P. *et al.* Microindel detection in short-read sequence data. *Bioinformatics* **26**, 722-9.
6. Chen, K. *et al.* PolyScan: an automatic indel and SNP detection approach to the analysis of human resequencing data. *Genome Res* **17**, 659-66 (2007).
7. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-74 (2006).
8. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869-73 (2010).
9. Totoki, Y. *et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet*.
10. Pleasance, E.D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-90 (2010).
11. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473-7 (2010).
12. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182-9 (2009).
13. Bentley, D.R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-9 (2008).
14. Shah, S.P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809-13 (2009).
15. Zaykin, D.V., Zhivotovsky, L.A., Westfall, P.H. & Weir, B.S. Truncated product method for combining P-values. *Genet Epidemiol* **22**, 170-85 (2002).
16. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-5 (2003).
17. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).