A Comprehensive Survey of Genomic Alterations in Gastric Cancer Reveals Systematic Patterns of Molecular Exclusivity and Co-Occurrence among Distinct Therapeutic Targets

(Supplementary Information)

Niantao Deng*[,1,2], Liang Kee Goh*[,1,3,4], Hannah Wang[1], Kakoli Das[1], Jiong Tao[1,5], Iain Beehuat Tan[1,2,4], Shenli Zhang[1], Minghui Lee[6], Jeanie Wu[6], Kiat Hon Lim[7], Zhengdeng Lei[8], Glenn Goh[1], Qing-Yan Lim[9], Angie Lay-Keng Tan[1], Dianne Yu Sin Poh[1], Sudep Riahi[10], Sandra Bell[10], Michael M. Shi[11], Ronald Linnartz[11], Feng Zhu[12], Khay Guan Yeoh[12], Han Chong Toh[4], Wei Peng Yong[13], Hyun Cheol Cheong[14], Sun Young Rha[14], Alex Boussioutas[15], Heike Grabsch[16], Steve Rozen[8], Patrick Tan[1,6,17,18,**]

[1]Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, Singapore
[2]NUS Graduate School for Integrative Sciences and Engineering/[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore
[4]Division of Medical Oncology, National Cancer Centre, Singapore
[5]Department of Physiology, National University of Singapore, Singapore
[6]Cellular and Molecular Research, National Cancer Centre, Singapore
[7]Dept of Pathology, Singapore General Hospital, Singapore
[8]Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, Singapore
[9]School of Biological Sciences, Nanyang Technological University, Singapore
[10]Section of Ophthalmology and Neuroscience, Leeds Institute for Molecular Medicine, Leeds, England
[11]Novartis Oncology, East Hanover, New Jersey, USA
[12]Department of Medicine/[13]National Cancer Institute Singapore, National University Health System, Singapore
[14]Department of Internal Medicine, Yonsei Cancer Centre, South Korea
[15]Cancer Genomics and Biochemistry Laboratory, Peter MacCallum Cancer Centre, Australia
[16]Department of Pathology and Tumour Biology, Leeds Institute for Molecular Medicine, Leeds, England
[17]Cancer Science Institute of Singapore, National University of Singapore, Singapore
[18]Genome Institute of Singapore, Singapore
*These authors contributed equally to this study
** **Correspondence to**
Patrick Tan
Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, 8 College Road, Singapore, 169857;
Tel: (65) 6516 1783; Fax: (65) 6221 2402
gmstanp@duke-nus.edu.sg

Keyword: copy number alterations, receptor tyrosine kinases, mutual exclusivity, targeted therapies

Supplementary Items

Materials and methods

Supplementary Text

Supplementary Tables

Supplementary Figures

Figure S11: *FGFR2* Overexpression in GCs Relative to Normal Gastric Samples

Figure S12: Inhibition of Soft Agar Colony Growth by Dovitinib (SNU-16)

Materials and Methods

Clinical Samples and Cell Lines

Primary gastric samples were obtained from the Singapore Health Services (SingHealth) and the National University Hospital System (NUHS) tissue repositories, with signed informed patient consent and approvals from the respective institutional Research Ethics Review Committees. Clinical information was collected with Institutional Review Board approval. There was no pre-specified sample size calculation since this is a hypothesis generating discovery study. Clinical characteristics of patients analyzed in this study are presented in Supplementary Table S1. GC cell lines were obtained from commercial sources (American Type Culture Collection, Japan Health Science Research Resource Bank) or from collaborators (Yonsei Cancer Centre, S. Korea).

DNA and RNA Extraction

Genomic DNA was extracted from flash-frozen tissues and cells using a Qiagen genomic DNA extraction kit. Total RNAs was extracted using Trizol (Invitrogen, CA), digested with RNase free DNase (RQ1 DNase, Promega), and subsequently purified using an RNeasy Mini kit (Qiagen,CA).

Copy Number Profiling and GISTIC Analysis

Genomic DNAs from gastric tumors and matched non-malignant gastric tissues (normal) were hybridized on Affymetrix SNP6 genotyping arrays and processed as follows:
Step 1) Normalization: Raw SNP6 CEL files were processed using Affymetrix Genotyping Console 4.0. A reference file was first created from the SNP6 CEL files of normal gastric samples (98 samples). The 193 tumor SNP6 CEL files were then normalized against this normal reference file.
Step 2) Segmentation: Copy number segmentation data was produced using the Circular Binary Segmentation (CBS) algorithm using the R package *DNAcopy* [1] for both tumor and normal gastric samples. The p value cutoff for detecting a change-point was 0.01, with a permutation number of 10000.

3) GISTIC Analysis: The GISTIC algorithm [2] was used to identify genomic regions with recurrent copy number alterations. GISITC was applied to the CBS-segmented files of tumors, and filtered through a CNV (copy number variation) file constructed from the segmented data of normal samples to identify somatic tumor-specific CNAs. GISTIC reports regions of interest with an associated q-value, which is obtained by multiple hypotheses correction. Genomic regions with q-value<0.25 for broad regions and q-value<0.001 for focal regions were considered significant. Proportions of CNA for individual normal and tumor sample was defined as: size of CBS regions with CNA per sample divided by the sum of all autosome lengths. Chromosomal instability values for GCs were estimated by the number of cytobands exhibiting CNA for each sample, calculated by averaging the CBS segmented value for each cytoband.
The SNP6 copy number data has been deposited into the National Centre for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) website, series accession number GSE31168.The Reviewer link is http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=tbqtnokgaooucjo&acc=GSE31168

DRP: Identification of Mutually Exclusive and Co-Altered CNAs

To identify significant relationships between regions of frequent CNA, we implemented a dimension reduction permutation (DRP) statistical algorithm adapted from a previous study analyzing patterns of somatic DNA mutations in tumor [3]. To determine the significance of any specific mutually exclusive (ME) or co-alteration (CA) interaction, we compared the numbers of samples exhibiting a particular ME or CA interaction against a null distribution of interactions obtained by randomly permuting the genomic alterations across samples and genes (100,000 permutations), while taking into consideration the prevalence of genomic alterations. Essentially, for each permutation, we constrained the number of samples with genomic alterations and the number of genes exhibiting alterations within each sample to be similar to the original data. Empirical p-values of <0.05 were considered significant. An in-depth description of the DRP methodology is presented in Text S1, and the DRP software can be downloaded from http://research.duke-nus.edu.sg/papers/DRP.zip.

FISH and Immunohistochemical Analysis

KRAS and *FGFR2* FISH was performed using BAC clones obtained from the BACPAC resources center (CHORI, Oakland, CA USA). BAC DNA was labeled using a Bioprime DNA labeling kit (Invitrogen, Carlsbad, CA, USA). FISH was performed on metaphase spreads (cell lines) or on FFPE sections after deparaffinization (clinical specimens). Target DNA probes were labeled using spectrum green and control probes in spectrum orange (centromeric CEP probes for chromosomes 10 and 12) (Abbott Molecular Inc, Des Plaines, IL, USA). Hybridized slides were counterstained with DAPI and analyzed using a Olympus BX50 fluorescence microscope. Nuclei were scored for amplification by comparing signals from internal controls (CEP probes) against target gene signals (*KRAS*, and *FGFR2*). For *ERBB2* immunohistochemistry, we analyzed 146 of the 193 tumors, representing all cases for which we were able to obtain full sections. The remaining 47 cases were not analyzed for a variety of reasons, including failure to retrieve the samples due to historical storage arrangements (archival samples are stored off-site at our center) and insufficient material due to exhaustion of the FFPE blocks (small tumors). Sections of archival formalin-fixed, paraffin-embedded tissue (3 μm) were placed on slides coated with poly-L-lysine. After deparaffinisation and blocking of endogenous peroxidase, *ERBB2* immunostaining was performed using rabbit anti-human c-erbB-2 oncoprotein as primary antibody (Dako Corp, Carpinteria, CA, USA) at 1/100 dilution. Binding of the primary antibody was revealed by means of the Dako Quick-Staining, Labelled Streptavidin–Biotin System (Dako), followed by the addition of diaminobenzidine as a chromogen. *ERBB2* immunoreactivity was evaluated by an experienced pathologist (LKH) according to the scoring system of [4]. CSMD1 immunohistochemistry was performed on full sections as described in [5]. Tumors were scored by two independent observers (HG, SB) and classified as *CSMD1* present (> 25% positive positive tumour cells) or *CSMD1* Absent/Reduced (<= 25% positive tumor cells).

DNA Sequencing, Mutation Genotyping and Quantitative PCR

DNA products corresponding to the coding regions of target genes were amplified by PCR and were subjected to cycle sequencing using the BigDye Terminator v3.1 Cycle

Sequencing Kit (Applied Biosystems, Foster City, CA, USA) on a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA). *KRAS* mutation genotyping was performed by both Sanger sequencing (139 GCs) and mass-spectrometry based genotyping (Sequenom MassARRAY) (94 GCs). Reference sequences were obtained from the Ensembl Genome Browser database. Quantitative real-time PCR was performed on an ABI 7900 HT instrument using *FGFR2* intron 2 primers. Reaction mixes consisted of 5ul SYBR green PCR master mix (ABI), 1ul *FGFR2/LINE1* primers, 20ng (0.5ul) of genomic DNA template in a final reaction volume of 10ul. All experiments were performed in triplicate. *FGFR2* cycle thresholds were normalized to the *LINE1* repeat element from the same samples, as an endogenous control. Normal human genomic DNA was chosen as the calibrator and for each analysis a negative control was also prepared using all reagents except DNA template.

Gene Expression Analysis

Of the 193 tumors profiled on Affymetrix SNP6 arrays (Affymetrix, Santa Clara, CA, USA), 156 tumors had corresponding gene expression data available along with 100 normal gastric samples on Affymetrix U133P2 arrays (this cohort is analyzed in Figure 4C). Additional details of the gene expression data set are presented in [6] and are publicly available at GEO under accession number GSE15460. To analyze *FGFR2* mRNA survival associations in Figure 4D, we analyzed a combined GC gene expression data set of 398 tumors. The 156 patients analyzed in Figure 4C form a subset of the 398 patients. To establish this combined data set, we combined gene expression data from GSE15460 and three other GC cohorts from Singapore (U133AB), Australia (AU) and the University of Leeds, UK (UK). Clinical information for these gene expression data sets is provided in Table S6. Briefly, individual arrays were normalized using the MAS5 algorithm, and batch effects removed using the COMBAT algorithm [7].

Clinico-Pathologic Correlation Analysis

Survival curves were estimated using the Kaplan-Meier method, with the duration of survival measured from the date of surgery to date of death or last follow-up visit. Overall survival was used as the outcome metric. Patients who were still alive or lost to

follow-up at time of analysis were censored at their last date of follow up. Univariate and multivariate survival analysis was performed using the Cox proportional hazards regression model. Besides genetic factors (e.g. *FGFR2*, *KRAS*), other clinical factors considered in the multivariate model included grade and stage which were also significant in univariate analysis. Associations with other clinical variables were performed using the Fisher Exact Test, at a significance threshold of $p<0.05$.

Reverse Transcription-PCR (RT-PCR) and Western Blotting Analysis

For mRNA analysis, equal quantities of RNA were reverse transcribed using SuperScript III Reverse Transcriptase enzyme and oligo$(dT)_{20}$ primers (Invitrogen). RT-PCR was performed with forward primers to *FGFR2* exon 8 (5'- GTGCTTGGCGGGTAATTCTA-3') and reverse primers to exon 9 (5'-TACGTTTGGTCAGCTTGTGC -3'). *GAPDH* was used as a loading control (forward primer (5'- GTGCTTGGCGGGTAATTCTA-3'); reverse primer (5'-TCCACCACCCTGTTGCTGTA-3'). For protein analysis, cells were harvested in lysis buffer (0.3M NaCl, 0.05M Tris-HCl pH8, 0.5% NP40, 0.1% SDS, Protease Inhibitor (Roche, Mannhein, Germany) and Halt Phosphatase Inhibitor Cocktail (Pierce, Rockford, IL, USA)). *FGFR2* immunoprecipitation was performed by incubating lysates with MAB6841 (R&D Systems, Minneapolis, MN, USA) for 4 hrs at room temperature; followed by incubation with protein A/G agarose beads (Pierce, Rockford, IL, USA) overnight at 4°C. After washing, 4X SDS loading buffer was added and the mixture was boiled at 95°C for 5 minutes. Antibodies against p-ERK, ERK, p-AKT, AKT and Caspase-3(8G10) were obtained from Cell Signaling Technology (Cell Signaling Technology, Danvers,. MA, USA). Other antibodies include 4G10 phosphotyrosine antibody (Upstate Biotechnology, Lake Placid, NY, USA β-actin (Millipore, Billerica, MA, USA) or α-tubulin (Cell Signaling Technologies, Danvers, MA, USA) were used as loading controls. Blots were incubated with DyLight Fluorescence secondary antibodies (Thermo Scientific) and imaged using LI-COR Odyssey. Experiments were repeated a minimum of three independent times.

Cell Proliferation Assays and Drug Treatments

Cell proliferation assays were performed using the CellTiter 96® AQ_{ueous} One Solution Assay kit (Promega) and the plates were measured using a PerkinElmer plate reader. Each assay was performed in triplicate, and the results were averaged over three independent experiments. Dovitinib was provided by Drs. D. Graus-Porta and C. Garcia-Echeverria (Novartis Institutes for Biomedical Research, Basel, Switzerland). GC cells were seeded in 96-well plates 24 hours prior to Dovitinib treatment. On the day of drug treatment, CellTiter reagent was added to one plate of cells to provide a measurement of the cell population at the time of drug addiction (T_z). Five serial 10-fold dilution mixtures of Dovitinib, beginning with a maximum concentration of 10^{-5} M, were added to the respective wells. The final DMSO concentration in the wells did not exceed 0.1% (v/v). GI50 values for Dovitinib, representing the concentration at which 50% cell growth inhibition is achieved for 48 hours of treatment, were computed using the GI50 calculation formula at http://dtp.nci.nih.gov/branches/btb/ivclsp.html.

Cell Death and Colony Formation Assays

Caspase 3/7 assays were performed using the Caspase-Glo® 3/7 Assay kit (Promega, WI, USA) and the plates were measured using a Tecan plate reader. Three independent experiments were performed and each assay was performed in triplicate. GC cells were seeded in 96-well black plates and treated with Dovitinib using the same method as the cell proliferation assays. For colony formation assays, base layers of 0.5% Gum Agar in 1x McCoy's 5A and 10% FBS were poured into 6-well plates and allowed to harden at 4°C. After siRNA transfection, overexpression, or drug treatment, 50 000 cells/well were seeded in complete media plus agar mixture at 42°C and seeded on top of the solidified base layer. Plates were incubated at 37°C in for 3-4 weeks, during which plates were fed drop-wise with complete media. After 3-4 weeks, plates were photographed using the Kodak GL 200 System (EpiWhite illumination). Each assay was performed in triplicate, and the results were averaged over three independent experiments.

Xenograft assays

Efficacy of dovitinib was evaluated and compared to the positive control drug 5-FU in a primary human gastric cancer xenograft model (n= 10 in each group). This tumor model

was derived from a primary gastric cancer from Chinese ethnicity and is confirmed with *FGFR2* gene amplification (26 copies of *FGFR2* by SNP6.0 array). Tumor fragments from stock mice inoculated with selected primary human gastric cancer tissues were harvested and used for inoculation into Balb/c nude mice. Each mouse was inoculated subcutaneously at the right flank with primary human gastric tumor fragment (2-3 mm in diameter) for tumor development. Treatments were started at day 24 after tumor inoculation when the average tumor size reached about 150 mm^3.

Text S1: Dimension Reduction Permutation (DRP): Identification of Mutually Exclusive and Co-Altered CNAs

Non-random associations between distinct genomic alterations (co-associated or mutually exclusive) may suggest synergistic or antagonistic biological event in carcinogenesis. To compute the significance of these associations, a dimension reduction permutation (DRP) algorithm was developed. It was adapted from a previous study analyzing patterns of somatic DNA mutations in tumor [3]. To determine the significance of any pair of mutually exclusive or co-altered CNAs, we used permutation testing, taking into consideration the prevalence of genomic alterations. Since we are testing for associations regardless of the level of alterations (i.e. focal or broad), we assigned each gene to either an amplification or deletion status, based on the mean aggregation of log ratio signals of all probes within each gene. To maintain a similar prevalence of genomic alterations observed in the original data, the number of samples with genomic alterations and the number of genes exhibiting the alterations were maintained in the permutations. Suppose the matrix is represented as genes (row) x samples (column). DRP permutes the genomic alterations by row or by column progressively, depending on which number of rows or columns is smaller. Permutations can start from the top row or the left column of the matrix while maintaining the marginal counts for genomic alterations in genes and samples to be similar to the original data. In effect, for each permutation, the algorithm traverses iteratively from top left to bottom right of the matrix, each time reducing the dimension by multiple numbers of rows and columns – hence the name Dimension Reduction Permutation. For each permutation, the number of samples with co-altered (N_{CA}) and mutually exclusive CNA (N_{ME}) was then recorded for each pair of genes and then compared with original data on co-altered (O_{CA}) and mutually exclusive genes (O_{ME}) respectively. Frequencies were summarized for co-altered ($N_{CA}>=O_{CA}$) and mutually exclusive associations ($N_{ME}>=O_{ME}$). Empirical p-values were then computed against these frequencies under the null hypothesis.

Supplementary Tables

Table S1. Clinical Characteristics of the GC Patient Cohort.

This table provides clinical data for 193 patients analyzed by Affymetrix SNP6 arrays. Stage categories were based on the AJCC 6th edition classification**.** 3 patients received neoadjuvant therapy, and of 131 patients where subsequent treatment information was available, 28 patients received 5-FU chemoradiation as adjuvant therapy.

	GC Samples (193)
Age	
Range	23-92
Mean,S.D	64.2, 12.6
Gender	
Male	123
Female	70
Lauren Classification	
Intestinal	99
Diffuse	73
Mixed/Others	21
Anatomical Location*	
Gastro-oesophageal junction	9
Cardia	13
Body	24
Greater Curve	17
Lesser Curve	37
Pylorus	12
Antrum	22
Incisura	2
Grade	
Undifferentiated	2
Poorly differentiated	117
Moderately differentiated	67
Well differentiated	5
Unknown	2
Stage	
1	32
2	26
3	71
4	64

*This is only for 136 patients where location information was reliably recorded.

Table S2: Concordance Table between *ERBB2* SNP6 and *ERBB2* IHC

9 of 132 (6.8%) *ERBB2* copy number neutral tumors exhibit ERBB2 protein expression (IHC 1-3+), while 8 of 13 (61.5%) tumors with *ERBB2* copy number gain also exhibit ERBB2 protein expression ($p<0.01$, Fisher's exact test).

	ERBB2 Immunohistochemistry				
ERBB2 SNP 6 Copy Number	Positive staining	0	1+	2+	3+
Loss (logRatio<-0.2)	0 / 1 (0 %)	1	0	0	0
Neutral (-0.2 < logRatio < 0.2)	9 / 132 (6.8%)	123	2	3	4
Gain (logRatio > 0.2)	8 / 13 (61.5%)	5	2	3	3

Table S3: DRP Analysis of Mutually Exclusive and Co-Amplification Interactions

This table lists all significant mutually exclusive (ME) and co-occurring (CO) interactions for a pair of genes ('Gene1' and 'Gene2'). The columns are : '#Gene1' and '#Gene2' are the observed frequency of amplification for each pair of genes. '#Both' indicates the observed number of cases of coamplification for this pair of genes, and '#OnlyOne' indicates the observed number of cases for amplification in only one of this pair of genes. '#BothExp' and '#OnlyOneExp' are the expected results from the DRP permutation for coamplification cases and non-coamplification cases. 'PvalueME' and 'PvalueCO' are the empirical pvalues for ME and CO interactions. 'QvalueME' and 'QvalueCO' are converted Storey's qvalue. Gene pairs related to RTK/RAS signaling are highlighted. Significant ME interactions are at the top of the list, while significant CO interactions are at the bottom.

Gene1	Gene2	#Gene1	#Gene2	#Both	#OnlyOne	#BothExp	#OnlyOneExp	PvalueME	QvalueME	PvalueCO	QvalueCO
FGFR2	*KLF5*	22	22	0	44	5.765	32.470	**0.001**	0.118	0.999	0.999
GATA4	*KLF5*	23	22	1	43	5.840	33.320	**0.010**	0.464	0.990	0.999
KRAS	*ERBB2*	21	17	1	36	5.191	27.619	**0.018**	0.464	0.982	0.999
FGFR2	*MET*	22	14	1	34	4.681	26.637	**0.028**	0.464	0.972	0.999
CCNE1	*MET*	23	14	1	35	4.692	27.615	**0.028**	0.464	0.972	0.999
ERBB2	*MET*	17	14	1	29	4.558	21.884	**0.031**	0.464	0.969	0.999
CCNE1	*GATA4*	23	23	2	42	5.904	34.193	**0.042**	0.470	0.958	0.999
CCNE1	*KRAS*	23	21	2	40	5.744	32.513	**0.048**	0.470	0.952	0.999
FGFR2	*KRAS*	22	21	2	39	5.696	31.607	**0.049**	0.470	0.951	0.999
KRAS	*EGFR*	21	21	2	38	5.634	30.733	0.052	0.470	0.948	0.999
GATA4	*ERBB2*	23	17	2	36	5.248	29.503	0.070	0.578	0.930	0.999
GATA6	*CDH12*	25	14	2	35	4.687	29.625	0.105	0.602	0.895	0.999
CCND1	*MET*	24	14	2	34	4.692	28.616	0.106	0.602	0.895	0.999
CDH12	*CCND1*	14	24	2	34	4.694	28.611	0.106	0.602	0.894	0.999
GATA4	*MET*	23	14	2	33	4.685	27.631	0.106	0.602	0.894	0.999

CCNE1	*CDK6*	23	26	3	43	6.034	36.933	0.112	0.602	0.888	0.999
GATA6	*KLF5*	25	22	3	41	5.917	35.167	0.119	0.602	0.881	0.999
FGFR2	*CCNE1*	22	23	3	39	5.826	33.349	0.127	0.602	0.873	0.999
KRAS	*CCND1*	21	24	3	39	5.791	33.418	0.131	0.602	0.869	0.999
FGFR2	*EGFR*	22	21	3	37	5.700	31.599	0.137	0.602	0.863	0.999
CDH12	*MET*	14	14	2	24	4.316	19.368	0.139	0.602	0.861	0.999
CDK6	*ERBB2*	26	17	3	37	5.282	32.435	0.181	0.741	0.820	0.999
FGFR2	*ERBB2*	22	17	3	33	5.232	28.536	0.187	0.741	0.813	0.999
CCNE1	*CCND1*	23	24	4	39	5.963	35.074	0.249	0.763	0.751	0.999
GATA4	*CCND1*	23	24	4	39	5.962	35.075	0.250	0.763	0.750	0.999
FGFR2	*CDH12*	22	14	3	30	4.690	26.620	0.258	0.763	0.742	0.999
CDH12	*GATA4*	14	23	3	31	4.684	27.633	0.260	0.763	0.740	0.999
GATA6	*MET*	25	14	3	33	4.684	29.632	0.260	0.763	0.740	0.999
KRAS	*CDH12*	21	14	3	29	4.670	25.660	0.262	0.763	0.738	0.999
CDH12	*EGFR*	14	21	3	29	4.669	25.663	0.263	0.763	0.737	0.999
KRAS	*GATA6*	21	25	4	38	5.817	34.366	0.268	0.763	0.732	0.999
EGFR	*ERBB2*	21	17	4	30	5.204	27.592	0.367	0.960	0.633	0.976
FGFR2	*CDK6*	22	26	5	38	5.950	36.101	0.431	0.960	0.569	0.909
KLF5	*CCND1*	22	24	5	36	5.888	34.224	0.440	0.960	0.561	0.909
EGFR	*CCND1*	21	24	5	35	5.786	33.427	0.459	0.960	0.541	0.909
GATA4	*EGFR*	23	21	5	34	5.750	32.500	0.466	0.960	0.534	0.909
KRAS	*GATA4*	21	23	5	34	5.735	32.529	0.468	0.960	0.532	0.909
CDH12	*CDK6*	14	26	4	32	4.689	30.622	0.472	0.960	0.528	0.909
KRAS	*MET*	21	14	4	27	4.663	25.674	0.474	0.960	0.526	0.909
EGFR	*KLF5*	21	22	5	33	5.698	31.604	0.475	0.960	0.525	0.909
CDH12	*ERBB2*	14	17	4	23	4.559	21.882	0.499	0.982	0.501	0.909
GATA6	*ERBB2*	25	17	5	32	5.273	31.455	0.561	0.982	0.439	0.897
ERBB2	*CCND1*	17	24	5	31	5.259	30.482	0.565	0.982	0.435	0.897
ERBB2	*KLF5*	17	22	5	29	5.231	28.539	0.568	0.982	0.432	0.897
GATA6	*CDK6*	25	26	6	39	6.201	38.598	0.568	0.982	0.432	0.897

CDK6	*GATA4*	26	23	6	37	6.052	36.896	0.596	0.982	0.404	0.897
FGFR2	*MYC*	22	46	6	56	6.023	55.955	0.604	0.982	0.397	0.897
CCNE1	*GATA6*	23	25	6	36	6.010	35.979	0.604	0.982	0.396	0.897
GATA6	*GATA4*	25	23	6	36	6.007	35.985	0.606	0.982	0.394	0.897
CDK6	*KLF5*	26	22	6	36	5.951	36.098	0.615	0.982	0.385	0.897
KRAS	*CDK6*	21	26	6	35	5.839	35.322	0.637	0.999	0.363	0.897
KRAS	*KLF5*	21	22	6	31	5.694	31.612	0.664	1.000	0.336	0.897
KLF5	*MET*	22	14	5	26	4.687	26.626	0.685	1.000	0.315	0.895
MYC	*ERBB2*	46	17	6	51	5.289	52.423	0.740	1.000	0.260	0.789
FGFR2	*GATA6*	22	25	7	33	5.922	35.157	0.780	1.000	0.220	0.714
GATA6	*EGFR*	25	21	7	32	5.815	34.370	0.797	1.000	0.203	0.685
CCNE1	*EGFR*	23	21	7	30	5.734	32.533	0.810	1.000	0.190	0.664
CDH12	*KLF5*	14	22	6	24	4.677	26.646	0.846	1.000	0.154	0.560
EGFR	*MET*	21	14	6	23	4.678	25.644	0.847	1.000	0.153	0.560
CDK6	*CCND1*	26	24	8	34	6.125	37.751	0.866	1.000	0.134	0.529
GATA6	*CCND1*	25	24	8	33	6.065	36.870	0.873	1.000	0.127	0.524
MYC	*KRAS*	46	21	8	51	5.894	55.211	0.892	1.000	0.108	0.466
FGFR2	*CCND1*	22	24	8	30	5.881	34.237	0.895	1.000	0.105	0.466
FGFR2	*GATA4*	22	23	8	29	5.847	33.307	0.898	1.000	0.102	0.466
CCNE1	*KLF5*	23	22	8	29	5.838	33.324	0.898	1.000	0.102	0.466
MYC	*CCND1*	46	24	9	52	6.222	57.556	0.932	1.000	0.068	0.365
MYC	*MET*	46	14	7	46	4.695	50.610	0.938	1.000	0.062	0.351
CCNE1	*ERBB2*	23	17	8	24	5.247	29.505	0.950	1.000	0.051	0.306
CCNE1	*CDH12*	23	14	8	21	4.676	27.649	0.982	1.000	**0.019**	0.140
MYC	*GATA6*	46	25	11	49	6.306	58.388	0.988	1.000	**0.012**	0.111
MYC	*GATA4*	46	23	11	47	6.124	56.752	0.991	1.000	**0.009**	0.090
MYC	*CCNE1*	46	23	11	47	6.129	56.741	0.991	1.000	**0.009**	0.090
CDK6	*EGFR*	26	21	11	25	5.849	35.303	0.995	1.000	**0.005**	0.062
MYC	*EGFR*	46	21	12	43	5.896	55.209	0.999	1.000	**0.001**	0.023
CDK6	*MET*	26	14	10	20	4.694	30.613	0.999	1.000	**0.001**	0.015

MYC	*CDH12*	46	14	10	40	4.698	50.603	0.999	1.000	**0.001**	0.015
MYC	*KLF5*	46	22	13	42	6.021	55.959	0.999	1.000	**5.00E-04**	0.015
MYC	*CDK6*	46	26	15	42	6.375	59.249	1.000	1.000	**1.00E-04**	0.005

Table S4a: Multivariate analysis comparing RTK amplification status with tumor stage, grade, adjuvant treatment and genome instability
(Outcome: overall survival, relative to patients lacking RTK amplification).

Model 1 (Predictors: RTK Amp, Stage ,Grade and Adjuvant Treatment)	**Hazard Ratio (95% CI)**	**P-value**
RTK Amp vs RTK Absent	1.966 (1.180, 3.279)	**0.01**
Stage 2 vs Stage 1	2.329 (0.867, 6.254)	0.09
Stage 3 vs Stage 1	6.522 (2.712, 15.686)	**2.8E-05**
Stage 4 vs Stage 1	8.576 (3.280, 22.425)	**1.2E-05**
Poorly Differentiated vs Moderately to well Differentiated	1.058 (0.642, 1.741)	0.8
Surgery alone vs Surgery + 5 FU	0.951 (0.556, 1.628)	0.3
Model 2 (Predictors: RTK Amp and Genomic Instability*)	**Hazard Ratio (95% CI)**	**P-value**
RTK Amp vs RTK Absent	1.495 (0.970, 2.304)	0.07
High CNA vs Low CNA	1.228 (0.823, 1.833)	0.3

Significant p-values are shown in bold type. *Genomic Instability was inferred based on the number of copy number altered cytobands for each tumor sample (methods).

Table S4b: Univariate analysis analyzing the prognostic impact of individual RTK amplifications
(Outcome: overall survival, relative to patients lacking RTK amplifications)

Model 3 (Predictors: RTK Amp vs RTK Absent)	**Hazard Ratio (95% CI)**	**P-value**
EGFR Amp vs RTK Absent	1.179 (0.589, 2.360)	0.6
ERBB2 Amp vs RTK Absent	2.824 (1.558, 5.119)	**0.0006**
FGFR2 Amp vs RTK Absent	1.098 (0.549, 2.196)	0.8
MET Amp vs RTK Absent	2.744 (1.190, 6.327)	**0.002**

Significant p-values are shown in bold type.

Table S4c: Multivariate analysis comparing individual RTK amplification status with tumor stage and grade
(Outcome: overall survival, relative to patients lacking RTK amplifications)

Model 4 (Predictors: RTK Amp, Stage and Grade)	**Hazard Ratio (95% CI)**	**P-value**
EGFR Amp vs RTK Absent	1.160 (0.570, 2.360)	0.7
ERBB2 Amp vs RTK Absent	3.691 (1.985, 6.863)	**3.7E-05**
FGFR2 Amp vs RTK Absent	1.227 (0.609, 2.471)	0.6
MET Amp vs RTK Absent	1.358 (0.564, 3.269)	0.5
Stage2 vs Stage 1	1.968 (0.816, 4.744)	0.1
Stage3 vs Stage 1	4.969 (2.325, 10.621)	**3.5E-05**
Stage4 vs Stage 1	8.414 (3.887, 18.213)	**6.5E-08**
Poorly Differentiated vs Moderately to well Differentiated	0.996 (0.665, 1.491)	1.0

Significant p-values are shown in bold type.

Table S5a: Univariate analysis of prognostic associations for individual RTK/*KRAS* amplifications
(Outcome: overall survival, relative to patients lacking RTK or *KRAS* amplifications)

Model 1 (Predictors: RTK/*KRAS* Amp vs RTK/KRAS Absent)	**Hazard Ratio (95% CI)**	**P-value**
EGFR Amp vs RTK/*KRAS* Absent	1.306 (0.647, 2.638)	0.5
ERBB2 Amp vs RTK/*KRAS* Absent	3.141 (1.714, 5.756)	**0.0002**
FGFR2 Amp vs RTK/*KRAS* Absent	1.217 (0.603, 2.453)	0.6
MET Amp vs RTK/*KRAS* Absent	2.993 (1.291, 6.940)	**0.01**
KRAS Amp vs RTK/*KRAS* Absent	2.116 (1.155, 3.879)	**0.02**

Significant p-values are highlighted in bold type.

Table S5b: Multivariate analysis comparing *KRAS* and RTK Amplifications with tumor stage and grade

Model 2 (Predictors: RTK/*KRAS* Amp, Stage and Grade)	**Hazard Ratio (95% CI)**	**P-value**
EGFR Amp vs RTK/*KRAS* Absent	1.231 (0.600, 2.528)	0.6
ERBB2 Amp vs RTK/*KRAS* Absent	3.909 (2.082, 7.340)	**2.2E-05**
FGFR2 Amp vs RTK/*KRAS* Absent	1.296 (0.639, 2.631)	0.5
MET Amp vs RTK/*KRAS* Absent	1.440 (0.594, 3.493)	0.4
KRAS Amp vs RTK/*KRAS* Absent	1.455 (0.790, 2.682)	0.2
Stage2 vs Stage 1	1.935 (0.802, 4.670)	0.1
Stage3 vs Stage 1	4.786 (2.230, 10.269)	**5.8E-05**
Stage4 vs Stage 1	8.053 (3.702, 17.515)	**1.4E-07**
Poorly Differentiated vs Moderately to well Differentiated	1.012 (0.675, 1.517)	1.0

Significant p-values are highlighted in bold type.

Table S6. Clinical Characteristics of GC Patient Cohorts Used in Gene Expression Analysis

	SG U133A (51)	SG U133B (248)	AU(70)	UK(29)
Age				
range	38-86	23-92	32-85	53-84
mean,S.D	64.0, 11.2	65.4, 12.5	65.5, 12.5	71.7, 9.11
Gender				
Male	33	161	48	16
Female	18	87	22	13
Lauren classification				
Intestinal	27	138	34	20
Diffuse	11	86	30	6
Mixed	13	24	6	3
Grade				
Moderate to well differentiated	20	96	24	13
Poorly differentiated	30	149	46	15
Unknown	1	3	0	1
Stage				
1	10	40	13	6
2	11	43	16	4
3	15	88	33	15
4	12	76	8	4
Unknown	3	1	0	0

Table S7: Multivariate analysis comparing high *FGFR2* gene expression (>2-fold mean level in normal gastric tissues) with tumor stage and grade

(Outcome: overall survival, relative to patients with low *FGFR2* expression (< 2-fold mean level in normal gastric tissues)

Model 1 (Predictors: *FGFR2* Expression, Stage and Grade)	**Hazard Ratio (95% CI)**	**P-value**
FGFR2 High Expression vs *FGFR2* Low Expression	1.321 (0.966, 1.807)	0.08
Stage 2 vs Stage 1	1.643 (0.924, 2.922)	0.09
Stage 3 vs Stage 1	4.593 (2.807, 7.514)	**1.3e-09**
Stage 4 vs Stage 1	8.440 (5.009, 14.221)	**1.1e-15**
Poorly Differentiated vs Moderately to well Differentiated	0.942 (0.718, 1.235)	0.7

Significant p-values are highlighted in bold type.

Supplementary Figures

Figure S1: Copy Number PRofiles of matched gastric tumor and non-malignant samples

Three representative paired primary GC tumor/normal samples are shown (IDs 2000068, 57689477 and 980021). The x-axis represents chromosomes 1 to 22 and chromosomes X and Y, y-axis represents the extent of copy number amplifications/deletions. The proportion of CNAs for each sample are indicated respectively as a percentage of the whole genome.

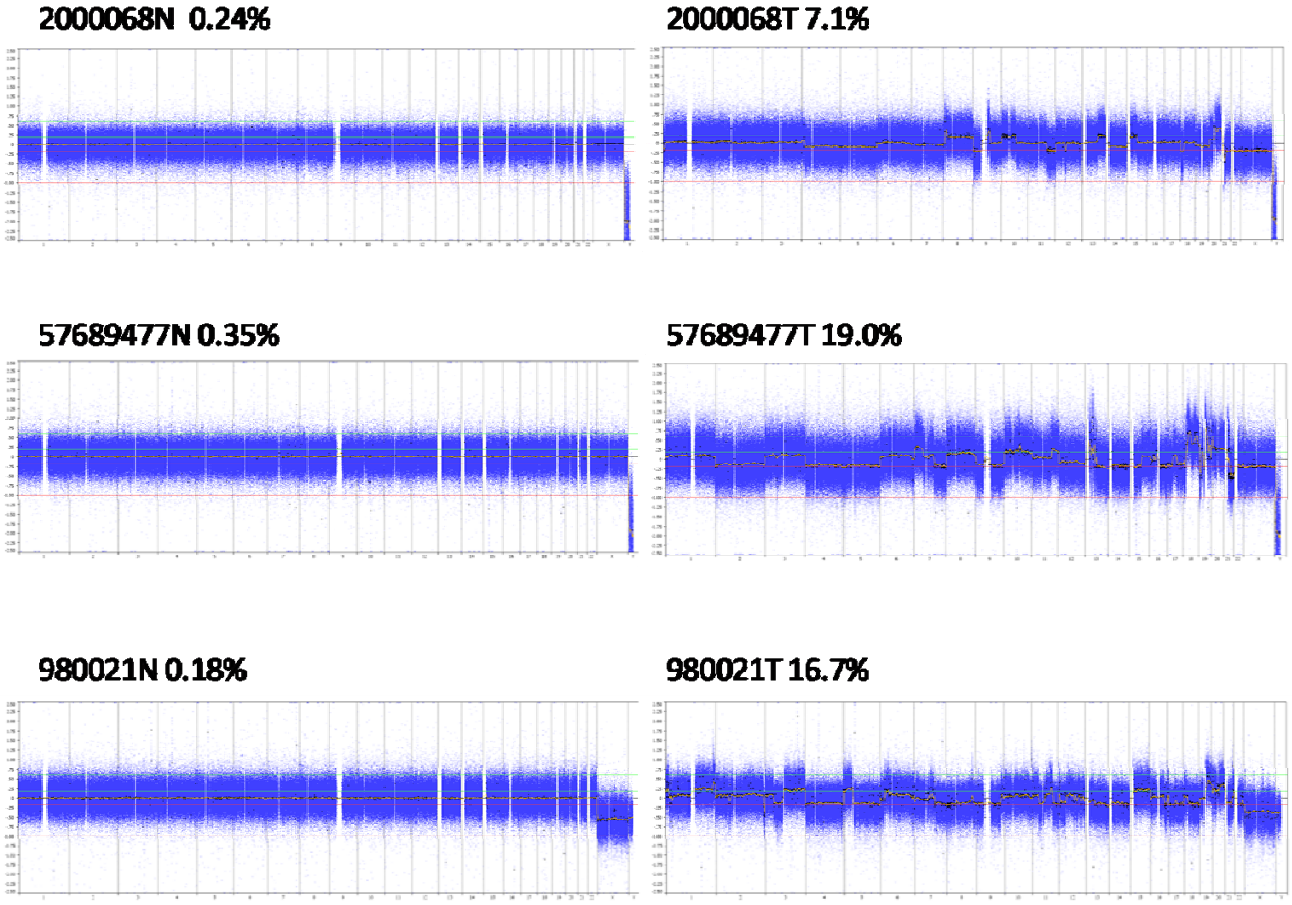


Figure S2: *ERBB2* Copy Number and Protein Expression in GC.

Two primary GCs are shown (IDs 970010 (A,B) and 2000472 (C,D)). (A) Tumor 970010 is predicted to exhibit *ERBB2* copy number amplification. The top graph represents a segment of Chromosome 17 where ERBB2 resides. The *ERBB2* region is marked by yellow boundaries. The y-axis represents the extent of copy number amplification. The bottom graph is a close up of the region, where the ERBB2 gene is marked by a red box. (B) Immunohistochemical (IHC) analysis of ERBB2 reveals high ERBB2 protein expression (IHC 3+) in 970010. (C) Tumor 2000472 is predicted to show normal/neutral ERBB2 copy number levels. Boundaries of the yellow and red boxes are the same as in (A). (D) IHC analysis of ERBB2 reveals absence of ERBB2 protein expression (IHC 0) in 2000472.

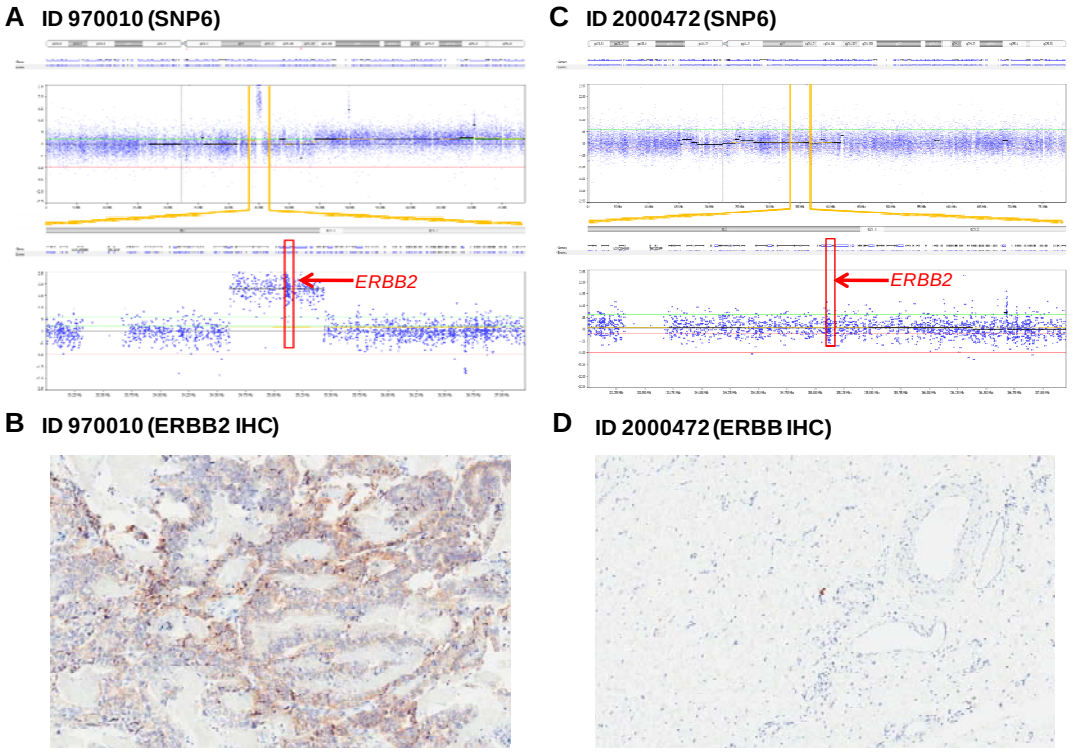


Figure S3: *CSMD1* Expression in GC

Full sections of GCs (n=42) were subjected to CSMD1 immunohistochemistry. (A) CSMD1 expression in normal gastric epithelium (black triangle) and loss of expression in intestinal metaplasia (blue triangle). (B) Loss of CSMD1 expression in a diffuse-type GC. Staining in adjacent normal gastric epithelial (black triangle) cells and within endothelial cells within the tumor serves as a positive internal control. (C) Strong membranous CSMD1 staining in an intestinal-type GC. Approximately 40% of GCs show absent or reduced CSMD1 expression relative to normal gastric epithelium.

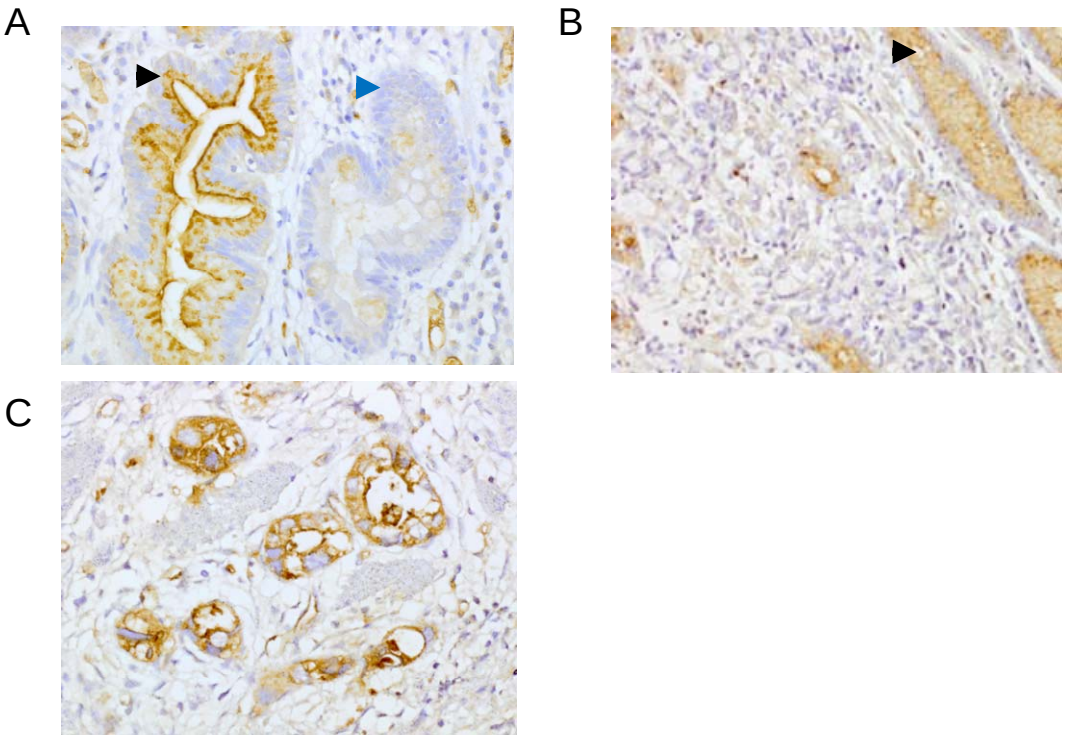


Figure S4. Hierarchical clustering of GCs using genes exhibiting recurrent focal amplifications

In the heatmap, each row represents a different focally amplified amplified gene from the highest recurrent regions (Table 1 in Main Text). Each column represents an individual tumor exhibiting amplifications of these genes (total 113 tumors). The red color gradient (top right) highlights the degree of copy number amplification. Hierarchical clustering was performed both row and column-wise. The highlighted region identified *ERBB2* and *CCNE1*, which exhibit a significant co-amplification pattern as identified by DRP.

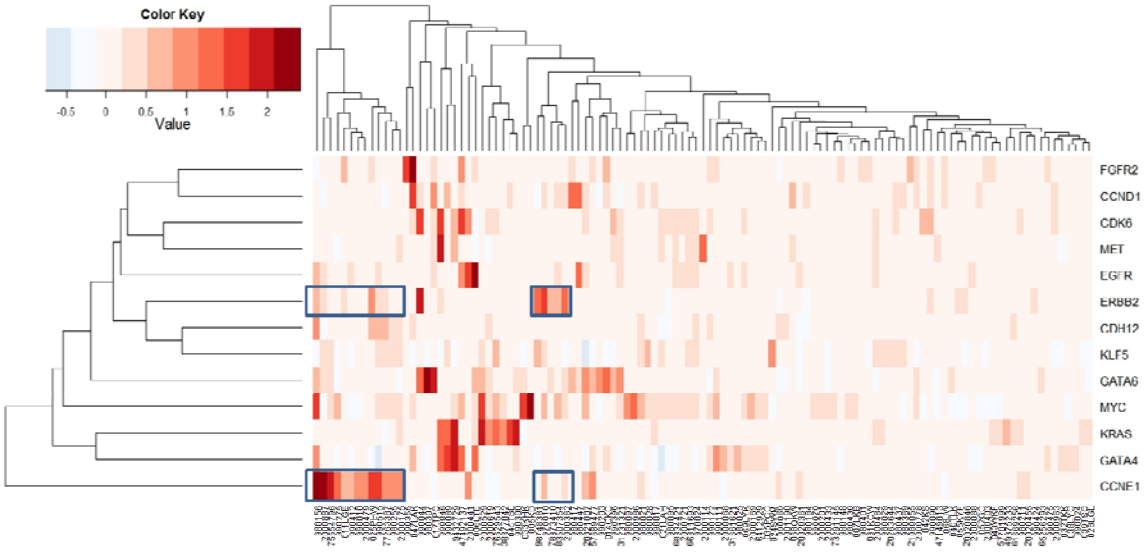


Figure S5: Network Diagram Showing Relationship of RTK Signaling to RAS

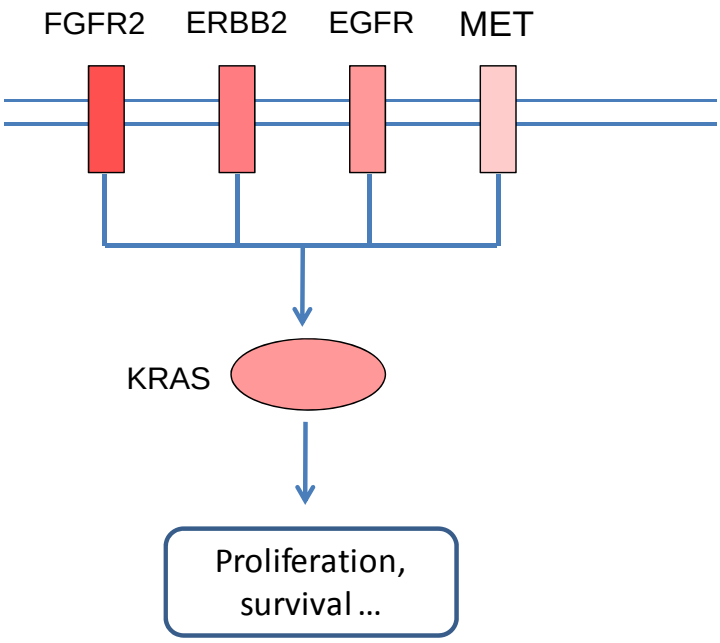


Figure S6: Kaplan-Meier Survival Analysis based on *KRAS* Copy Number Status
A) KM survival graph comparing outcomes of patient with tumors exhibiting KRAS amplification against patients with no/low KRAS CNA irrespective of RTK amplification status. The 17 KRAS-amplified patients correspond to the same patients identified in the Figure 3A heat-map presented in the Main Text.
B) KM survival graph comparing outcomes of patients with tumors exhibiting high KRAS copy number, defined as the top 25% of patients exhibiting a high SNP6 logRatio (high KRAS LR) vs the remaining 75% of patients (low KRAS LR).

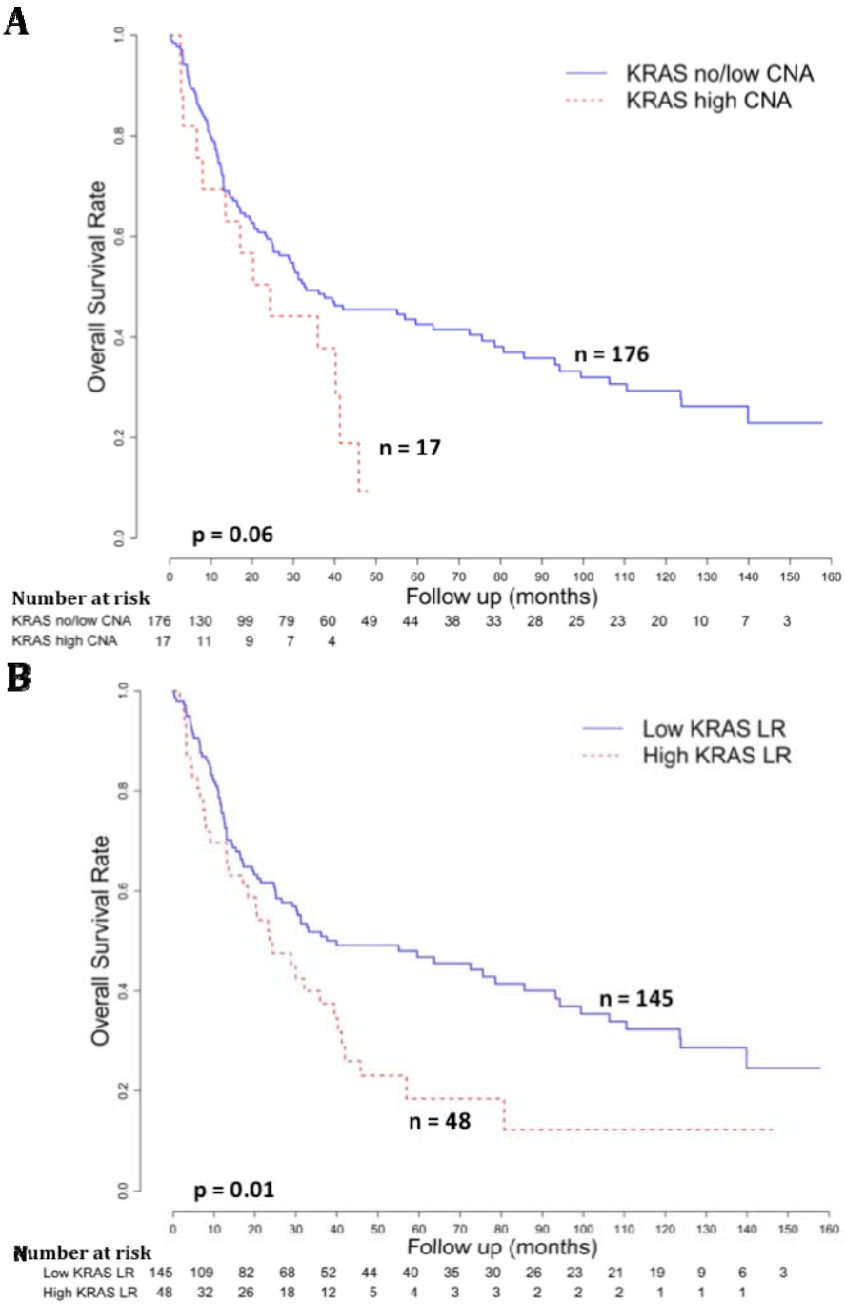


Figure S7: Phenotypic Effects of *KRAS* siRNA Knockdown in *KRAS*-amplified, Mutated and Wild-type GC Lines

KRAS siRNAs or Control Scrambled siRNAs were applied to four GC cell lines – YCC1 and MKN1 (KRAS-amplified), AGS (KRAS-mutated; G12D), and TMK1 (KRAS non-amplified and wild-type). For each cell line, KRAS knockdown was confirmed at the protein level (Western blots – not treated (--), scrambled siRNA (Ctl), *KRAS* siRNA (KRAS)). Cell proliferation was measured 48-96 h after knockdown, comparing *KRAS* siRNA-treated cells to control siRNA treated cells (Numbers above bars are p-values comparing KRAS siRNA vs control siRNA treated cells). Significant reductions in cell proliferation are observed in KRAS-amplified and KRAS-mutated lines (P<0.05), but no significant effects are seen in wild-type TMK1 cells. Similar effects were observed with two non-overlapping *KRAS* siRNAs. All experiments were repeated a minimum of three independent times.

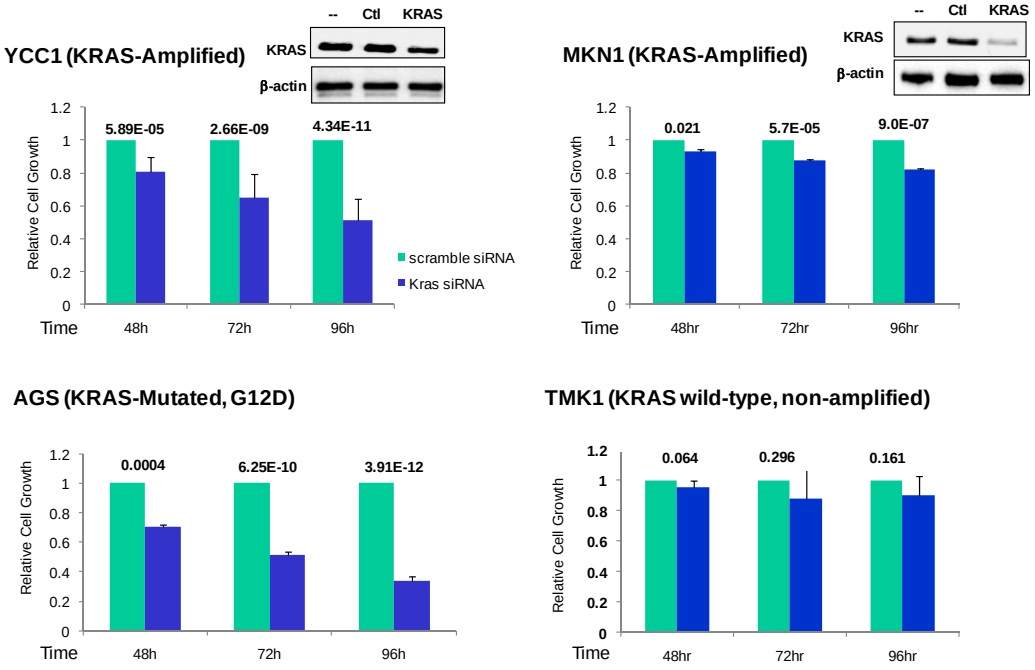


Figure S8: qPCR Analysis of *FGFR2* Amplification in GC

Quantitative PCR of genomic DNA from 63 GC primary tumors, performed using *FGFR2* primers flanking the GISTIC identified amplification peak in intron 2. A) The X-axis shows samples classified into three categories - normal (black), tumors without *FGFR2* amplification (grey), and tumors with *FGFR2* amplification (red, including samples with high copy number level (Figure 3A) and intron 2 copy number). The Y-axis indicates the qPCR DNA level. The horizontal broken black line indicates the cutoff for qPCR amplification. A Fisher exact test shows that samples with high *FGFR2* qPCR values are associated with *FGFR2* amplification (p = 0.0006). Samples were internally normalized against a LINE1 control. B) An X-Y scatter plot of *FGFR2* qPCR values and *FGFR2* copy number based on SNP arrays. x-axis indicates qPCR value and y-axis represents the copy number logRatio. Red, orange and grey colored samples represent high CNA (Figure 3A), focal high CNA (intron 2) and no/low CNA samples respectively. The Spearman correlation is 0.84, showing a positive correlation between *FGFR2* copy number and qPCR values (p < 2.2e-16)

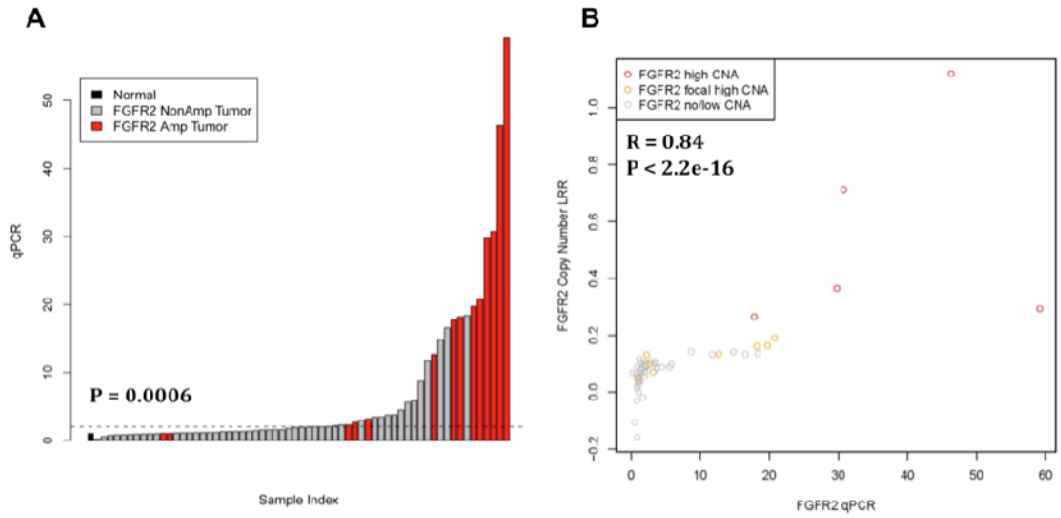


Figure S9. Scatter plot of gene expression and copy number for FGFR2

The figure shows an XY scatter plot of *FGFR2* gene expression and *FGFR2* copy number. x-axis - log2 transformed mRNA expression values; y-axis - copy number logRatio. Red, grey and blue colored samples represent high CNA, low/no CNA, and normal samples respectively. Spearman correlation value is indicated as R = 0.38, with p value = 3.3e-7.

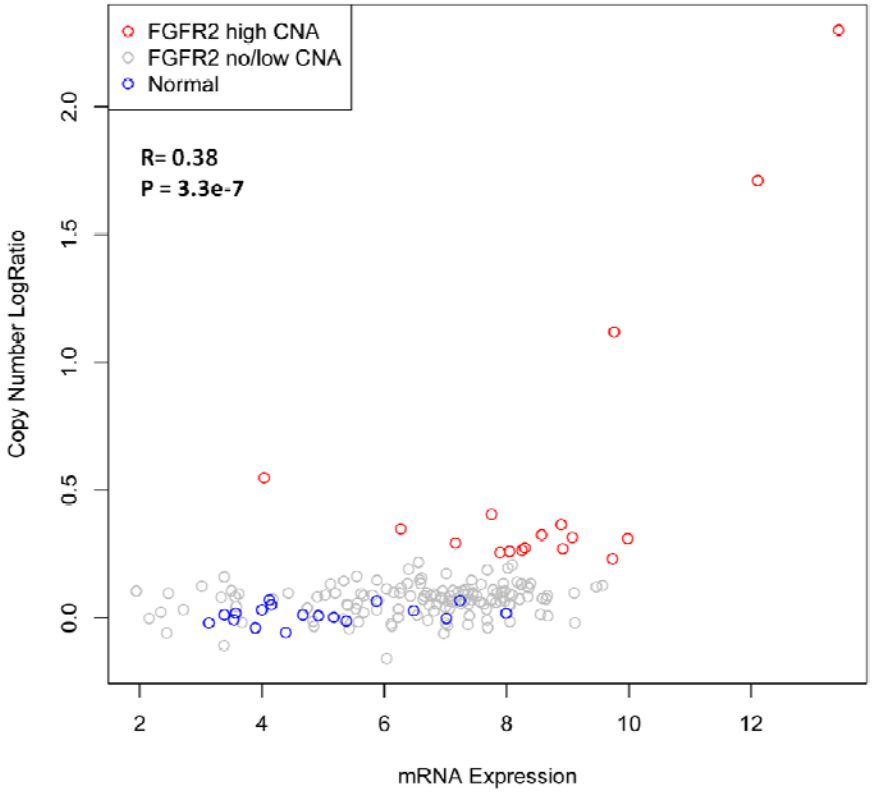


Figure S10: Relationship between Copy Number and Gene Expression for ATE1 and BRWD2, Genes Adjacent to FGFR2

Primary GCs exhibiting genomic amplification of the *FGFR2* locus were also assessed for relationships between copy number status and gene expression in A,C) *ATE1* (upstream of *FGFR2*) and B,D) *BRWD2* (downstream of *FGFR2*). For each gene, mRNA expression was compared across three categories, each represented by a box-plot - non-malignant gastric tissues (normal) (n =100 for A,B, n =18 samples with available copy number information for C,D), tumors exhibiting no/low *FGFR2* gene locus CNA (n = 139), and tumors exhibiting high *FGFR2* gene locus CNA (n = 17). *ATE1* and *BRWD2* expression was inferred from Affymetrix microarrays (*ATE1* 234584_s; *BRWD2* probe 218090_s_at).

A) *ATE1* expression levels in amplified tumors are observed to be significantly higher than normal samples (P=0.004, Wilcoxon test, underlined). However this significance level is weaker than that observed for *FGFR2* (p=1.7e-7, see Main Text).

B) *BRWD2* expression levels in amplified tumors are not significantly higher than normal samples (P=0.3, Wilcoxon test, underlined).

C) XY scatter plot of *ATE1* expression with copy number information. Spearman correlation R is 0.16 with p value = 0.04.

D) XY scatter plot of *BRWD2* expression with copy number information. Spearman correlation R is 0.16 with p value = 0.04.

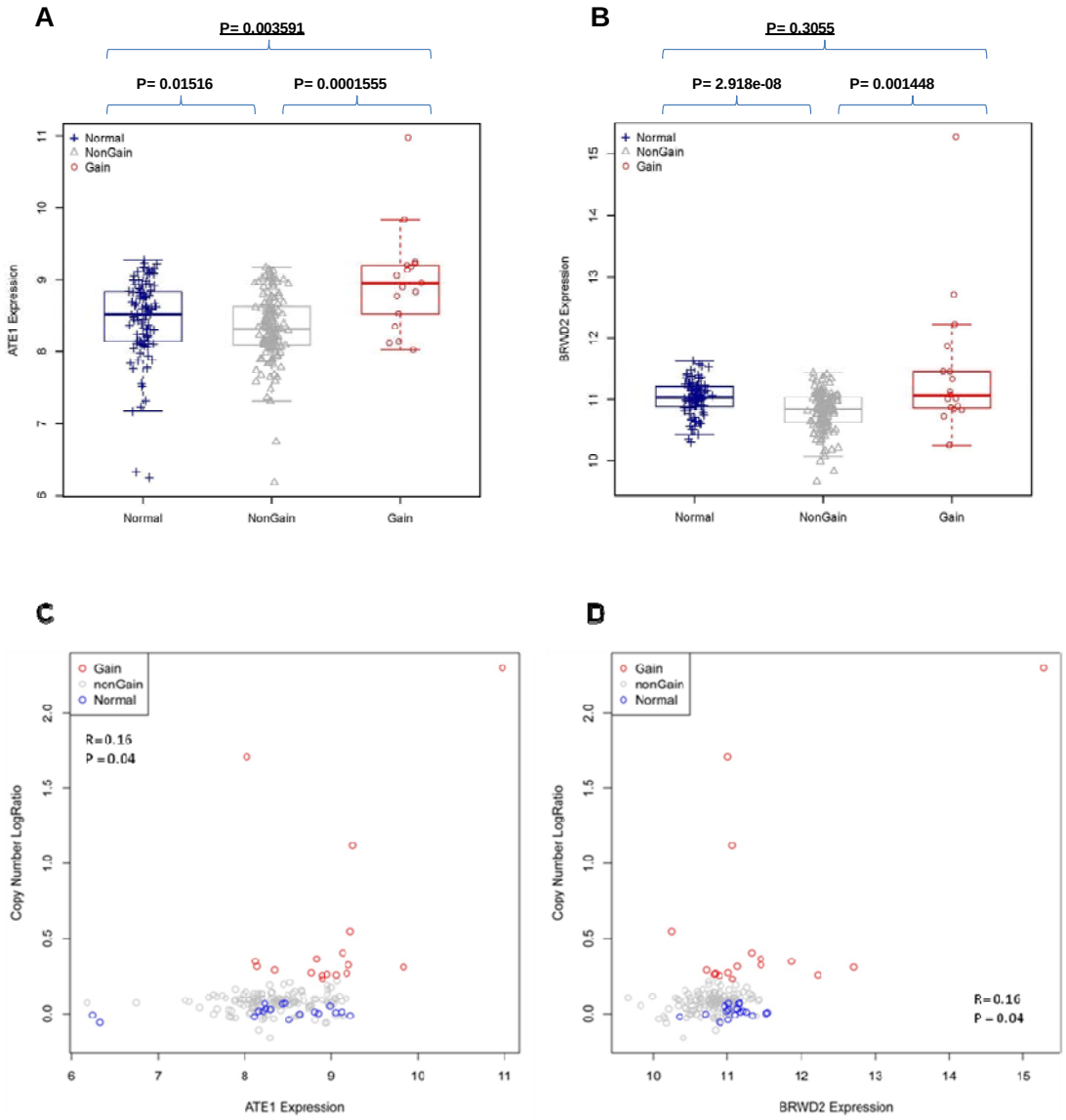

A
P= 0.003591
P= 0.01516
P= 0.0001555
Normal
NonGain
Gain
ATE1 Expression
B
P= 0.3055
P= 2.918e-08
P= 0.001448
BRWD2 Expression
C
Gain
nonGain
Normal
R=0.16
P=0.04
Copy Number LogRatio
ATE1 Expression
D
BRWD2 Expression

Figure S11: *FGFR2* Overexpression in GCs Relative to Normal Gastric Samples

The graph depicts 236 normal gastric tissues and 399 primary gastric tumors, arranged along the x-axis in ascending order of their *FGFR2* expression level. FGFR2 gene expression levels were inferred using Affymetrix microarrays (FGFR2 probe 211401_s_at). At the cut-off threshold level of >2x the average level in normal tissues (dotted line), approximately 18% of gastric tumors exhibit high *FGFR2* levels (marked in red).

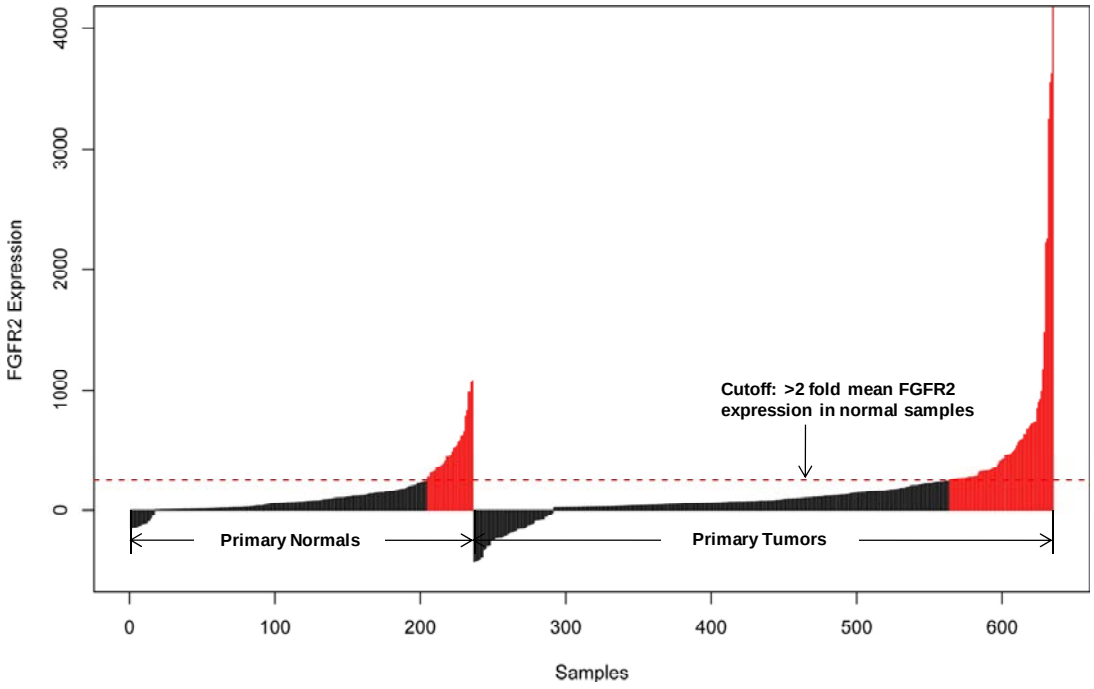
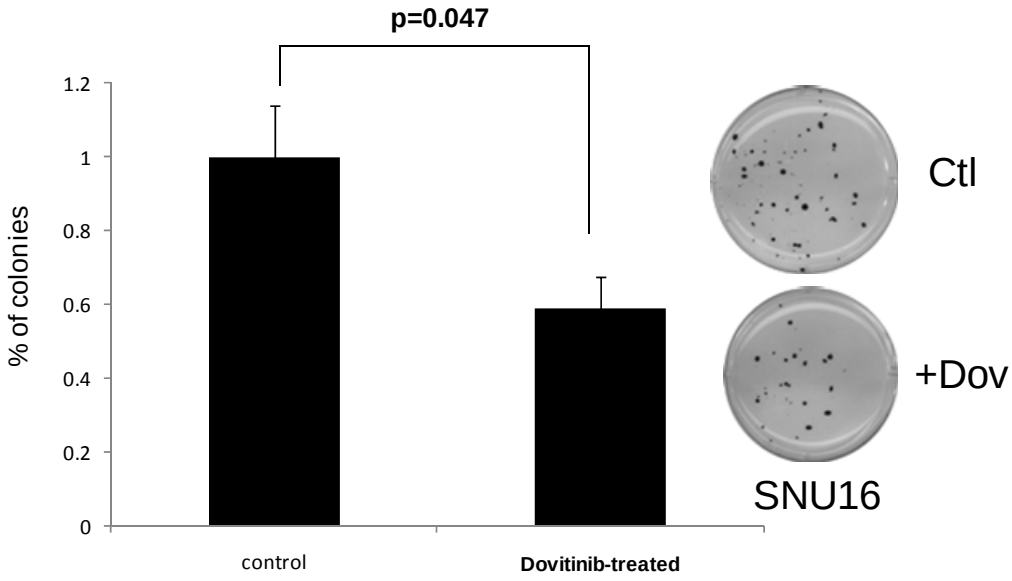


Figure S12: Inhibition of Soft Agar Colony Growth by Dovitinib (SNU-16)

FGFR2-amplified SNU16 cells were treated with dovitinib at the GI50 concentration (0.17uM) for 48 hrs, and soft-agar colony formation monitored over the subsequent 3-4 weeks. Representative plates are shown (Ctl : mock treated, + Dov : Dovitinib treated). Bar graphs depict results from a minimum of three independent experiments.



References

1. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data.* Biostatistics, 2004. **5**(4): p. 557-72.
2. Beroukhim, R., et al., *Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 20007-12.
3. Ding, L., et al., *Somatic mutations affect key pathways in lung adenocarcinoma.* Nature, 2008. **455**(7216): p. 1069-75.
4. Hofmann, M., et al., *Assessment of a HER2 scoring system for gastric cancer: results from a validation study.* Histopathology, 2008. **52**(7): p. 797-805.
5. Kamal, M., et al., *Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma.* Breast Cancer Res Treat, 2010. **121**(3): p. 555-63.
6. Ooi, C.H., et al., *Oncogenic pathway combinations predict clinical prognosis in gastric cancer.* PLoS Genet, 2009. **5**(10): p. e1000676.
7. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.