Supplementary Information

Supplementary Figures 1-10
Supplementary Tables 1-16
Supplementary Notes

Exome Sequencing of Gastric Adenocarcinoma Identifies Recurrent Somatic Mutations in Cell Adhesion and Chromatin Remodeling Genes

Zhi Jiang Zang[1,2,20], Ioana Cutcutache[3,20], Song Ling Poon[4], Shen Li Zhang[2], John R McPherson[3], Jiong Tao[2,5], Vikneswari Rajasegaran[4], Hong Lee Heng[4], Niantao Deng[2,6], Anna Gan[4], Kiat Hon Lim[7], Choon Kiat Ong[4], DaChuan Huang[4], Sze Yung Chin[8], Iain Beehuat Tan[6,9], Cedric Chuan Young Ng[4], Willie Yu[4,6], Yingting Wu[10], Minghui Lee[1], Jeanie Wu[1], Dianne Poh[2], Wei Keat Wan[7], Sun Young Rha[11], Jimmy So[12], Manuel Salto-Tellez[8,13], Khay Guan Yeoh[14], Wai Keong Wong[15], Yi-Jun Zhu[16], P Andrew Futreal[17], Brendan Pang[5,8], Yijun Ruan[18], Axel M Hillmer[18], Denis Bertrand[18], Niranjan Nagarajan[18], Steve Rozen[3,#], Bin Tean Teh[2,4,19,#]& Patrick Tan [2,5,18,#]

[1]Cellular and Molecular Research, National Cancer Centre, Singapore.
[2]Cancer and Stem Cell Biology Program, Duke-National University of Singapore (NUS) Graduate Medical School, Singapore.
[3]Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, Singapore.
[4]National Cancer Centre Singapore-Van Andel Research Institute Translational Research Laboratory, National Cancer Centre, Singapore.
[5]Cancer Sciences Institute of Singapore, Singapore.
[6]Graduate School for Integrative Sciences and Engineering; National University of Singapore, Singapore.
[7]Department of Pathology, Singapore General Hospital, Singapore.
[8]Department of Pathology, National University of Singapore, Singapore.
[9]Division of Medical Oncology, National Cancer Centre, Singapore.
[10]Computation and Systems Biology, Singapore-MIT Alliance, Singapore.
[11]Department of Internal Medicine, Yonsei Cancer Centre, Yonsei University College of Medicine, Seoul, South Korea.
[12]Department of Surgery, Yong Loo Lin School of Medicine (NUS), Singapore.
[13]Centre for Cancer Research and Cell Biology, Queen's University Belfast, UK
[14]Department of Medicine, Yong Loo Lin School of Medicine (NUS), Singapore.
[15]Department of General Surgery, Singapore General Hospital, Singapore.
[16]Department of Pathology, Northwestern University, Feinberg School of Medicine, Chicago, USA.
[17]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, UK.
[18]Genome Institute of Singapore, Singapore
[19]Laboratory of Cancer Genetics, Van Andel Research Institute, Grand Rapids, Michigan, USA.
[20]These authors contributed equally to this work.

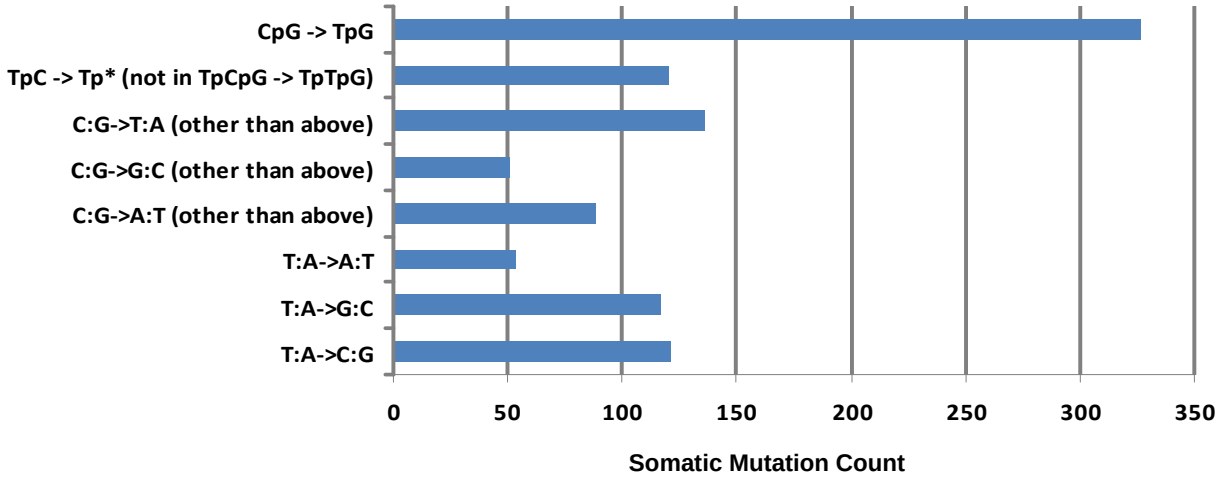
#Address Correspondence to Patrick Tan (gmstanp@duke-nus.edu.sg), Bin Tean Teh (teh.bin.tean@singhealth.com.sg), or Steve Rozen (steve.rozen@duke-nus.edu.sg)

Nature Genetics: doi:10.1038/ng.2246

Supplementary Figures

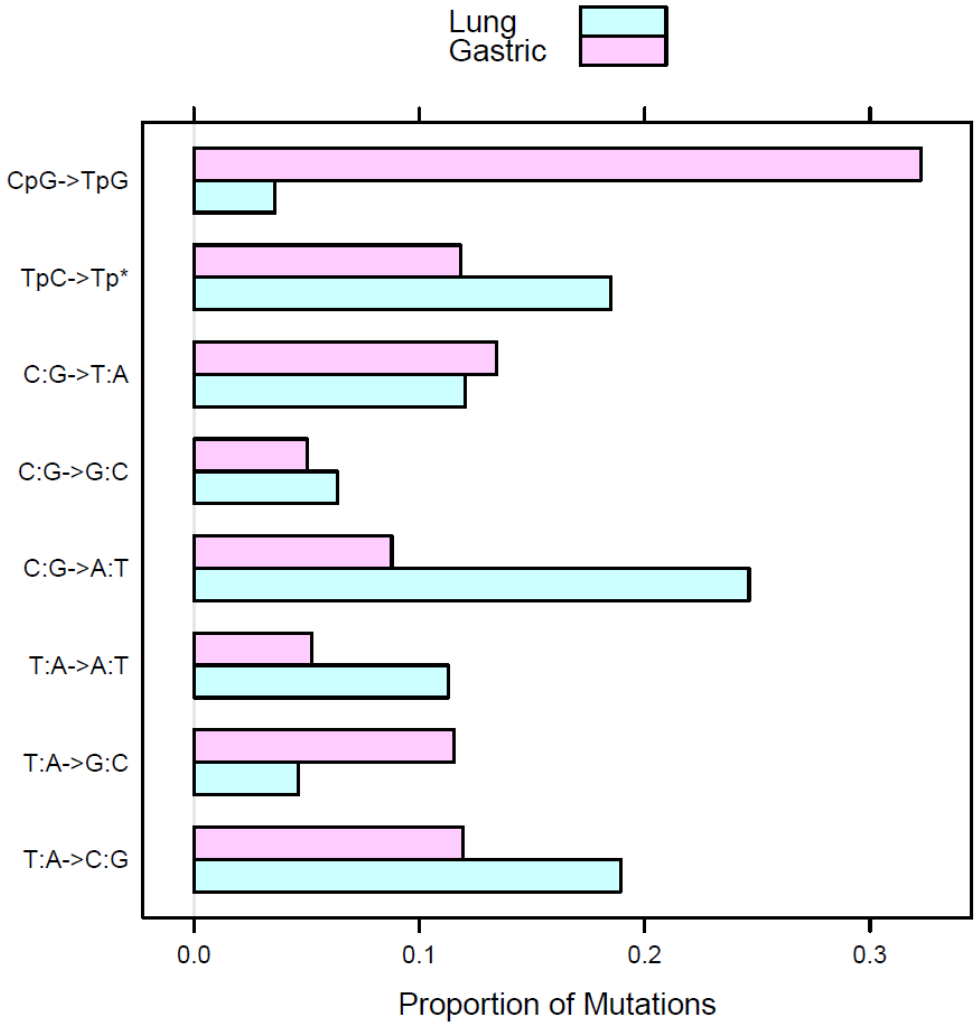
Supplementary Figure 1

Mutation spectra of 15 gastric adenocarcinomas. Mutation categories include their reverse complements – i.e. CpG -> TpG also includes CpG -> CpA, TpC->Tp* (mutation of C to any other base in the context of TpC) also includes GpA -> *pA, C:G -> T:A includes both C->T and G->A, and so forth. Counts include all in-target silent and non-silent somatic mutations passing quality filters, totaled over the 15 tumors.



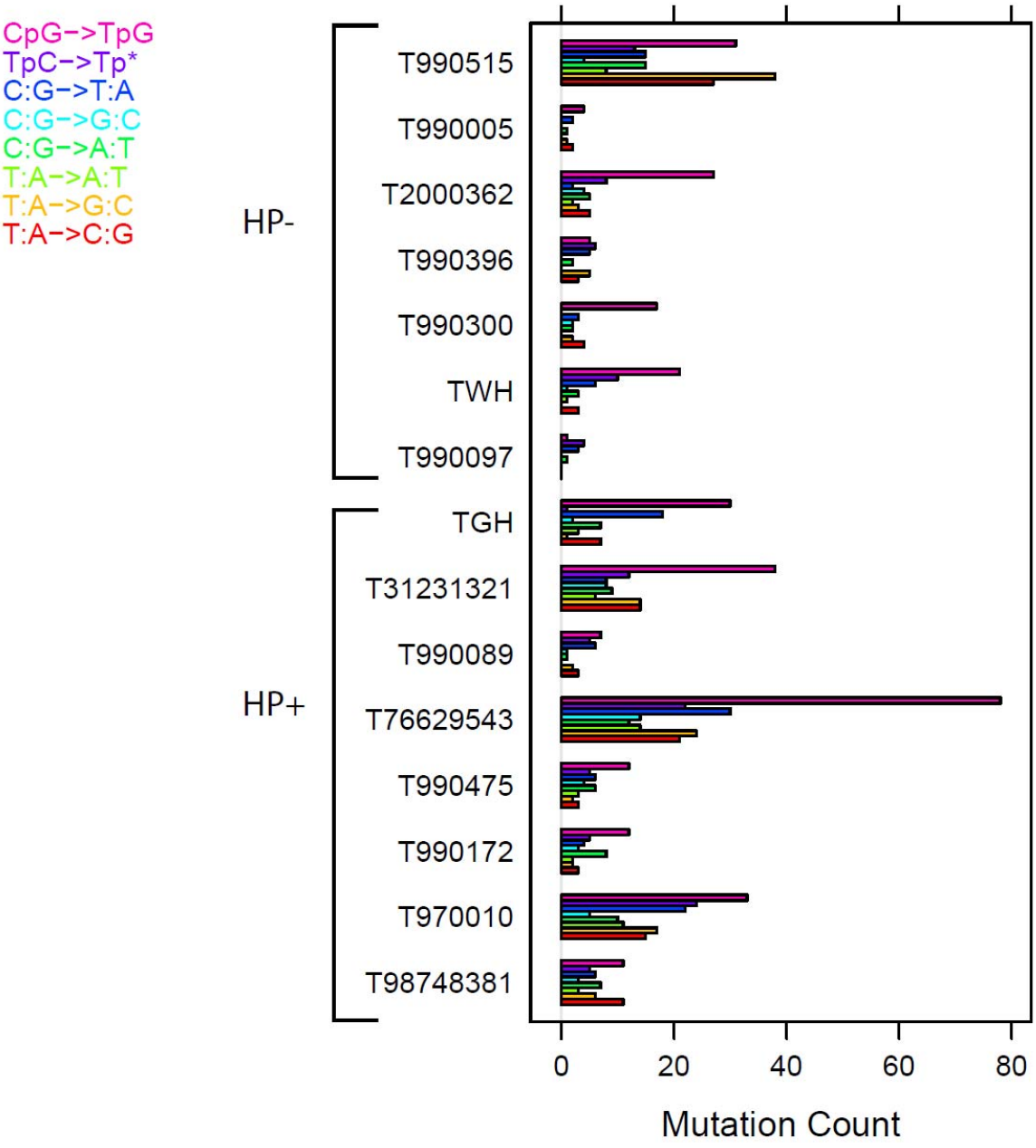
Supplementary Figure 2

Mutation spectra of gastric cancer and lung cancers. Mutations belonging to the different mutation categories are presented. Gastric cancers exhibit a significantly different mutational spectra compared to tobacco-associated lung cancers ($p < 10^{-16}$, chi-square test, see **Online Methods**).



Supplementary Figure 3

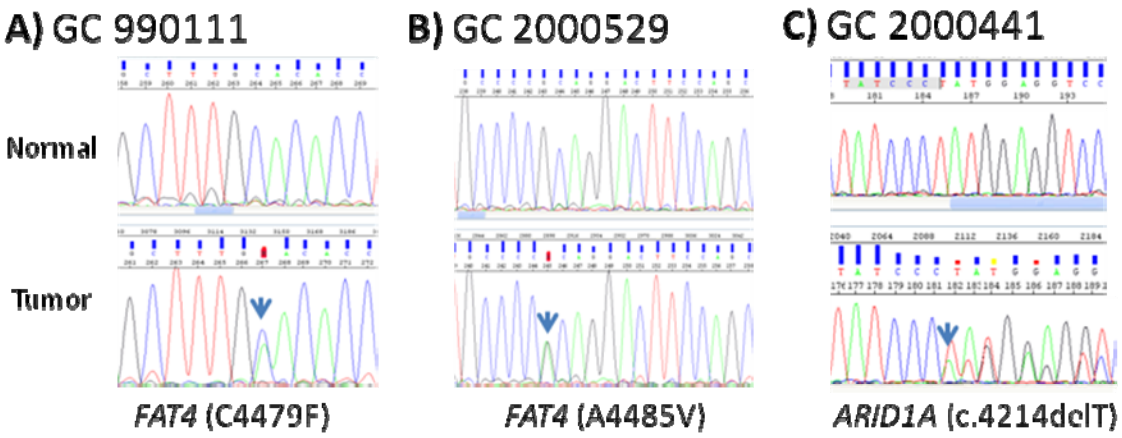
Mutation spectra in *H.-pylori* (HP)-negative and positive tumors. HP-negative tumors are shown at the top and HP-positive tumors at the bottom. Mutation categories are shown in the color legend. Considerable heterogeneity is observed across the individual tumors.



Nature Genetics: doi:10.1038/ng.2246

Supplementary Figure 4

Representative Sanger sequencing trace files of *FAT4* (a, b) and *ARID1A* (c) mutated validation samples. Comparisons of the sequenced regions between normal (top) and matched tumor samples (bottom) confirm the somatic nature of the alterations. Specific mutations are listed at the bottom of the trace images, and highlighted by the arrows in the images. del=deletion.



Supplementary Figure 5

Loss of heterozygosity at *FAT4* in tumors that also have missense somatic mutations in *FAT4*.

Panels A, B, and C refer to tumor 990515 and panels D, E, and F refer to 2000068. Both tumors have somatic mutations in *FAT4* (**Supplementary Table 11**) located in the middle of the long arm of chromosome 4. The analysis shows Affymetrix SNP Array 6.0 chip data from matched tumor and non-malignant DNA analyzed by ASCAT 2.0 (Allele-Specific Copy number Analysis of Tumors, Van Loo *et. al.*, 2010, see **Online Methods**).

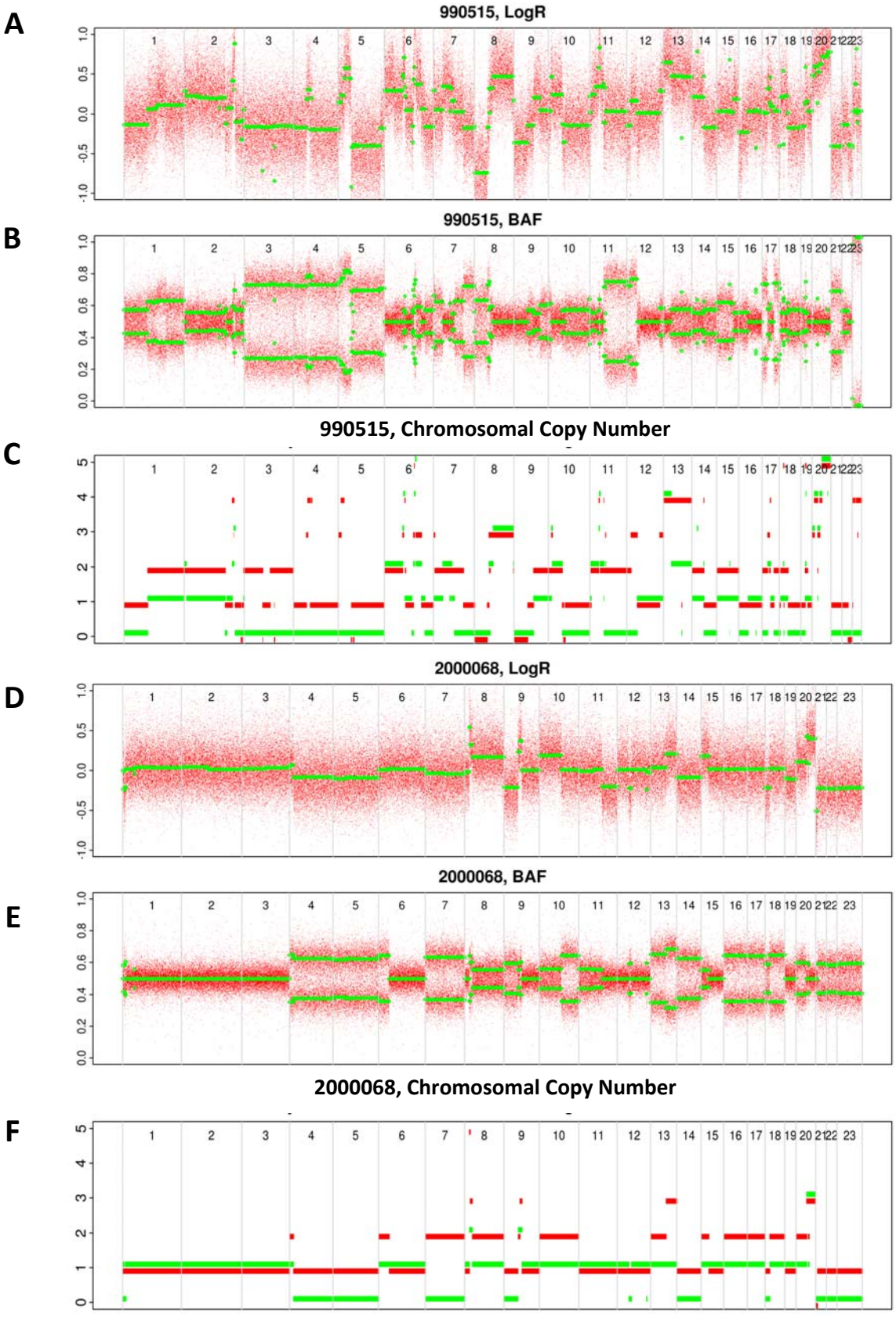
A, D) (LogR, "logR ratio") Each red dot shows, for a single SNP, the $\log_2$ ratio of the total (i.e. both alleles combined) probe intensities in the tumor to the total probe intensities in the matched non-malignant sample. This estimates the ratio of total copy number in the tumor to total copy number in the non-malignant sample. The overlaid green dots show the segmentation (i.e. smoothing) of these data.

B, E) (BAF, "B allele frequency") Each red dot shows the proportion of non-reference alleles in the tumor sample at sites that are heterozygous in the matched non-malignant sample. As for panels A and B, the green dots show the segmentation of these data. Regions where the green dots are simultaneously displaced to values higher and lower than 0.5 are regions of LOH or allelic imbalance. In both 990515 and 2000068 the green dots are widely separated across all of chromosome 4, indicating LOH.

C, F) ASCAT estimates of the genomic copy number of the two parental copies of each chromosome (arbitrarily colored red and green). The genomic copy number is shown along the *y* axis. Note that in both 990515 and 2000068, ASCAT estimates that one copy of chromosome 4 (the "green" copy) is completely deleted (copy number 0), leading to LOH.

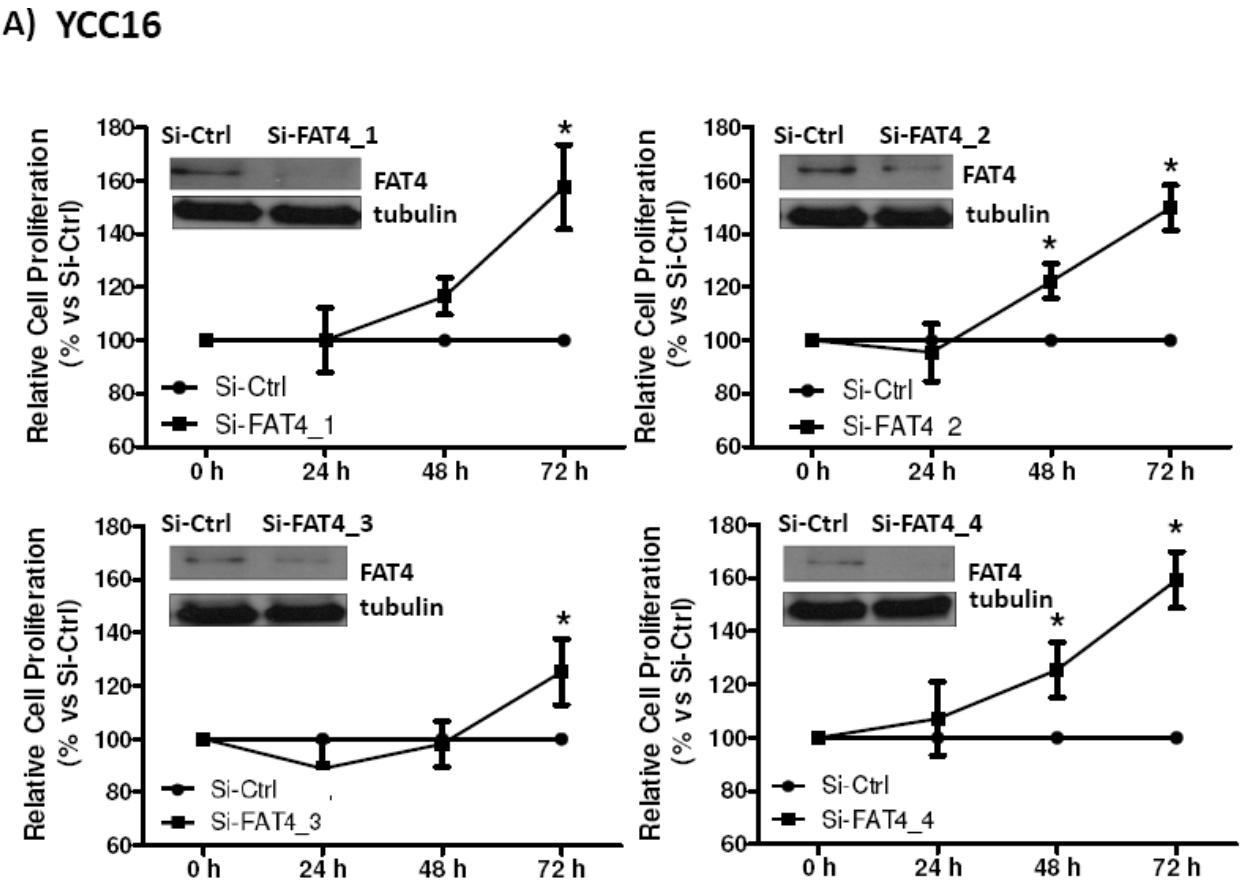
Nature Genetics: doi:10.1038/ng.2246

Supplementary Figure 5 (Legend on preceding page)



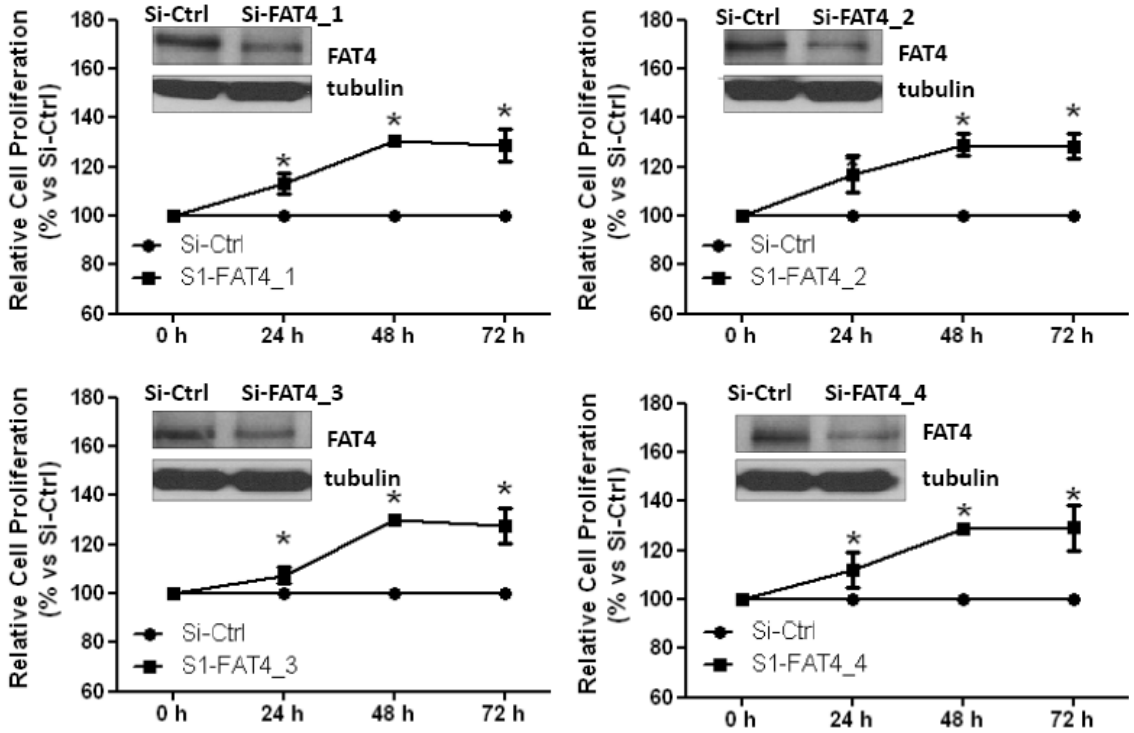
Supplementary Figure 6

***FAT4* silencing using individual *FAT4* siRNAs.** Four independent and non-overlapping *FAT4* siRNAs (Si-FAT4_1, Si-FAT4_2, Si-FAT4_3, Si-FAT4_4) were used to silence *FAT4* in three gastric cancer cell lines with wild-type *FAT4* gene (YCC16, HGC27, Hs746T). Western blotting was performed on samples after 48 hours of siRNA treatment. All experiments were performed a minimum of three independent times, and compared against a scrambled siRNA control (Si-Ctrl). Results are presented as mean ± standard error of the mean (SEM). **p*<0.05 compared with respective controls (Bonferroni post test). (A) YCC16, (B) HGC27, (C) Hs746T. *FAT4* silencing was observed to enhance cell proliferation in *FAT4* wild-type lines.

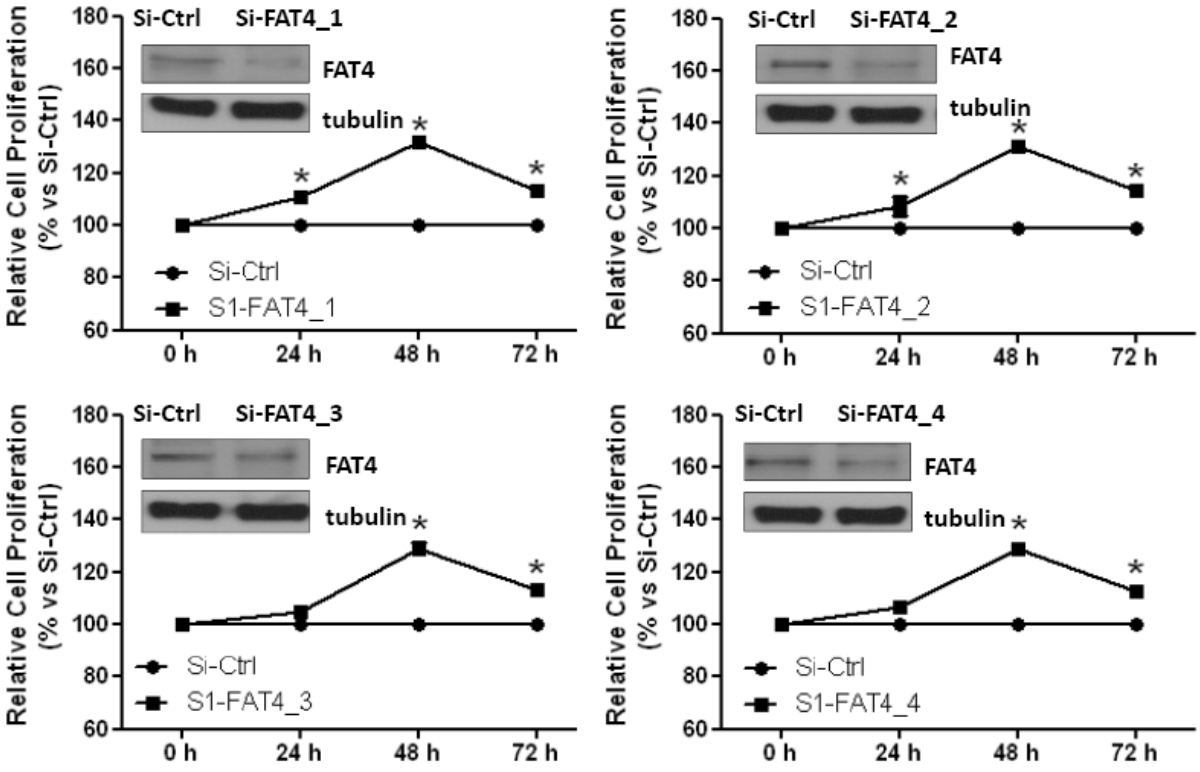


Nature Genetics: doi:10.1038/ng.2246

B) **HGC27**



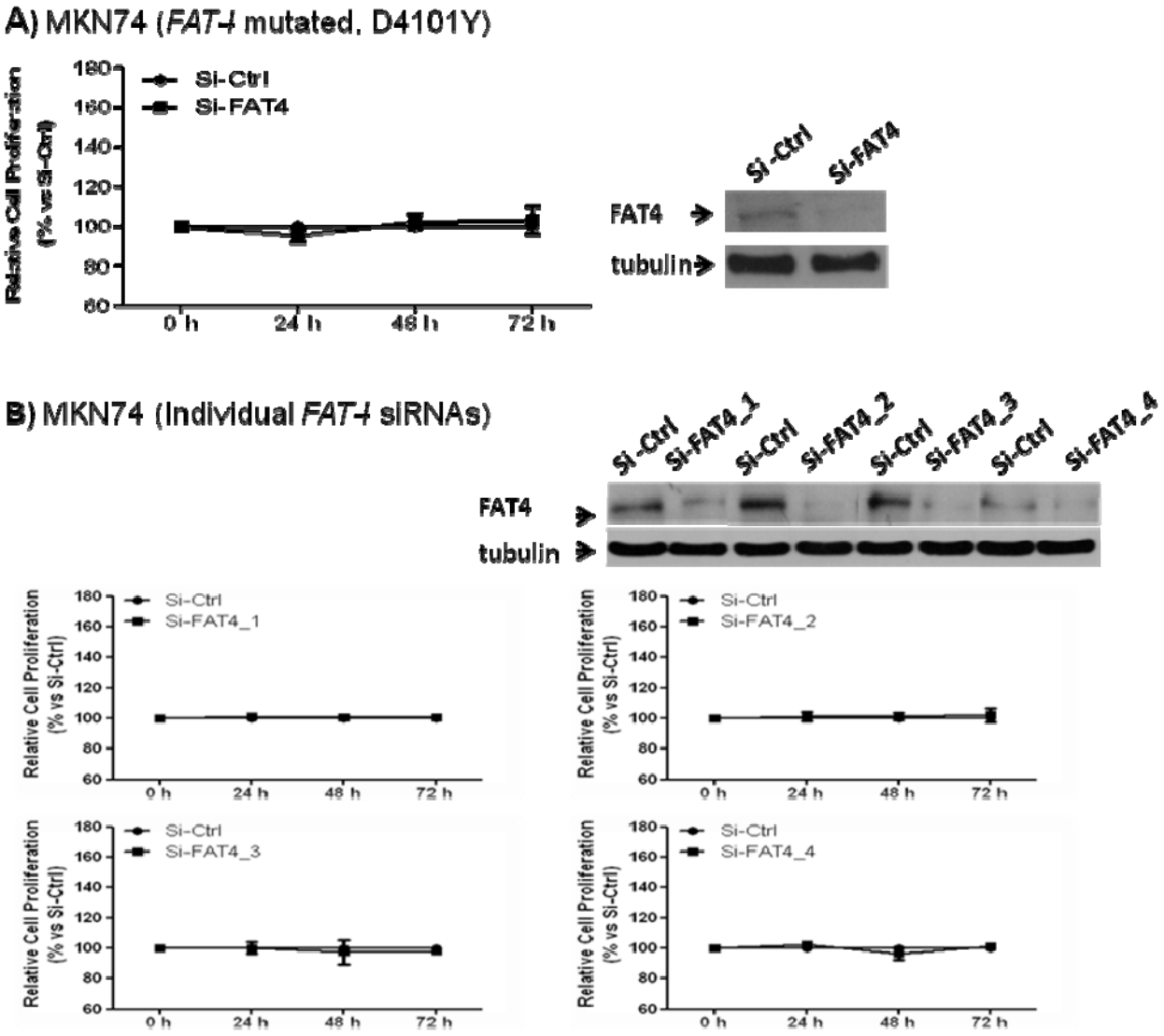
C) **HS746T**



Nature Genetics: doi:10.1038/ng.2246

Supplementary Figure 7

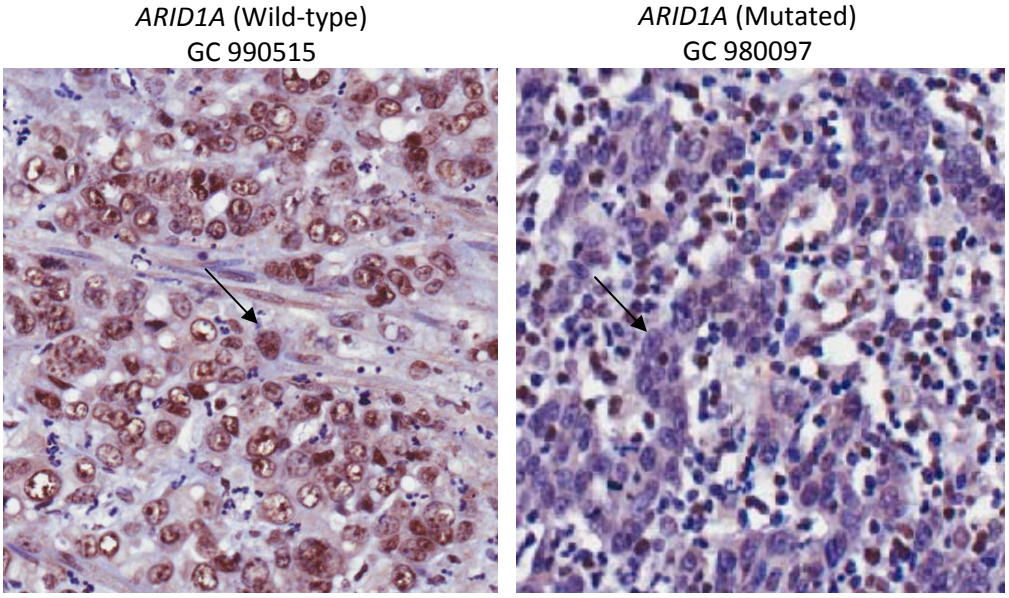
***FAT4* siRNA silencing in *FAT4*-mutated gastric cancer cells.** *FAT4* siRNAs were used to silence *FAT4* in MKN74 cells harboring a novel *FAT4* genetic alteration predicted to be "damaging" by Polyphen software. Pooled *FAT4* siRNAs were used in A), and multiple (n=4) individual siRNAs were used in B). Western blotting was performed on samples after 48 hrs of siRNA treatment. All experiments were performed a minimum of three independent times. **p*<0.05 compared with respective controls (Student's t-test). *FAT4* silencing in MKN74 did not alter cell proliferation significantly.



Nature Genetics: doi:10.1038/ng.2246

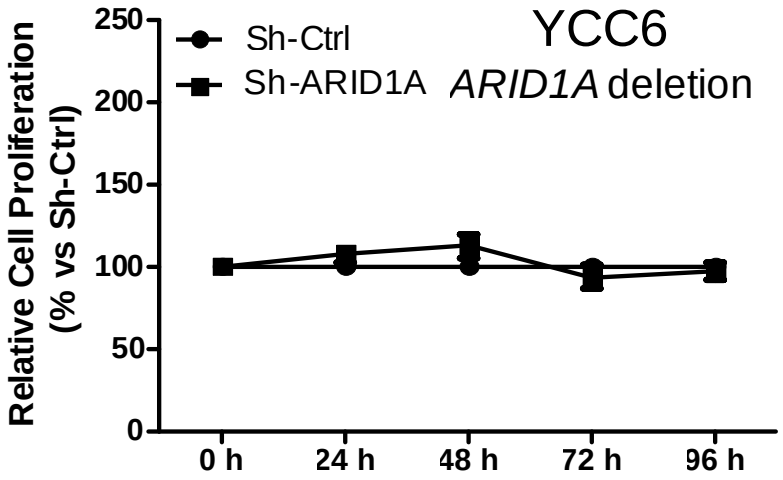
Supplementary Figure 8

ARID1A protein expression in *ARID1A* mutated and wild-type gastric cancers. (Left) Positive ARID1A expression in a *ARID1A*-wild-type gastric cancer (Tumor 990515). Strong nuclear immunoreactivity is observed in cancer cells (arrow). (Right) Loss of ARID1A expression in a *ARID1A*-mutated gastric cancer (Tumor 980097). Nuclear ARID1A immunereactivity is not observed in gastric cancer cells (arrow). However, ARID1A expression is still observed in lymphocytes, providing a positive control for the ARID1A immunohistochemistry staining protocol. Sections are shown at 200× magnification.



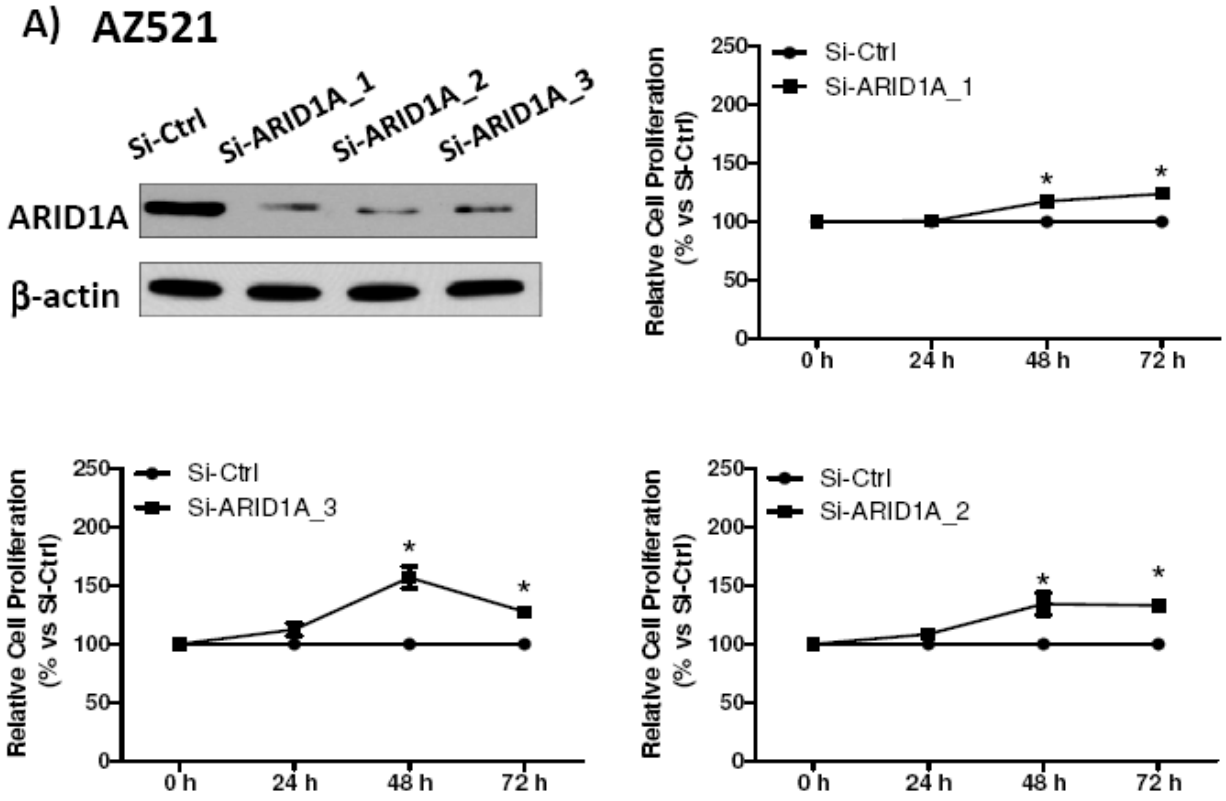
Supplementary Figure 9

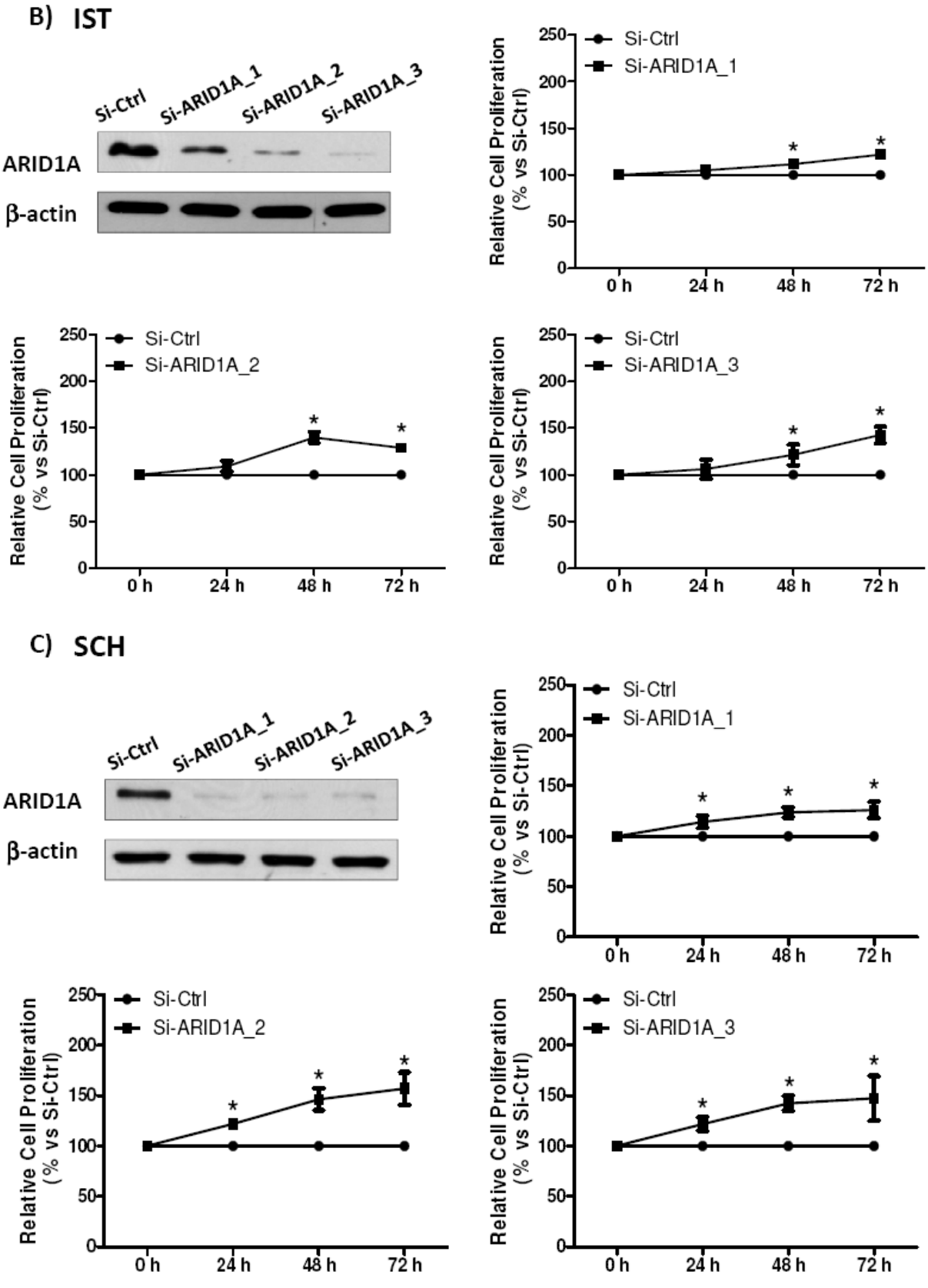
***ARID1A* shRNA-mediated silencing in YCC6 cells.** YCC6 cells exhibit genomic deletion of *ARID1A* and do not express ARID1A protein (see Figure 3b in Main Text). Treating YCC6 cells with *ARID1A* shRNAs does not elicit significant effects on cell proliferation.



Supplementary Figure 10

***ARID1A* silencing using individual *ARID1A* siRNAs.** Three independent non-overlapping *ARID1A* siRNAs (Si-ARID1A_1, Si-ARID1A_2, Si-ARID1A_3) were used to silence *ARID1A* in three gastric cancer cell lines with wild-type *ARID1A* genes. Western blotting was used to measure ARID1A protein in *ARID1A*-siRNA treated or scrambled-siRNA treated (Si-Ctrl) cells. Results are expressed as mean ± standard error of the mean (SEM) of at least three independent experiments. *p<0.05 (Bonferroni post test) compared with respective controls (i.e. Si-Ctrl). *ARID1A* silencing was observed to enhance cell proliferation in *ARID1A* wild-type line (A) AZ521, (B)IST1, (C) SCH cells.




B) IST
Si-Ctrl
Si-ARID1A_1
Si-ARID1A_2
Si-ARID1A_3
ARID1A
β-actin
Relative Cell Proliferation (% vs Si-Ctrl)
0 h
24 h
48 h
72 h
C) SCH

Supplementary Tables

Supplementary Table 1

Clinicopathologic features of the gastric adenocarcinomas used for exome sequencing

ID	Age (Y)	Ge	Site of Tumor	Stage (T)	Stage (N)	Stage (M)	Stage (AJC C6)	Grade	Lauren Classification	*H. pylori* Status	MSI status
990515	59	m	lesser curve	T3	N3	m0	4	moderately differentiated	intestinal	no	MSS
990475	70	m	antrum	T2a	N1	m0	2	well differentiated	intestinal	yes	MSS
990396	64	m	greater curve	T3	N2	m0	3B	poorly differentiated	diffuse	no	MSS
990300	50	fe	body	T3	N2	m0	3B	poorly differentiated	intestinal	no	MSS
990172	60	m	lesser curve	T2b	N2	m0	3A	poorly differentiated	intestinal	yes	MSS
990097	46	m	cardia	T3	N2	m0	3B	poorly differentiated	intestinal	no	MSS
990089	56	fe	incisura	T3	N2	m1	4	poorly differentiated	intestinal	yes	MSS
990005	60	m	greater curve	T3	N2	m0	3B	poorly differentiated	diffuse	no	MSS
98748381	67	m	lesser curve	T3	N2	m0	3B	poorly differentiated	mixed	yes	MSS
970010	72	m	unknown	T3	N1	m0	3A	moderately differentiated	intestinal	yes	MSS
76629543	92	m	pylorus	T3	N1	m0	3A	moderately differentiated	intestinal	yes	MSS
31231321	72	m	GE junction	T3	N3	m0	4	poorly differentiated	diffuse	yes	MSS
2000362	51	m	cardia	T3	N2	m0	3B	moderately differentiated	intestinal	no	MSS
TGH	77	m	unknown	T3	N2	M0	3B	moderately differentiated	intestinal	yes	MSI
TWH	77	fe	unknown	T2a	N0	M0	1B	moderately-differentiated	intestinal	No	MSS

Supplementary Table 2

Summary of Illumina high throughput sequencing run analysis for 15 gastric adenocarcinoma exome pairs.

Sample	# of lanes	# of targeted bases	Bases sequenced	Bases mapped to genome	Bases mapped to targeted regions	Avg # of reads per targeted base	Targeted bases with at least 10 reads (%)	Known SNPs in targeted regions
N2000362	1	37806033	6969949783	5993546924	2962684844	78.37	91.91	22702
T2000362	1	37806033	5952677564	5061464528	2374774990	62.81	89.04	21813
N970010	2	37806033	7620174864	6594141332	3514610856	92.96	93.68	23075
T970010	2	37806033	7966946116	6908151304	3511126514	92.87	93.27	21490
N98748381	2	37806033	8154209036	6868014296	3675685420	97.22	94.45	23381
T98748381	2	37806033	7997067668	6711668946	3673771187	97.17	94.01	23208
N990515	1	37806033	7112388101	6179862886	3282176362	86.82	92.56	22822
T990515	1	37806033	4971281902	4077149010	2224905393	58.85	87.63	20624
N990172	2	37806033	8822829080	7361206906	4112900499	108.79	93.98	23046
T990172	2	37806033	8885987340	7617146858	4080063737	107.92	94.26	22895
N990475	2	37806033	9158287952	7654141012	4217295619	111.55	94.04	23065
T990475	2	37806033	8922401556	7324436158	3899950839	103.16	94.15	23094
N76629543	2	37806033	8543062016	6838063388	3308377286	87.51	89.67	21881
T76629543	2	37806033	9014876544	7743622404	3895184824	103.03	92.82	20730
N990089	1	37806033	4723502530	4183591592	1866604323	49.37	86.83	21676
T990089	1	37806033	5487972740	4872029114	2029863030	53.69	84.38	21088
N990300	2	37806033	9493701388	7091387218	3421655930	90.51	86.95	21238
T990300	2	37806033	8616916384	7395259782	3664436634	96.93	93.96	23309
N31231321	1	37806033	7645168089	6926398191	3142873586	83.13	89.02	21585
T31231321	1	37806033	7537716552	6672370765	2941004599	77.79	91.36	21883
N990396	1	37806033	8578223997	7740824778	3517519956	93.04	93.1	22624
T990396	1	37806033	9824092296	7559966096	3651699490	96.59	95.13	23144
N990097	1	37806033	1.9477E+10	1.5342E+10	7404084400	195.84	98.95	24398
T990097	1	37806033	2.0252E+10	1.592E+10	7774890897	205.65	99.07	24449
N990005	1	37806033	1.8763E+10	1.447E+10	7345214551	194.29	99.19	24300
T990005	1	37806033	2.116E+10	1.6899E+10	8250976521	218.24	99.11	24317
NGH	from genome	37806033	3709526850	3694298900	1754099669	46.4	98.44	23670
TGH	from genome	37806033	2103769475	2093644500	964167935	25.5	95.69	22621
NWH	from genome	37806033	2908010325	2898412800	1340077599	35.45	95.66	23085
TWH	from genome	37806033	2596614725	2587745475	1222284205	32.33	96.81	23241

Supplementary Table 3

Non-synonymous somatic point mutations affecting exons or splice sites in 15 gastric adenocarcinomas. Coordinates are based on Hg18**.**

The table is submitted separately as an Excel file

Supplementary Table 4

Somatic small insertions or deletions (indels) affecting exons in 15 gastric adenocarcinomas. Coordinates are based on Hg18.

Genes	Confirmation by Sanger sequencing	Type	ID	Chr	Start	End	Affects Exon Bases
ACSBG2	Yes	+A	990005	chr19	6141638	6141638	Yes
APC	Yes	-T	990172	chr5	112205631	112205632	Yes
APC	Yes	-TGA	990172	chr5	112204790	112204793	Yes
ARID1A	Yes	+A	990097	chr1	26979722	26979722	Yes
C10orf68	Yes	-AAAT	970010	chr10	33057901	33057905	Yes
C6orf130	Yes	-A	TGH	chr6	41145764	41145765	Yes
C7orf63	Yes	-C	990396	chr7	89755519	89755520	Yes
C9orf128	Yes	-G	TGH	chr9	35816096	35816097	Yes
COIL	Yes	-T	TGH	chr17	52383014	52383015	Yes
DSCAM	Yes	-G	TWH	chr21	40307175	40307176	Yes
FBXO34	Yes	+T	TWH	chr14	54888307	54888307	Yes
IL15RA	Yes	-C	990097	chr10	6042368	6042369	Yes
KCTD3	Yes	-AA	990005	chr1	213860519	213860521	Yes
NLGN4Y	Yes	-A	990172	chrY	15451621	15451622	Yes
NUP210	Yes	-TG	76629543	chr3	13338709	13338711	Yes
PPRC1	Yes	+C	990097	chr10	103891066	103891066	Yes
PTEN	Yes	-A	TGH	chr10	89707749	89707750	Yes
PZP	Yes	+G	98748381	chr12	9246203	9246203	Yes
ROCK2	Yes	-AT	990515	chr2	11344163	11344165	Yes
RUNX1T1	Yes	-CTTT	970010	chr8	93057311	93057315	Yes
SHROOM3	Yes	+C	98748381	chr4	77879182	77879182	Yes
SLC9A11	Yes	-T	990005	chr1	171759835	171759836	Yes
SLITRK3	Yes	-C	98748381	chr3	166389349	166389350	Yes
TP53	Yes	-G	990515	chr17	7520075	7520076	Yes
TP53	Yes	-G	76629543	chr17	7520440	7520441	Yes
USP11	Yes	+C	TGH	chrX	46992223	46992223	Yes
USP21	Yes	-TCTGGCA	98748381	chr1	159399052	159399059	Yes
VPS13B	Yes	-A	98748381	chr8	100901372	100901373	Yes
ZNF202	Yes	+T	31231321	chr11	123104160	123104160	Yes
ZNF527	Yes	-CA	76629543	chr19	42571692	42571694	Yes

Supplementary Table 5

Numbers of non-silent (nonsense, missense, splice site somatic mutations, indels) and synonymous mutations detected in 15 gastric adenocarcinomas.

	Nonsense	Splice	Missense	Sub-total	Indel	Total (Non-Silent)	Nonsyn[a] Somatic Mutations per Screened Mb	Syn Mutations	Syn Somatic Mutations per Screened Mb
970010	3	3	84	90	2	92	2.514	39	1.127
990005	1	1	7	9	3	12	0.217	1	0.027
990089	3	1	16	20	0	20	0.602	6	0.19
990097	0	0	7	7	3	10	0.191	2	0.054
990172	1	2	26	29	3	32	0.777	6	0.173
990300	0	2	16	18	0	18	0.497	7	0.218
990396	2	1	15	18	1	19	0.495	6	0.175
990475	2	0	31	33	0	33	0.949	6	0.172
990515	3	8	103	114	2	116	3.124	30	0.893
2000362	4	1	35	40	0	40	1.16	18	0.535
31231321	5	7	75	87	1	88	2.442	21	0.641
76629543	5	6	121	132	3	135	3.803	64	1.932
98748381	3	7	31	41	5	46	0.974	9	0.258
TGH	5	0	43	48	5	53	1.331	17	0.471
TWH	1	0	31	32	2	34	0.917	9	0.258
Total	**38**	**39**	**641**	**718**	**30**	**748**		**241**	
Average	**2.5**	**2.6**	**42.7**	**47.8**	**2**	**49.8**	**1.333**	**16.1**	**0.475**

[a] Nonsense plus missense

Supplementary Table 6

Recurrently mutated genes found from exome sequencing in 15 gastric adenocarcinomas.

Genes	# of samples bearing mut	# of somatic non-silent mut	# of somatic silent mut	Ratio of Non-syn/syn	Known function of the genes	Previously Implicated in gastric cancer? (Pubmed ID is listed)
TP53	11	11	0	NA	Tumor suppressor gene	PMID:2032245
ARID1A	3	3	0	NA	SWI/SNF component, chromatin remodeling	This study
PIK3CA	3	3	0	NA	Phosphatidylinositol 3-kinase, catalytic subunit	PMID:15784156; PMID:15994075
PKHD1	3	3	0	NA	Receptor protein that might play a cytoskeleton and adhesion role	This study
TTN	3	5	0	NA	Key component in the assembly and functioning of vertebrate striated muscles	This study
ABCC9	2	2	0	NA	ATP-sensitive potassium channel in vascular and non-vascular smooth muscle	This study
APOB	2	2	0	NA	Functions as a recognition signal for cellular binding and internalization	This study
ATP4A	2	2	0	NA	Gastric H+, K+-ATPase that maintains an acidic environment within the stomach	PMID:16048577; PMID:10764766
BAI3	2	2	0	NA	G-protein coupled receptor	This study
C1orf168	2	2	0	NA	Function is not known	This study
CCNL1	2	2	0	NA	Pre-mRNA splicing processes	This study
CNTN1	2	2	0	NA	Cell adhesion molecule, a ligand of NOTCH1	This study
CSMD3	2	4	1	4	Function is not known	This study
CTNNB1	2	2	0	NA	regulation of cell adhesion and Wnt signaling	PMID:21541555
DBR1	2	3	0	NA	RNA lariat debranching enzyme	This study
DMD	2	2	2	1	Anchors extracellular matrix to the cytoskeleton via F-actin	This study
ERCC6	2	2	0	NA	Transcription-coupled nucleotide excision repair	This study
FAT4	2	3	0	NA	Cadherin family member. Involved in the non-canonical Wnt Signaling pathway	This study
FMN2	2	2	0	NA	Cytoskeletal organization and cell polarity	This study
HTR1E	2	2	0	NA	Receptor for 5-hydroxytryptamine	This study
LRP1B	2	4	0	NA	Cell surface protein that internalizes ligands	PMID:20095042
MLL3	2	2	0	NA	Histone methyltransferase	This study
OR4C15	2	2	1	2	Odorant receptor	This study
OR4C46	2	2	0	NA	Odorant receptor	This study
RIT2	2	3	0	NA	Binds and exchanges GTP and GDP	This study
SHROOM3	2	2	0	NA	Involved in regulating cell shape	This study
SPTA1	2	2	1	2	Cytoskeletal component of RBC plasma membrane	This study
ZSWIM3	2	2	0	NA	Function is not known	This study
ZZEF1	2	2	0	NA	Function is not known	This study

Supplementary Table 7

Ranking of genes by probability that they are affected by more somatic mutations than expected based on the background mutation rate. NS = non-synonymous single nucleotide substitutions. P=p value by Genome MuSiC's likelihood-ratio test (see **Online Methods**). FDR=false discovery rate. Genes in **Supplementary Table 6** are highlighted in orange. Genes in the GO "cell adhesion" category (GO: 0007155) are marked with *.

Rank	Gene	Rank among Recurrently Mutated Genes (Sup. Table 6)	Rank among Recurrently Mutated Cell Adhesion genes	Mutation Counts		P	FDR
				NS	Indels		
1	*TP53*	*1*		9	2	0.0000	0.0000
2	*DBR1*	*2*		3	0	2.40E-06	0.0198
3	*RIT2*	*3*		2	0	4.22E-06	0.0233
4	*CCNL1*	*4*		2	0	8.67E-06	0.0359
5	*APC**			0	2	1.60E-05	0.0422
6	*ATF5*			2	0	1.64E-05	0.0422
7	*PKHD1**	*5*	1	3	0	1.79E-05	0.0422
8	*ADAM18*			2	0	2.22E-05	0.0460
9	*HTR1E*	*6*		2	0	2.65E-05	0.0487
10	*ARID1A*	*7*		2	1	4.14E-05	0.0685
11	*NFAT5*			2	0	6.04E-05	0.0908
12	*PPP1R1B*			1	0	0.0001	0.1481
13	*ZNF193*			1	0	0.0001	0.1481
14	*CERK*			2	0	0.0001	0.1481
15	*OR4C46*	*8*		2	0	0.0001	0.1578
16	*OR4C15*	*9*		2	0	0.0002	0.1592
17	*PIK3CA*	*10*		3	0	0.0002	0.1604
35	*APOB*	*11*		2	0	0.0008	0.3225
37	*SHROOM3*	*12*		1	1	0.0009	0.3225
40	*NLGN4Y**			0	1	0.0009	0.3225
42	*LRP1B*	*13*		4	0	0.0010	0.3225
47	*ZSWIM3*	*14*		2	0	0.0010	0.3225
51	*C1orf168*	*15*		2	0	0.0012	0.3225
52	*CTNNB1**	*16*	2	2	0	0.0012	0.3225
63	*TNC**			1	0	0.0014	0.3225
67	*CNTN1**	*17*	3	2	0	0.0015	0.3225
73	*FAT4**	*18*	4	3	0	0.0016	0.3225
96	*ERCC6*	*19*		2	0	0.0021	0.3225
116	*ATP4A*	*20*		2	0	0.0026	0.3225
124	*FMN2*	*21*		2	0	0.0027	0.3225
127	*BAI3*	*22*		2	0	0.0028	0.3225
233	*ABCC9*	*23*		2	0	0.0048	0.3225
261	*CSMD3*	*24*		3	0	0.0052	0.3225
293	*SPTA1*	*25*		2	0	0.0058	0.3241
342	*ZZEF1*	*26*		2	0	0.0068	0.3269
467	*DMD*	*27*		2	0	0.0110	0.3888
528	*TTN*	*28*		2	0	0.0146	0.4566
671	*MLL3*	*29*		1	0	0.0558	1.0000
...							

Nature Genetics: doi:10.1038/ng.2246

~16,000 additional genes, most with no somatic mutations
One *MLL3* variant in **Supplementary Table 6** is in a splice site and not analyzed here. For *TTN*, *CSMD3*, and *RIT2*, **Supplementary Table 6** includes variants that are not included in the analysis here, which only considers the longest transcript of each gene.

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 8

Summary of pathway analysis. Biological terms associated with cell adhesion were found to be the most significantly enriched pathways, using DAVID Bioinformatics Resources 6.7 (http://david.abcc.ncifcrf.gov/). P-values were corrected for multiple hypothesis testing using the Benjamini method (second last column). Only annotation clusters passing a Benjamini-Hochberg p-value of <0.05 are listed.

Category	Term	Count	%	P-Value	Genes	List Total	PoP Hits	Pop Total	Fold Enrich-ment	Bonfer-roni	Benja-mini	FDR
GOTERM_BP_FAT	GO:0007155~cell adhesion	58	8.46	1.77E-07	POSTN, SDC3, CTNNB1, PCDHGA1, CDH20, COL12A1, ROBO2, COL11A1, DSCAM, RET, MAGI1, PTPRF, PCDH11Y, PCDHB2, LEF1, PTPRT, NRXN1, PCDH7, CTNNA2, SIGLEC1, CNTN1, ADAM17, RELN, DST, ROM1, DCHS2, ADAMTS13, PKHD1, TNC, COL3A1, DSCAML1, PCDHGC4, PCDHGC3, CDH4, PCDHAC2, LAMB4, CDH7, LGALS3BP, COL17A1, COL7A1, FAT4, COL27A1, COL6A3, PKD2, APC, LGALS4, NLGN1, NID1, ITGA4, COL5A3, PRPH2, PCDH18, DSG4, TSC1, NLGN4Y, DSG2, LAMA5, CD209, TROAP	538	700	13528	2.08	4.47E-04	4.47E-04	3.12E-04
GOTERM_BP_FAT	GO:0022610~biological adhesion	58	8.46	1.81E-07	POSTN, SDC3, CTNNB1, PCDHGA1, CDH20, COL12A1, ROBO2, COL11A1, DSCAM, RET, MAGI1, PTPRF, PCDH11Y, PCDHB2, LEF1, PTPRT, NRXN1, PCDH7, CTNNA2, SIGLEC1, CNTN1, ADAM17, RELN, DST, ROM1, DCHS2, ADAMTS13, PKHD1, TNC, COL3A1, DSCAML1, PCDHGC4, PCDHGC3, CDH4, PCDHAC2, LAMB4, CDH7, LGALS3BP, COL17A1, COL7A1, FAT4, COL27A1, COL6A3, PKD2, APC, LGALS4, NLGN1, NID1, ITGA4, COL5A3, PRPH2, PCDH18, DSG4, TSC1, NLGN4Y, DSG2, LAMA5, CD209, TROAP	538	701	13528	2.08	4.59E-04	2.30E-04	3.21E-04
GOTERM_BP_FAT	GO:0016337~cell-cell adhesion	27	3.94	4.07E-05	DCHS2, PKHD1, DSCAML1, PCDHGC4, PCDHGC3, CDH4, PCDHAC2, PCDHGA1, CTNNB1, CDH7, CDH20, FAT4, ROBO2, COL11A1, RET, PTPRF, PCDH11Y, NLGN1, PCDHB2, PTPRT, ITGA4, PCDH7, PCDH18, CTNNA2, SIGLEC1, DSG4, DSG2, CD209	538	276	13528	2.46	0.098	0.009	0.072

Nature Genetics: doi:10.1038/ng.2246

Category	Term	Count	%	P-Value	Genes	List Total	Pop Hits	Pop Total	Fold Enrich-ment	Bonfer-roni	Benja-mini	FDR
GOTERM_BP_FAT	GO:0000904~cell morphogenesis involved in differentiation	27	3.94	4.59E-06	DCC, NDN, ERBB3, PKHD1, DSCAML1, CDH4, CTNNB1, MYCBP2, HOXA2, CXCR4, ROBO2, SEMA3A, DCLK1, DSCAM, APC, NTF3, LEF1, NRXN1, CTNNA2, SLITRK2, SLITRK3, LAMA5, NOTCH4, CYFIP1, RELN, DST, KALRN	538	244	13528	2.78	0.012	0.003	0.008
GOTERM_BP_FAT	GO:0031175~neuron projection development	27	3.94	1.09E-05	DCC, NDN, ERBB3, PKHD1, DSCAML1, CDH4, PTEN, MYCBP2, HOXA2, CXCR4, DMD, ROBO2, SEMA3A, DCLK1, DSCAM, APC, GNAO1, NTF3, NRXN1, CTNNA2, SLITRK2, SLITRK3, MAP2, CYFIP1, RELN, DST, KALRN	538	256	13528	2.65	0.027	0.005	0.019
GOTERM_BP_FAT	GO:0048812~neuron projection morphogenesis	24	3.50	1.30E-05	DCC, NDN, NTF3, ERBB3, PKHD1, DSCAML1, NRXN1, CDH4, MYCBP2, CTNNA2, SLITRK2, HOXA2, SLITRK3, CXCR4, DMD, CYFIP1, ROBO2, RELN, SEMA3A, DST, DCLK1, DSCAM, APC, KALRN	538	213	13528	2.83	0.032	0.005	0.023
GOTERM_BP_FAT	GO:0000902~cell morphogenesis	33	4.81	1.40E-05	DCC, SHROOM3, NDN, PKHD1, ERBB3, MAEL, DSCAML1, CDH4, MYCBP2, CTNNB1, HOXA2, ANK1, CXCR4, DMD, ROBO2, SEMA3A, DCLK1, IDUA, DSCAM, APC, NTF3, LEF1, NRXN1, CTNNA2, SLITRK2, SLITRK3, LAMA5, PRICKLE2, NOTCH4, CYFIP1, RELN, DST, KALRN	538	356	13528	2.33	0.035	0.005	0.025
GOTERM_BP_FAT	GO:0032989~cellular component morphogenesis	35	5.10	2.09E-05	DCC, SHROOM3, NDN, PKHD1, ERBB3, MAEL, DSCAML1, TTN, CDH4, MYCBP2, CTNNB1, HOXA2, ANK1, CXCR4, DMD, ROBO2, SEMA3A, DCLK1, IDUA, DSCAM, APC, NTF3, LEF1, MYH6, NRXN1, CTNNA2, SLITRK2, SLITRK3, LAMA5, PRICKLE2, NOTCH4, CYFIP1, RELN, DST, KALRN	538	397	13528	2.21	0.051	0.007	0.037
GOTERM_BP_FAT	GO:0030030~cell projection organization	33	4.81	2.73E-05	DCC, DNAH9, NDN, PKHD1, ERBB3, DSCAML1, PTEN, CDH4, MYCBP2, HOXA2, CXCR4, DMD, ROBO2, SEMA3A, DCLK1, FGD4, DSCAM, APC, MYO1A, GNAO1, NTF3, NRXN1, VAV2, CTNNA2, SLITRK2, SLITRK3, TSC1, LAMA5, MAP2, CYFIP1, RELN, DST, KALRN	538	368	13528	2.25	0.067	0.008	0.048
GOTERM_BP_FAT	GO:0007409~axonogenesis	22	3.21	2.74E-05	DCC, NDN, NTF3, ERBB3, PKHD1, DSCAML1, NRXN1, CDH4, MYCBP2, CTNNA2, SLITRK2, HOXA2, SLITRK3, CXCR4, CYFIP1, ROBO2, RELN, SEMA3A, DST, DCLK1, APC, KALRN	538	193	13528	2.86	0.067	0.007	0.048
GOTERM_BP_FAT	GO:0048667~cell morphogenesis involved in neuron differentiation	23	3.35	2.97E-05	DCC, NDN, NTF3, ERBB3, PKHD1, DSCAML1, NRXN1, CDH4, MYCBP2, CTNNA2, SLITRK2, HOXA2, SLITRK3, CXCR4, CYFIP1, ROBO2, RELN, SEMA3A, DST, DCLK1, DSCAM, APC, KALRN	538	209	13528	2.76	0.073	0.007	0.053

Nature Genetics: doi:10.1038/ng.2246

Category	Term	Count	%	P-Value	Genes	List Total	Pop Hits	Pop Total	Fold Enrich-ment	Bonfer-roni	Benja-mini	FDR
GOTERM_BP_FAT	GO:0048858~cell projection morphogenesis	25	3.65	4.28E-05	DCC, NDN, ERBB3, PKHD1, DSCAML1, CDH4, MYCBP2, HOXA2, CXCR4, DMD, ROBO2, SEMA3A, DCLK1, DSCAM, APC, NTF3, NRXN1, CTNNA2, SLITRK2, SLITRK3, LAMA5, CYFIP1, RELN, DST, KALRN	538	245	13528	2.56	0.103	0.008	0.076
GOTERM_BP_FAT	GO:0032990~cell part morphogenesis	25	3.65	8.55E-05	DCC, NDN, ERBB3, PKHD1, DSCAML1, CDH4, MYCBP2, HOXA2, CXCR4, DMD, ROBO2, SEMA3A, DCLK1, DSCAM, APC, NTF3, NRXN1, CTNNA2, SLITRK2, SLITRK3, LAMA5, CYFIP1, RELN, DST, KALRN	538	256	13528	2.45	0.195	0.015	0.151
GOTERM_BP_FAT	GO:0048666~neuron development	29	4.23	2.06E-04	DCC, NDN, ERBB3, PKHD1, DSCAML1, CDH4, PTEN, MYCBP2, HOXA2, CXCR4, DMD, ROBO2, SEMA3A, LHX8, DCLK1, DSCAM, APC, RET, GNAO1, NTF3, NRXN1, CTNNA2, SLITRK2, SLITRK3, MAP2, CYFIP1, RELN, DST, KALRN	538	339	13528	2.15	0.407	0.032	0.364

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 9

Integrated table of clinicopathologic data of gastric cancers analyzed in this study.

ID	Age (Y)	Gen	Site of Tumor	Stage T	Stage N	Stage M	Stage (AJCC6)	Grade	Lauren Classifi-cation	*H. pylori* Status	*TP53 Mut*	*PIK3CA Mut*	*FAT4 Mut*	*ARID1A Mut*	MSI status	Seque-ncing
990515	59	M	lesser curve	T3	N3	m0	4	moderately differentiated	intestinal	no	c.336delC	WT	Q4872H, P4454L	WT	MSS	Exome
990475	70	M	antrum	T2a	N1	m0	2	well differentiated	intestinal	yes	V272M	WT	WT	WT	MSS	Exome
990396	64	M	greater curve	T3	N2	m0	3B	poorly differentiated	diffuse	no	F134C	WT	WT	WT	MSS	Exome
990300	50	F	body	T3	N2	m0	3B	poorly differentiated	intestinal	no	R248Q	WT	WT	WT	MSS	Exome
990172	60	M	lesser curve	T2b	N2	m0	3A	poorly differentiated	intestinal	yes	V272L	WT	WT	WT	MSS	Exome
990097	46	M	CARDIA	T3	N2	m0	3B	poorly differentiated	intestinal	no	WT	E545K	WT	c.6748insA	MSS	Exome
990089	56	F	incisura	T3	N2	m1	4	poorly differentiated	intestinal	yes	R213*	WT	WT	WT	MSS	Exome
990005	60	M	greater curve	T3	N2	m0	3B	poorly differentiated	diffuse	no	WT	WT	WT	R1276*	MSS	Exome
98748381	67	M	lesser curve	T3	N2	m0	3B	poorly differentiated	mixed	yes	R248Q	WT	WT	WT	MSS	Exome
970010	72	M	unknown	T3	N1	m0	3A	moderately differentiated	intestinal	yes	R280S	T1052K	L1062R	WT	MSS	Exome
76629543	92	M	PYLORUS	T3	N1	m0	3A	moderately differentiated	intestinal	yes	c.80delC	WT	WT	WT	MSS	Exome
31231321	72	M	GE junction	T3	N3	m0	4	poorly differentiated	diffuse	yes	R273C	WT	WT	WT	MSS	Exome
2000362	51	M	CARDIA	T3	N2	m0	3B	moderately differentiated	intestinal	no	WT	WT	WT	WT	MSS	Exome
TGH	77	M	unknown	T3	N2	M0	3B	moderately differentiated	intestinal	yes	WT	R93Q	WT	R1721*	MSI	Exome
TWH	77	F	unknown	T2a	N0	M0	1B	moderately-differentiated	intestinal	No	R273C	WT	WT	WT	MSS	Exome
990424	70	M	antrum	T2b	N1	m0	2	moderately differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
990413	67	M	greater curve	T2a	N1	m0	2	poorly differentiated	diffuse	no	WT	WT	WT	WT	MSS	Sanger
990412	64	M	greater curve	T3	N1	m0	3A	poorly differentiated	intestinal	no	R196*	Unknown	WT	WT	MSS	Sanger
990355	34	F	incisura	T3	N3	m1	4	poorly differentiated	diffuse	yes	WT	WT	WT	WT	MSS	Sanger
990275	71	M	lesser curve	T3	N0	m0	2	moderately differentiated	intestinal	no	Y163C	Unknown	WT	WT	MSS	Sanger

Nature Genetics: doi:10.1038/ng.2246

ID	Age (Y)	Gen	Site of Tumor	Stage T	Stage N	Stage M	Stage (AJCC6)	Grade	Lauren Classification	*H. pylori* Status	*TP53 Mut*	*PIK3CA Mut*	*FAT4 Mut*	*ARID1A Mut*	MSI status	Sequencing
990228	40	F	lesser curve	T3	N1	m0	3A	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
990205	72	F	lesser curve	T2b	N1	m0	2	poorly differentiated	intestinal	no	WT	Unknown	WT	c.400delG; c.3216delA	MSI	Sanger
990203	79	M	CARDIA	T3	N1	m0	3A	poorly differentiated	intestinal	no	R248W	Unknown	WT	WT	MSS	Sanger
990195	31	M	antrum	T3	N1	m1	4	moderately differentiated	intestinal	no	c.529_546 delCCCCACCATGAGCGCTGC	Unknown	WT	WT	MSS	Sanger
990170	41	F	body	T3	N1	m0	3A	poorly differentiated	intestinal	no	WT	Unknown	WT	WT	MSS	Sanger
990129	77	F	greater curve	T3	N1	m0	3A	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
990119	57	M	greater curve	T3	N1	m0	3A	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
990111	74	F	lesser curve	T1	N0	m0	1A	moderately differentiated	intestinal	no	WT	Unknown	C4479F	WT	MSI	Sanger
990108	65	M	body	T2b	N1	m0	2	moderately differentiated	intestinal	yes	c.511del G	Unknown	WT	WT	MSS	Sanger
990090	64	M	lesser curve	T3	N1	m0	3A	moderately differentiated	intestinal	no	S106R	Unknown	WT	WT	MSS	Sanger
990073	59	M	greater curve	T3	N2	m0	3B	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
990071	71	F	anastomosis	T4	N0	m0	4	moderately differentiated	intestinal	no	WT	WT	WT	WT	MSS	Sanger
990070	43	M	lesser curve	T3	N3	m0	4	poorly differentiated	diffuse	yes	WT	Unknown	WT	WT	MSS	Sanger
990068	74	M	body	T3	N1	m0	3A	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
990060	26	F	body	T3	N1	m0	3A	poorly differentiated	diffuse	yes	WT	Unknown	WT	WT	MSS	Sanger
990046	56	M	antrum	T3	N2	m0	3B	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
990044	69	M	greater curve	T3	N2	m0	3B	moderately differentiated	intestinal	yes	WT	WT	WT	WT	MSS	Sanger
990041	41	M	GE junction	T4	N1	m0	4	poorly differentiated	intestinal	yes	WT	WT	WT	WT	MSS	Sanger
990024	68	M	lesser curve	T3	N1	m0	3A	moderately differentiated	intestinal	yes	WT	R88Q	WT	c.1650delC; c.3978InsC	MSI	Sanger
990015	61	M	lesser curve	T1	N0	m0	1A	moderately differentiated	intestinal	no	WT	Unknown	WT	WT	MSS	Sanger
990010	70	M	body	T4	N3	m1	4	poorly differentiated	intestinal	no	H179R	WT	WT	WT	MSS	Sanger
980447	68	M	lesser curve	T3	N3	m0	4	poorly differentiated	intestinal	no	c.545_549del GCTCA	Unknown	WT	WT	MSS	Sanger
980442	61	F	unknown	T2a	N1	m0	2	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
980437	67	F	incisura	T3	N3	m0	4	poorly differentiated	intestinal	no	c.250_262 delGCCCCTGCACCAG	WT	WT	WT	MSS	Sanger

Nature Genetics: doi:10.1038/ng.2246

ID	Age (Y)	Gen	Site of Tumor	Stage T	Stage N	Stage M	Stage (AJCC6)	Grade	Lauren Classifi-cation	*H. pylori* Status	*TP53 Mut*	*PIK3CA Mut*	*FAT4 Mut*	*ARID1A Mut*	MSI status	Seque-ncing
980436	64	F	lesser curve	T3	N1	m0	3A	moderately differentiated	intestinal	no	WT	E545K	WT	WT	MSS	Sanger
980418	87	M	greater curve	T3	N1	m0	3A	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSI	Sanger
980401	83	F	unknown	T3	N1	m0	3A	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
980390	77	F	unknown	T1	N1	m0	1B	moderately differentiated	intestinal	no	P219L; Y220C	Unknown	WT	WT	MSS	Sanger
980386	76	F	CARDIA	T3	N2	m0	3B	poorly differentiated	diffuse	yes	WT	Unknown	WT	WT	MSS	Sanger
980369	46	F	body	T2a	N0	m0	1B	poorly differentiated	diffuse	no	WT	WT	WT	WT	MSS	Sanger
980344	69	F	unknown	T3	N1	m1	4	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
980327	69	F	unknown	T3	N1	m0	3A	poorly differentiated	mixed	no	WT	Unknown	WT	WT	MSS	Sanger
980319	67	M	unknown	T3	N1	m0	3A	poorly differentiated	mixed	yes	WT	WT	WT	WT	MSS	Sanger
980305	39	F	unknown	T3	N3	m0	4	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
980252	66	M	body	T3	N1	m1	4	poorly differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
980251	53	F	unknown	T1	N0	m0	1A	poorly differentiated	diffuse	yes	A74T	Unknown	WT	WT	MSS	Sanger
980211	63	M	unknown	T3	N2	m0	3B	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
980161	66	F	lesser curve	T4	N1	m0	4	moderately differentiated	intestinal	no	WT	E545K	WT	c.5547 insG	MSS	Sanger
980097	65	M	unknown	T2a	N1	m0	2	undifferentiated	mixed	no	WT	Unknown	WT	Q812*	MSS	Sanger
980095	49	F	CARDIA	T4	N3	m0	4	poorly differentiated	diffuse	no	WT	WT	WT	WT	MSS	Sanger
980048	46	F	unknown	T3	N0	m1	4	poorly differentiated	diffuse	yes	WT	Unknown	WT	WT	MSS	Sanger
980029	76	M	unknown	T3	N1	m0	3A	poorly differentiated	diffuse	no	Y220C	WT	WT	WT	MSS	Sanger
980028	61	M	unknown	T2a	N0	m0	1B	moderately differentiated	intestinal	no	WT	Unknown	WT	c.3999-4001 delGCA	MSI	Sanger
980025	80	M	unknown	T3	N3	m0	4	poorly differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
980021	59	M	unknown	T3	N1	m0	3A	moderately differentiated	intestinal	no	Y234C	Unknown	WT	WT	MSS	Sanger
980011	74	M	unknown	T4	N1	m0	4	poorly differentiated	intestinal	no	WT	E545K	WT	WT	MSS	Sanger
970017	57	F	body	T3	N0	m0	2	poorly differentiated	diffuse	yes	WT	WT	WT	WT	MSS	Sanger
970005	68	M	GE junction	T4	N2	m0	4	moderately differentiated	intestinal	no	c.445delT	Unknown	WT	WT	MSS	Sanger

Nature Genetics: doi:10.1038/ng.2246

ID	Age (Y)	Gen	Site of Tumor	Stage T	Stage N	Stage M	Stage (AJCC6)	Grade	Lauren Classi-fication	*H. pylori* Status	*TP53 Mut*	*PIK3CA Mut*	*FAT4 Mut*	*ARID1A Mut*	MSI status	Seque-ncing
87622942	56	M	CARDIA	T4	N3	m0	4	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
82973565	69	M	body	T3	N1	m0	3A	poorly differentiated	mixed	no	WT	WT	WT	WT	MSS	Sanger
77263387	75	M	CARDIA	T2b	N1	m0	2	unknown	mixed	yes	N239T	Unknown	WT	WT	MSS	Sanger
73291145	48	F	antrum	T2b	N1	m0	2	well differentiated	intestinal	yes	I195T	WT	WT	WT	MSS	Sanger
66811693	68	M	lesser curve	T1	N1	m0	1B	moderately differentiated	intestinal	no	R175H	WT	WT	WT	MSS	Sanger
57701999	88	M	GE junction	T3	N1	m0	3A	moderately differentiated	intestinal	no	c.719_721 insACA	Unknown	WT	WT	MSS	Sanger
57689477	84	F	greater curve	T1	N0	m0	1A	moderately differentiated	intestinal	no	R248Q	WT	WT	WT	MSS	Sanger
38877042	65	M	lesser curve	T3	N1	m0	3A	poorly differentiated	diffuse	no	WT	WT	WT	WT	MSS	Sanger
31661621	84	F	lesser curve	T3	N2	m0	3B	moderately differentiated	intestinal	yes	c.456_457 insC	Unknown	WT	WT	MSS	Sanger
21080055	54	M	lesser curve	T2b	N1	m0	2	poorly differentiated	intestinal	yes	WT	WT	WT	WT	MSS	Sanger
20263644	64	F	lesser curve	T2b	N1	m0	2	poorly differentiated	diffuse	yes	WT	Unknown	WT	WT	MSS	Sanger
20021007	53	M	GE junction	T2b	N0	m0	1B	poorly differentiated	intestinal	no	H179Y	Unknown	WT	WT	MSS	Sanger
20020846	45	M	lesser curve	T2b	N1	m0	2	poorly differentiated	diffuse	yes	WT	Unknown	WT	WT	MSS	Sanger
20020455	81	F	body	T2b	N1	m1	4	moderately differentiated	intestinal	no	WT	Unknown	WT	WT	MSS	Sanger
2001190	67	F	body	T3	N2	m0	3B	poorly differentiated	diffuse	no	WT	Q546K	WT	WT	MSS	Sanger
2001123	56	M	CARDIA	T3	N2	m0	3B	moderately differentiated	intestinal	no	WT	Unknown	WT	WT	MSS	Sanger
2001086	79	M	antrum	T3	N1	m0	3A	poorly differentiated	diffuse	yes	WT	Unknown	WT	WT	MSS	Sanger
2000892	71	F	lesser curve	T2a	N1	m0	2	moderately differentiated	intestinal	no	WT	Unknown	G10S	WT	MSS	Sanger
2000877	45	M	CARDIA	T2a	N1	m0	2	poorly differentiated	intestinal	no	R175H	WT	WT	WT	MSS	Sanger
2000696	79	F	greater curve	T1	N0	m0	1A	moderately differentiated	intestinal	yes	WT	WT	WT	WT	MSS	Sanger
2000676	84	F	antrum	T3	N1	m0	3A	moderately differentiated	intestinal	yes	C242Y	WT	WT	WT	MSI	Sanger
2000618	67	M	lesser curve	T2a	N1	m0	2	moderately differentiated	intestinal	no	WT	Unknown	WT	WT	MSS	Sanger
2000529	76	M	lesser curve	T3	N2	m1	4	moderately differentiated	intestinal	yes	R213*	Unknown	A4485V	WT	MSS	Sanger
2000521	66	F	greater curve	T3	N3	m0	4	poorly differentiated	diffuse	no	WT	WT	WT	WT	MSS	Sanger

Nature Genetics: doi:10.1038/ng.2246

ID	Age (Y)	Gen	Site of Tumor	Stage T	Stage N	Stage M	Stage (AJCC6)	Grade	Lauren Classi-fication	*H. pylori* Status	*TP53 Mut*	*PIK3CA Mut*	*FAT4 Mut*	*ARID1A Mut*	MSI status	Seque-ncing
2000479	62	M	PYLORUS	T3	N2	m0	3B	moderately differentiated	intestinal	no	WT	WT	WT	WT	MSS	Sanger
2000472	81	M	PYLORUS	T3	N1	m0	3A	poorly differentiated	diffuse	no	WT	WT	WT	WT	MSS	Sanger
2000441	53	M	body	T3	N2	m0	3B	poorly differentiated	diffuse	no	V173E	Unknown	WT	c.4214 delT	MSS	Sanger
2000433	56	F	PYLORUS	T3	N1	m0	3A	poorly differentiated	diffuse	yes	F134C	Unknown	WT	WT	MSS	Sanger
2000403	67	M	GE junction	T2b	N1	m0	2	poorly differentiated	diffuse	yes	R213*	WT	WT	WT	MSS	Sanger
2000286	65	M	body	T3	N0	m0	2	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
2000256	61	F	lesser curve	T3	N1	m0	3A	poorly differentiated	mixed	no	K132T	WT	WT	WT	MSS	Sanger
2000242	67	F	lesser curve	T1	N0	m0	1A	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
2000238	74	F	greater curve	T3	N0	m0	2	poorly differentiated	diffuse	no	WT	Unknown	WT	WT	MSS	Sanger
2000201	70	M	lesser curve	T2b	N0	m0	1B	well differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
2000178	71	F	antrum	T2a	N1	m0	2	poorly differentiated	intestinal	yes	WT	WT	WT	WT	MSS	Sanger
2000175	64	M	antrum	T3	N1	m1	4	moderately differentiated	intestinal	yes	WT	WT	WT	WT	MSS	Sanger
2000159	53	F	greater curve	T2b	N0	m0	1B	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
2000114	53	M	lesser curve	T3	N3	m0	4	moderately differentiated	intestinal	yes	WT	WT	WT	WT	MSS	Sanger
2000088	53	M	PYLORUS	T3	N1	m1	4	poorly differentiated	diffuse	no	R248Q	Unknown	WT	WT	MSS	Sanger
2000085	53	M	lesser curve	T2a	N0	m0	2	moderately differentiated	intestinal	yes	WT	Unknown	WT	WT	MSS	Sanger
2000068	65	F	greater curve	T1	N0	m0	1A	moderately differentiated	intestinal	yes	R196*	WT	L1900V	WT	MSS	Sanger
2000040	70	M	body	T4	N3	m0	4	poorly differentiated	diffuse	no	R158H	WT	WT	WT	MSS	Sanger

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 10

Association of clinicopathologic variables with *FAT4* and *ARID1A* mutation status

	FAT4		P value	ARID1A		P value
	Mut(6)	WT(104)		Mut(9)	WT(101)	
Age[1]			0.204			0.703
Mean	69.5	63.6		63.1	64.0	
S.D	6.3	12.9		9.5	13.0	
Gender			0.677			0.478
Male	3	64		7	60	
Female	3	40		2	41	
StageT[2]			0.380			0.687
T1	2	7		0	9	
T2	1	25		3	23	
T3	3	63		5	61	
T4	0	9		1	8	
StageN[3]			1			0.462
N0	2	18		1	19	
N1	2	53		4	51	
N2	1	20		4	17	
N3	1	13		0	14	
StageM			0.477			0.594
M0	5	94		9	90	
M1	1	10		0	11	
Stage(AJCC6)[4]			0.402			1
Stage1	2	13		1	14	
Stage2	1	21		2	20	
Stage3	1	44		5	40	
Stage4	2	26		1	27	
Grade[5]			**6.06E-03**			1
Undifferentiated	0	1		1	0	
Poorly differentiated	0	60		4	56	
Moderately differentiated	6	39		4	41	
Well differentiated	0	3		0	3	
Unknown	0	1		0	1	

Nature Genetics: doi:10.1038/ng.2246

	FAT4		P value	ARID1A		P value
	Mut(6)	WT(104)		Mut(9)	WT(101)	
Lauren			0.148			0.505
Diffuse	0	37		2	35	
Intestinal	6	60		6	60	
Mixed	0	7		1	6	
HP			1			0.296
Yes	3	44		2	45	
No	3	60		7	56	
TP53			0.200			0.149
TP53 WT	2	66		8	60	
TP53 Mutant	4	38		1	41	
PIK3CA			1			**1.31E-03**
PIK3CA WT	3	40		1	42	
PIK3CA Mutant	0	8		4	4	
Unknown	3	56		4	55	
MSI			0.332			**6.80E-04**
MSI	1	6		4	3	
MSS	5	98		5	98	

Remarks:
1. Associations with age were tested by two-sided Wilcoxon Sum Rank Test, the rest were tested using a two-sided Fisher Exact Test
2. Stage T associations were examined by combining T1&T2 versus T3&T4
3. Stage N association were examined by combining N0&N1 versus N2&N3
4. Stage (AJCC6) associations were examined by combining Early Stage(Stage1&Stage2) versus Late Stage(Stage3&Stage4)
5. Grade associations were tested among the two groups: Undifferentiated or poorly differentiated versus Moderately to well differentiated

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 11

Summary of *FAT4* somatic mutations in gastric adenocarcinomas and associated clinicopathologic information.
Data represent somatic mutations found in both the discovery set and prevalence set. We observed 7 non-synonymous and 0 synonymous mutations, whereas the expected values are 5.28 and 1.72, with a trend toward an excess of non-synonymous mutations at p = 0.19. When compared against the background mutation rate, *FAT4* harbors a greater than expected number of non-synonymous mutations (p = 0.0016, FDR = 0.32, **Supplementary Table 7**).

ID	*FAT4*	Age (y)	Gen	Site of Tumor	Stage T	Stage N	Stage M	Stage (AJCC6)	Grade	Lauren	*H. pylori* Status	LOH
2000529	A4485V	76	male	lesser curve	T3	N2	m1	4	moderately differentiated	intestinal	yes	yes
990111	C4479F	74	female	lesser curve	T1	N0	m0	1A	moderately differentiated	intestinal	no	no
2000892	G10S	71	female	lesser curve	T2a	N1	m0	2	moderately differentiated	intestinal	no	no
970010	L1062R	72	male	unknown	T3	N1	m0	3A	moderately differentiated	intestinal	yes	yes
2000068	L1900V	65	female	greater curve	T1	N0	m0	1A	moderately differentiated	intestinal	yes	yes
990515	P4454L Q4872H	59	male	lesser curve	T3	N3	m0	4	moderately differentiated	intestinal	no	yes

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 12

Summary of *ARID1A* somatic mutations in gastric adenocarcinomas and associated clinicopathologic information.
Data represent somatic mutations found in both the discovery set and prevalence set. WT means no somatic mutation was found. We observed 3 non-synonymous and 0 synonymous mutations, whereas the expected values are 1.85 and 1.15, with a trend toward an excess of non-synonymous mutations at p = 0.0555. Note that this analysis does not consider indels. The analogous analysis of *FAT4* and *ARID1A* combined yields a p value of 0.038. When compared against the background mutation rate, *ARID1A* harbored an excess of non-silent mutations ($p= 4.14\times10^{-5}$, FDR = 0.069, **Supplementary Table 7**).

ID	*ARID1A*	*TP53*	*PIK3CA*	Age (y)	Gen	Site of Tumor	Stage T	Stage N	Stage M	Stage (AJCC6)	Grade	Lauren	*H. pylori*	MSI status
990024	c.1650delC; c.3978InsC	WT	R88Q	68	M	lesser curve	T3	N1	m0	3A	moderately differentiated	intestinal	yes	MSI
980028	c.3999_4001 delGCA	WT	Unknown	61	M	unknown	T2a	N0	m0	1B	moderately differentiated	intestinal	no	MSI
990205	c.400delG; c.3216delA	WT	Unknown	72	Fe	lesser curve	T2b	N1	m0	2	poorly differentiated	intestinal	no	MSI
2000441	c.4214delT	V173E	Unknown	53	M	body	T3	N2	m0	3B	poorly differentiated	diffuse	no	MSS
980161	c.5547insG	WT	E545K	66	Fe	lesser curve	T4	N1	m0	4	moderately differentiated	intestinal	no	MSS
990097	c.6748insA	WT	E545K	46	M	cardia	T3	N2	m0	3B	poorly differentiated	intestinal	no	MSS
980097	Q812*	WT	Unknown	65	M	unknown	T2a	N1	m0	2	undifferentiated	mixed	no	MSS
990005	R1276*	WT	WT	60	M	greater curve	T3	N2	m0	3B	poorly differentiated	diffuse	no	MSS
TGH	R1721*	WT	R93Q	77	M	unknown	T3	N2	M0	3B	moderately differentiated	intestinal	yes	MSI

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 13

ARID1A protein expression status in *ARID1A*-mutated and wild-type samples. The table lists 9 *ARID1A*-mutated samples and 12 *ARID1A* wild-type samples. *ARID1A* genetic status was determined either by exome or Sanger sequencing. Representative IHC images are presented in **Supplementary Figure 8.**

Tumor ID	*ARID1A* Genetic Status	Sequncing Platform	ARID1A Protein Expression (IHC)
990097	*ARID1A* mutated	Exome Sequenced	Negative
990005	*ARID1A* mutated	Exome Sequenced	Negative
TGH	*ARID1A* mutated	Exome Sequenced	Weak
990024	*ARID1A* mutated	Sanger	Weak
980161	*ARID1A* mutated	Sanger	Weak
980097	*ARID1A* mutated	Sanger	Negative
2000441	*ARID1A* mutated	Sanger	Positive
980028	*ARID1A* mutated	Sanger	Positive
990205	*ARID1A* mutated	Sanger	ND
2000362	*ARID1A* Wild type	Exome Sequenced	Positive
970010	*ARID1A* Wild type	Exome Sequenced	ND
98748381	*ARID1A* Wild type	Exome Sequenced	Positive
990515	*ARID1A* Wild type	Exome Sequenced	Positive
990172	*ARID1A* Wild type	Exome Sequenced	Positive
990475	*ARID1A* Wild type	Exome Sequenced	Positive
76629543	*ARID1A* Wild type	Exome Sequenced	Positive
990089	*ARID1A* Wild type	Exome Sequenced	Positive
990300	*ARID1A* Wild type	Exome Sequenced	Positive
31231321	*ARID1A* Wild type	Exome Sequenced	Positive
990396	*ARID1A* Wild type	Exome Sequenced	Positive
TWH	*ARID1A* Wild type	Exome Sequenced	Positive

ND=not done due to IHC technical difficulties

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 14

Sequencing analysis of *ARID1A* cDNA clones. GC990024, a tumor harboring two *ARID1A* inactivating mutations (1650delC and 3978insC) was subjected to RT-PCR using *ARID1A*-specific primers flanking both mutations. Amplified RT-PCR products were cloned into PCR4-Topo vectors, followed by Sanger sequencing of individual ARID1A cDNA clones. Sanger sequencing was performed on individual *ARID1A* cDNA clones. The table lists the sequencing results for 27 individual clones. WT-WT: Wild-type at both mutation sites; 1650delC-WT: Presence of 1650delC mutation but no 3978insC mutation; WT-3978insC : Presence of 3978insC mutation but no 1650delC mutation; 1650delC-3978insC : Presence of both mutations. Eight clones contained the 1650delC mutation but were negative for the 3978insC mutation. 10 clones contained the 3978insC mutation but not the 1650 delC mutation. No clones contained both mutations. 9 clones were wild type, likely reflecting contributions from normal cells. This result strongly suggests that the *ARID1A* mutations are likely to affect both *ARID1A* alleles.

Genotype	Number of clones	%
WT-WT	9	33.3
1650delC-WT	8	29.6
WT-3978insC	10	37
1650delC-3978insC	0	0
Total	27	100

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 15

Primer sequences used for prevalence screening of *FAT4* (whole coding region), *PIK3CA* (exon 1, 9 and 20) and *TP53* (exon 4, 5, 6, 7, 8 and 9).

Gene/Exon	Forward Primer	Reverse Primer
FAT4Exon1_1	gtctcagaccaagccgtcag	tgcaggcgatagcggatgtc
FAT4Exon1_2	ccagctcctggttgaggtggag	tctgggaagcaagttcacggtcc
FAT4Exon1_3	gtgtacagagtgaacctgagcga	tacccacatgatatcccagcgcca
FAT4Exon1_4	caggtcaatgctcagagtgg	ctgtcttgaagttcacggtcca
FAT4Exon1_5	tgaagtcacccttctgagtc	ggacacgagttcaccaattctc
FAT4Exon1_6	agacagtacctctggtcaggta	tgaaggtgtctgtgtcccca
FAT4Exon1_7	cggtatgctttgaagaacgtg	ctctgttagcctttgacagaaca
FAT4Exon2	tggcagtcttgattgcgtggat	gaaaggagcacacatgcttaacct
FAT4Exon3	catacaaagcttttaccgccttcg	aacttgccgttggtcacacagcta
FAT4Exon4	ctgaggtccactggatatctga	ttgtgacagctcaacacaata
FAT4Exon5_1	tcctaaagtgtgtttggatgcacg	atgccatagcgaacctgtcca
FAT4Exon5_2	ttgggaccattgatggtgaagtg	ggaagcgtaaacgctgtagaa
FAT4Exon6	tgaagccttccaaaaatgctgtgc	aagtgctggggttacaggcatgag
FAT4Exon7	aggatcccctgccttgactgg	aagcacccagtgtagtcccaaacc
FAT4Exon8	tgtttggtggccttgcacaa	ccatcagacaaaacaacctg
FAT4Exon9_1	aaggtcacatcaagcagcaa	tatctgagaaaacctaggggca
FAT4Exon9_2	ctgtttcggcaatggacaaggac	ggttcaagtcagaaatcagggagg
FAT4Exon9_3	tggacagtaatgacaatgcacctc	gtttggaaacaaaacggggcac
FAT4Exon9_4	ctgtacagtcatctgacagtgacc	aactggtggcaggacctgtg
FAT4Exon9_5	ccctcagctacaggtagtgc	cttcgtagacagctccctcca
FAT4Exon9_6	cttagggactgctgtgcaactg	aagtgcttagaacagtatttgacc
FAT4Exon10	tggtaatggtgtaacggtgtt	ccaatgcaatgtattcaatggg
FAT4Exon11	ggtgtctttggaaaacactgc	agtatagtcctgacatggcct
FAT4Exon12	agtcatgggtagtgttggctg	accagtcatctgaaactgcc
FAT4Exon13	tcggaagtctaacaactctg	catatctctctgtgctggaa
FAT4Exon14	acttccatgcttagaacagtgc	tttgattacatacttgggtgca
FAT4Exon15	ggacattaattcccaaaacgacat	gtgcccagcattagtcggaac
FAT4Exon16	tctatctgctgtggactggtc	agctctcgagaaagtggtggcta
FAT4Exon17_1	gaggtgattgtggagcttctc	ttcaggctgcttcctcacagtc
FAT4Exon17_2	agaacgttgcttttgatgaccc	aatcggctgctttaggcgactc
FAT4Exon17_3	agtcctcaggccatggcatcac	aatgtgacccaacctagtcagc
TP53 Exon4	cctggtcctctgactgctct	gccaggcattgaagtctcat
TP53 Exon5-Eonx6	ctagctcgctagtgggttg	aggagaaagccccctactg
TP53 Exon7	tgcttgccacaggtctcc	aagctccagctccaggtagg
TP53 Exon8-Exon9	ttccttactgcctcttgctt	gaaaacggcattttgagtg
PIK3CA Exon1	cttctgtagttgtgtctcttttaacat	agaaagctatttaagattacgaagg
PIK3CA Exon9	tgaaaatgtatttgctttttctgt	tgtaaattctgctttatttattcc
PIK3CA Exon20	catttgctccaaactgacca	atgctgtttaattgtgtggaag

Nature Genetics: doi:10.1038/ng.2246

Supplementary Table 16

Primer sequences used for *ARID1A, E2F1, Cyclin E1* and *GAPDH* real-time PCR and *ARID1A* cDNA sequencing

Gene and application	Forward	Reverse
ARID1A real-time RCP primer	AAGCCACCAACTCCAGCATCCA	CGCTTCTGGAATGTGGAGTCAC
E2F1 real-time PCR primer	GGACCTGGAAACTGACCATCAG	CAGTGAGGTCTCATAGCGTGAC
Cyclin E1 real-time PCR primer	GGTTAATGGAGGTGTGTGAAGTC	CCATCTGTCACATACGCAAACT
GAPDH real-time PCR primer	GAGTCAACGGATTTGGTC	GACAAGCTTCCCGTTCTCAG
ARID1A cDNA sequencing primer	AGCCTCCACATCAGCAGTC	ATTGCCATAGGAATCATGTCG

Nature Genetics: doi:10.1038/ng.2246

Supplementary Note

Gastric cancer tissues and cell lines

Primary gastric and matched normal tissues were obtained from the Singhealth Tissue Repository or National University Hospital Singapore Tissue Repository, after approvals from Institutional Research Ethics Review Committees, and with signed patient informed consent. Gastric cancer cell lines were obtained from the American Type Culture Collection (Manassas, VA) or from collaborating institutions. Cell lines were grown at 37°C in a humidified atmosphere of 5% CO_2. The cells were subcultured when they reached 90% confluence using a trypsin/EDTA solution (0.05% trypsin, 0.5 mM EDTA).

DNA and RNA extraction

Genomic DNA was extracted from frozen tissues and cell lines using a Qiagen genomic DNA extraction kit. For real-time PCR, total RNA was extracted using a TriPure kit (Roche, IN). For expression microarrays, total RNA was extracted using a Qiagen Rneasy Kit.

Cell line transfections

FAT4 and *ARID1A* siRNAs (pooled or individual) or control siRNAs (Dharmacon, IL) were transfected into gastric cancer cell lines using Lipofectamine (Invitrogen, CA) according to the manufacturer's conditions. Cell lines stably expressing either *ARID1A* short hairpin RNAs (shRNA) or scrambled control shRNAs were established by a Block-iT Letiviral Pol II miR RNAi system (Invitrogen, CA) and selected by puromycin (2.5 ug/ml) (Sigma-Aldrich, MO) to generate stable, pooled populations. *ARID1A*-expressing vectors or empty vector controls were transfected into gastric cancer cells using Effectene Transfection Reagent (Qiagen Inc., CA).

Cell proliferation, invasion and migration assays

Cells were seeded into 96-well plates (2×10^3 cells per well) 24 hours prior to transfection with recombinant DNA constructs or siRNAs. Growth of untransfected or 24, 48, 72 and 96 hours post-transfected cells were determined by a colorimetric 3- (4,5-dimethylthiazol-2yl)-5-(3-carboxymethoxyphenyl)-(4-sulfophenyl)-2H-tetrazoluim assay according to the manufacturer's instruction (MTS; Promega, WI). Optical densities were measured using a PerkinElmer plate reader. Experiments were performed in triplicate, and the results were averaged over three independent experiments. Results are shown as the mean ± SEM of three independent experiments and were analyzed by student's T test using PRISM software (GraphPad Software Inc., CA). Means were considered significantly different from each other at $P<0.05$. For cellular invasion and migration assays, Gastric cancer cells were treated with 50 nM FAT4 siRNA (Si-FAT4) or 50 nM control siRNA (Si-Ctrl) for 24 h prior to seeding into transwells. Seeded cells were observed for an additional 12 h to assess migration rates. For invasion assays, transfected cells were seeded into Matrigel-coated transwells and cultured for 48 h. After this time period, non-migrating/invading cells were wiped from the upper side of the filter and nuclei of migrating/invading cells were stained with DAPI and counted. Results are expressed as mean ± SEM of at least three independent experiments. *$p<0.05$ compared with respective controls (i.e. Si-Ctrl) using Student's t-test.

Nature Genetics: doi:10.1038/ng.2246

Colony formation assays
For soft-agar colony formation assays, cells were treated with *FAT4* or control siRNAs for 48 hours. The treated cells (6000 cells for MKN74 and 5000 cells for YCC16) were plated in 0.3% agar in medium and layered over 0.5% agar covered in six-well plates in triplicate. Cells were allowed to grow for 21 days to form colonies, which were then stained with Thiazolyl Blue Tetrazolium Bromide (2 mg/ml; Sigma). Images were taken using the Kodak GL 200 System (EpiWhite illumination).

Adhesion assays
Adhesion assays were performed using the InnoCyte^{TM} ECM adhesion assay kit (Calbiochem) and performed according to the manufacturer's instructions. Briefly, 10,000 cells (HGC27, YCC16 or MKN74 cells) were transfected *FAT4* siRNAs (Si-FAT4) or control siRNA in serum free medium, added to fibronectin-coated 96 well plates, and allowed to adhere at 37 C for 1.5 h. After incubation, the plates were washed twice with cold PBS, and adherent cells were stained with Calcein-AM green fluorescent dye and incubated for one additional hour at 37 C in a cell culture incubator. Flouresence was measured using a fluorescence plate reader at excitation wavelength 485 nm and emission wavelength 520 nm. Results are expressed as fold change versus Si-Ctrl.

Immunohistochemistry (IHC)
ARID1A IHC was performed using ARID1A antibodies (Sigma, HPA005456, 1:50). Tumor sections were treated with citrate buffer (pH 6.0) and pressure cooking (3 min) for antigen retrieval. Primary antibody incubations were 1 hour at room temperature. Manual DAKO Envision kits were used for visualization. Normal epithelial and lymphoid cell nuclei were used as positive internal controls. Loss or decreased nuclear staining in tumor cells relative to normal cells was interpreted as ARID1A protein loss.

Real-time RT-PCR and cDNA Sequencing
Total purified RNA was reverse-transcribed to cDNA using iScript™ cDNA Synthesis Kit (Bio-Rad, CA) following the manufacturer's instructions. Real-time PCR was performed with SsoFast EvaGreen Supermix according to the manufacturer's recommendations using a CFX96™ Real-Time PCR Detection System (Bio-Rad, CA) and the primers for *ARID1A, E2F1, CyclinE1, and GAPDH* (**Supplementary Tables 16)**. All RT-qPCR experiments were run in triplicate, and a mean value was used for the determination of mRNA levels. Negative controls containing water instead of sample cDNAs were used in each experiment. Relative quantification of the mRNA levels of *ARID1A*, *E2F1*, *cyclinE1* was performed using the comparative Cq method with *GAPDH* as the reference gene and with the formula $2^{-\Delta\Delta Cq}$.

To sequence *ARID1A*-mutated cDNAs of tumor 990024, total RNA was extracted from *ARID1A*-mutated tumor (990024) and subjected to *ARID1A* RT-PCR using primers designed to flank mutation sites (**Supplementary Tables 16**). Amplified RT-PCR products were cloned into PCR4-Topo vectors (Invitrogen CA). Individual cDNA clones were isolated and processed for Sanger sequencing.

Nature Genetics: doi:10.1038/ng.2246

Western blot analysis

Cellular proteins were extracted with lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM Na_2EDTA, 1 mM EGTA, 1% Triton, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na_3VO_4 and 1 μg/mL leupeptin) containing freshly added protease inhibitor cocktail (Sigma-Aldrich). Cell lysates were sonicated, centrifuged for 30 min at 20,000 *g* and supernatants were collected and stored at -70°C. Protein concentrations were quantified using the DC assay (Bio-Rad Laboratories, Mississauga, ON) with bovine serum albumin as a standard. Western blotting was performed on supernatants using the following antibodies, FAT4 (P-17) and ARID1A (sc-32761) (Santa Cruz, CA); β-actin (A-1978) (Sigma-Aldrich, MO); tubulin antibody (Cell Signaling Technology, MA). All antibodies were diluted in TBS with 5% (w/v) skim milk and 0.1% (v/v) Tween-20. Equal amounts of protein were subjected to reducing SDS polyacrylamide gel electophoresis prior to being electrotransferred to nitrocellulose membrane (Bio-Rad Laboratories). Membranes were blocked for 1 h at room temperature in TBS with 5% skim milk and 0.1% Tween-20, and incubated with primary antibodies overnight at 4°C. After four washes with TBS containing 0.1% Tween-20, the membranes were incubated with secondary antibodies conjugated with horseradish peroxidase, washed four times and visualized with SuperSignal West Pico chemiluminescent substrate (Pierce, IL) and Kodak BioMax XAR film (VWR International).

Microsatellite instability (MSI) testing

Gastric cancer tissues were evaluated for MSI status by microsatellite sequencing. Tumor genomic DNAs were processed using the Promega MSI Analysis System Version 1.2 (Promega Corp, USA) measuring five mononucleotide markers (NR-21, BAT-26, BAT-25, NR-24, MONO-27) and two pentanucleotide markers (Penta C and Penta D) for sample identification. MSI analysis was performed according to the manufacturer's protocol and PCR products were analyzed using an ABI 3130XL Genetic Analyzer (Applied Biosystems, USA). Tumors exhibiting shifts in two or more microsatellites were designated as MSI-high (MSI-H), and tumors with less than two or no shifts were microsatellite stable (MSS). This method is technically better than the original NCI-reported panel for MSI assessment [1, 2].

Nature Genetics: doi:10.1038/ng.2246

Supplementary references

1. Murphy KM, et al. Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers. J Mol Diagn. **8**, 305-311 (2006)
2. Salto-Tellez M, et al. Microsatellite instability in colorectal cancer: considerations for molecular diagnosis and high-throughput screening of archival tissues. Clin Chem. **50,** 1082-1086 (2004)

Nature Genetics: doi:10.1038/ng.2246