Description of 250K SNP Array Analysis

250K SNP arrays were analyzed as described in Beroukhim et al.[1] A brief overview of our analysis pipeline using the GenePattern analysis platform (www.broadinstitute.org/cancer/software/genepattern) is provided below.[2] Chromosomal gains and losses across chromosome 4 are presented in this study. For a full genomic analysis of all chromosomal aberrations in this panel of 733 NSCLC samples, see Beroukhim et al. (in review).[3]

1. Signal Intensity Normalization

To generate SNP-level intensities from raw probe-level intensities, the SNPFileCreator module in Genepattern was used. Here, raw probe-level signal intensity data from cel files were normalized to a baseline array with median intensity, using the quantile normalization method.[4] The signal intensity of each SNP was then obtained using a model-based (PM/MM) method.[5]

Parameters:
Chip type: 250K Sty
normalization method: Quantile Normalization (dChip)
reference method: Median
model method: PM/MM Difference Model (dChip)
allele specific: Non Allele-Specific
sort snp file: Sort

Input Files: *.cel file for each sample.
Output File: *.snp file containing SNP-level intensities for each sample.

2. Preprocessing and Copy Number Estimation

Signal intensity data was further preprocessed using the GISTICPreprocess module with default parameters in Genepattern to perform quality control and estimate copy number in tumor samples. For quality control, low-quality samples are identified, using signal intensity data histograms that do not display distinct copy number peaks, and discarded. Duplicate samples identified by similar SNP genotype profiles are also discarded. To determine copy number from SNP-level intensity values, tumor profiles are normalized against a set of non-tumor normal control samples that share similar noise profiles to the tumor sample being normalized. For each tumor, profiles are normalized against five samples from a reference collection of normal samples that are closest to the tumor being normalized measured by Euclidean distance between the $\log_2$-transformed signal profiles.

Parameters:
use paired normals: 0
quality control tumors: 1
normal selection method: 1
n closest n: 5
correct batch effect: 0

batch correct pvalue: 0.001
verbosity: 0

Input File: *.snp file created by SNPFileCreator and a sample info file (*.txt) containing information about the arrays.
Output Files: *.cn file containing preprocessed intensity values for each sample and a core array list file (*.txt) containing samples suggested for use in the GISTIC module (i.e. duplicates, normals, and low-quality samples removed).

3. Segmentation

Copy number profiles were segmented with the Gain and Loss of DNA (GLAD) algorithm[6] with default parameters using the GLAD module in Genepattern.

Input Files *.cn file created by GISTICPreprocess.
Output File: *.seg file that describes each altered region: start, end, number of SNPs, and smoothed copy number.

4. GISTIC

To determine recurrent chromosomal aberrations, normalized and segmented tumor profiles were analyzed with the GISTIC algorithm[1] in Genepattern.

Parameters:
refgene file: Human Hg18
amplifications threshold: 0.3
deletions threshold: 0.3
join segment size: 6
qv thresh: 0.25
remove X: Yes

Input Files: *.cn file created by GISTICPreprocess, a markers file (*.txt) that identifies the names and positions of the markers in the original dataset (before segmentation), and a cnv file (*.txt) containing markers from regions with known germline copy number variations as listed in http://projects.tcag.ca/variation.

Output Files: All lesions file (all_lesions_file.txt) summarizing results from the GISTIC run, Amplification genes file (Amp_genes.txt) listing all significant amplifications found, and a GISTIC scores file (scores.gistic.txt) listing information about aberrations found.

References

1. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A 2007; 104:20007-12.

2. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet 2006; 38:500-1.

3. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of copy number alterations across multiple human cancers. Nature 2009;

4. Li C and Wong WH. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2001; 2.

5. Li C and Wong WH. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. Proc Natl Acad Sci U S A 2001; 98:31-6.

6. Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 2004; 20:3413-22.

Supplementary Figure Legends

Figure S1. Raw (black) and smoothed (red) copy number data for NCI-H1703 defining a 4q12 amplification of *PDGFRA*. Estimated copy number values (y axis) are plotted according to position on chromosome 4 (x axis). Genomic positions of *SCFD2*, *FIP1L1*, *LNX1*, *CHIC2*, *PDGFRA*, and *KIT* are shown along the x axis.

Figure S2. (A) Immunostaining for KIT expression in primary sample 9-112-114, an adenocarcinoma with high-level amplification of 4q12. Strong cytoplasmic and membrane staining of KIT is seen in tumor cells. (B) Summary of IHC staining for KIT by level of staining and 4q12 amplification status (determined by FISH).

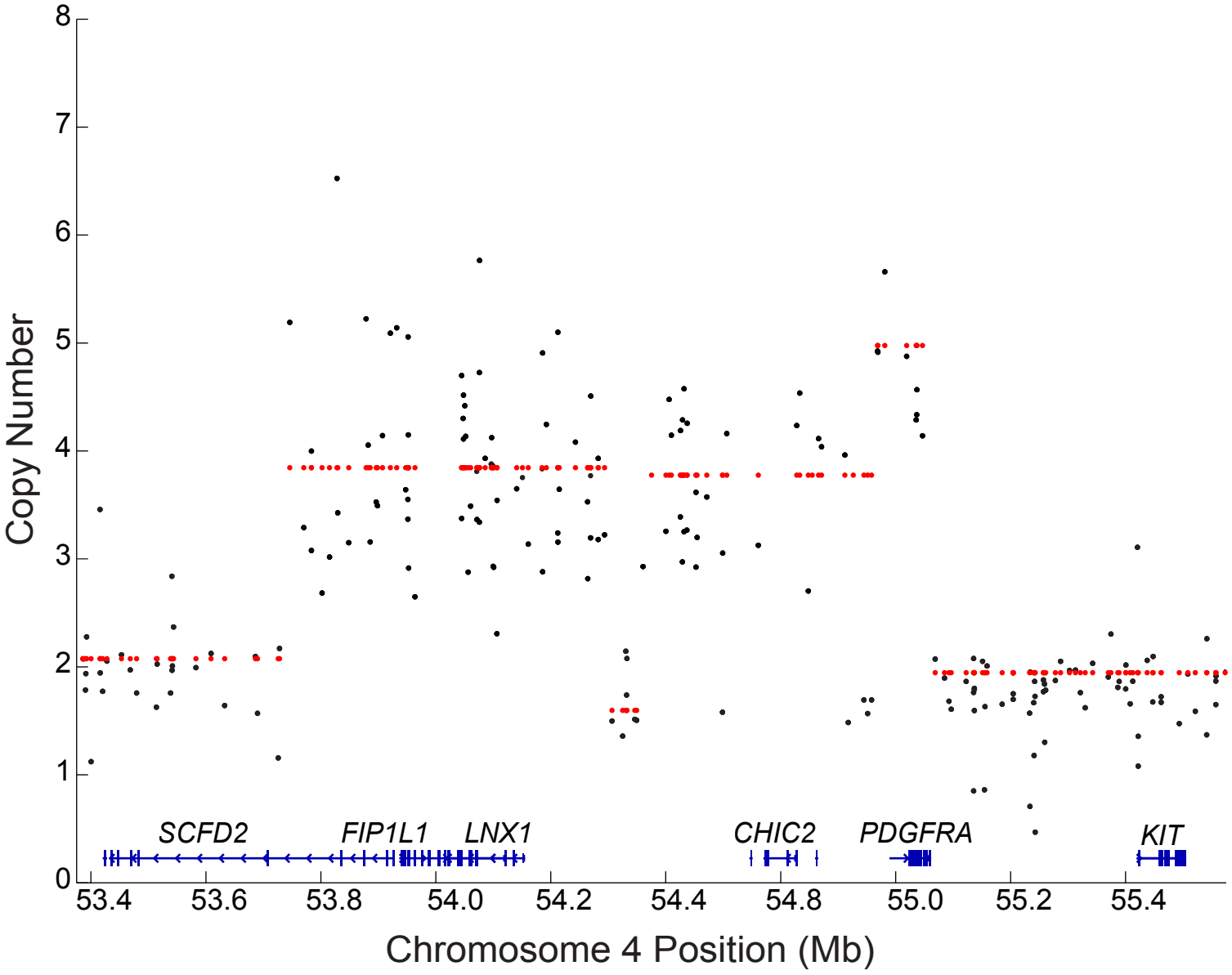
Figure S3. (A) Western blot analysis of PDGFRα in six different 4q12 amplified (NCI-H1703, NCI-H661, NCI-H1819, NCI-H1838, NCI-H23 and HCC366) and one non-amplified (HCC15) NSCLC cell lines. NCI-H1703 cells show increased PDGFRα expression as compared to other NSCLC cell lines. (B) *PDGFRA* transcript fold change determined by quantitative RT-PCR, *PDGFRA* gene copy number status determined by quantitative PCR, and 4q12 amplicon length determined by SNP array are reported.

Figure S4. Anchorage independent growth of NCI-H1703 cells is dependent on PDGFRA activity. (A and B) Infection with three independent hairpins (#1, #3 and #5) inhibited colony formation in soft agar in NCI-H1703 cells overexpressing PDGFRA, but not in HCC15 cells. All results are normalized to survival or colony formation by cells infected with shGFP.

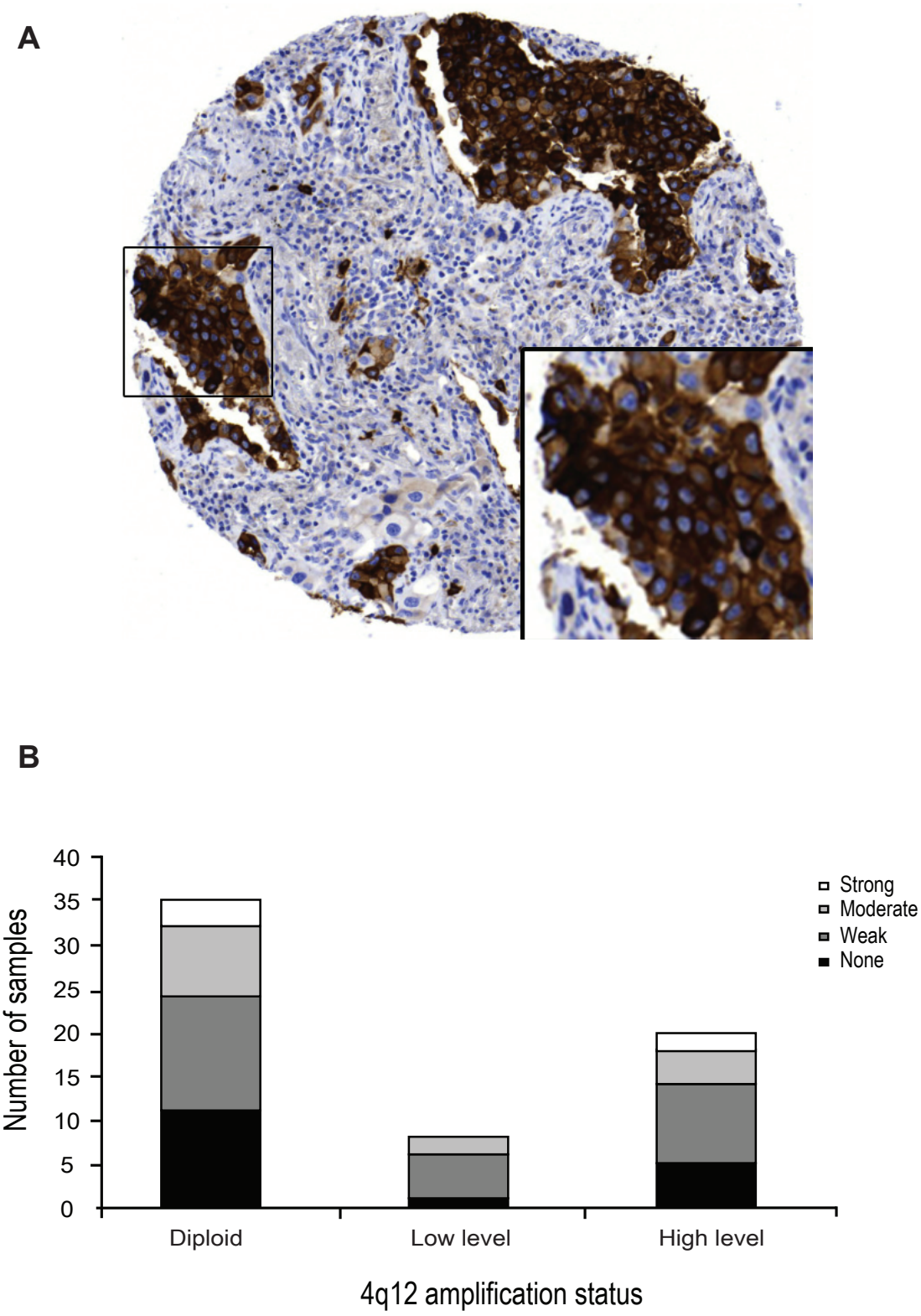
Figure S5. (A and B) WST survival assays performed with seven NSCLC cell lines (NCI-H1703, NCI-H1819, HCC15, NCI--H23, NCI-H366, NCI-H1838 and NCI-H661) after 4 days treatment with sunitinib and imatinib. IC50s are indicated.

Figure S6. Treatment of NCI-H1703 cells with kinase inhibitors decreases colony formation ability. (A and B) Treatment of NCI-H1703 cells with sunitinib and imatinib resulted in a marked decrease in colony formation in soft agar with IC50s in the 20 nM and 81 nM range, respectively, whereas similar treatment of the HCC15 cell line without *PDGFRA* amplification had no significant effect.

Supplementary Figure S1.

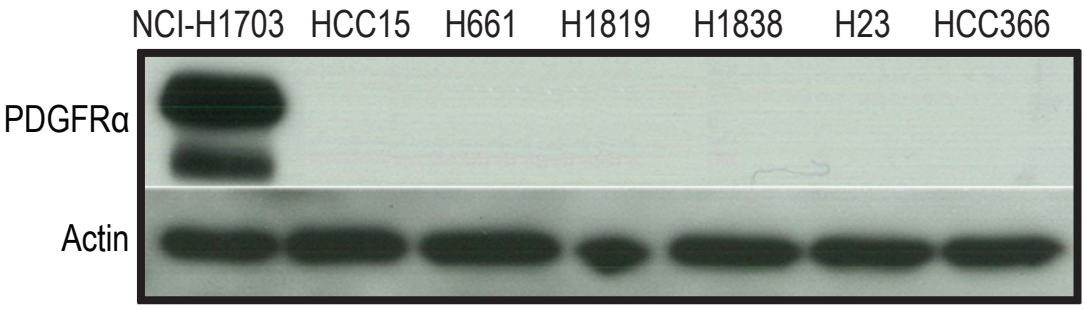


Supplementary Figure S2.



Supplementary Figure S3.

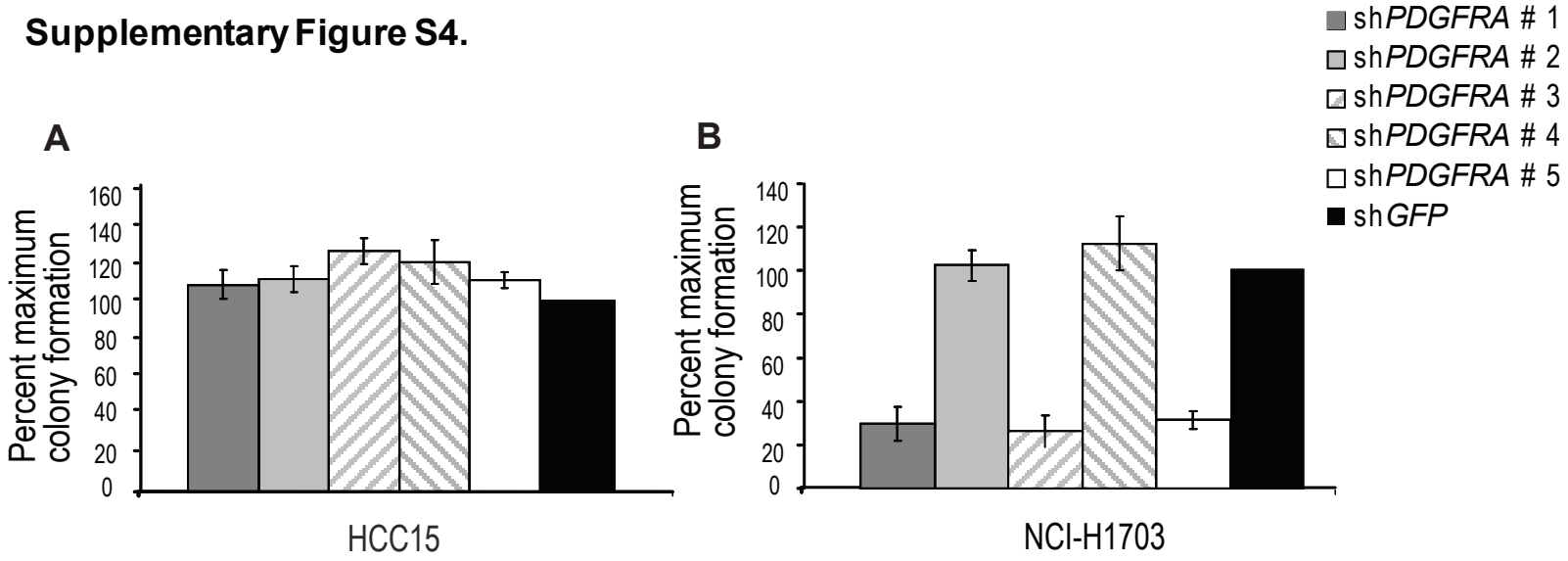
A



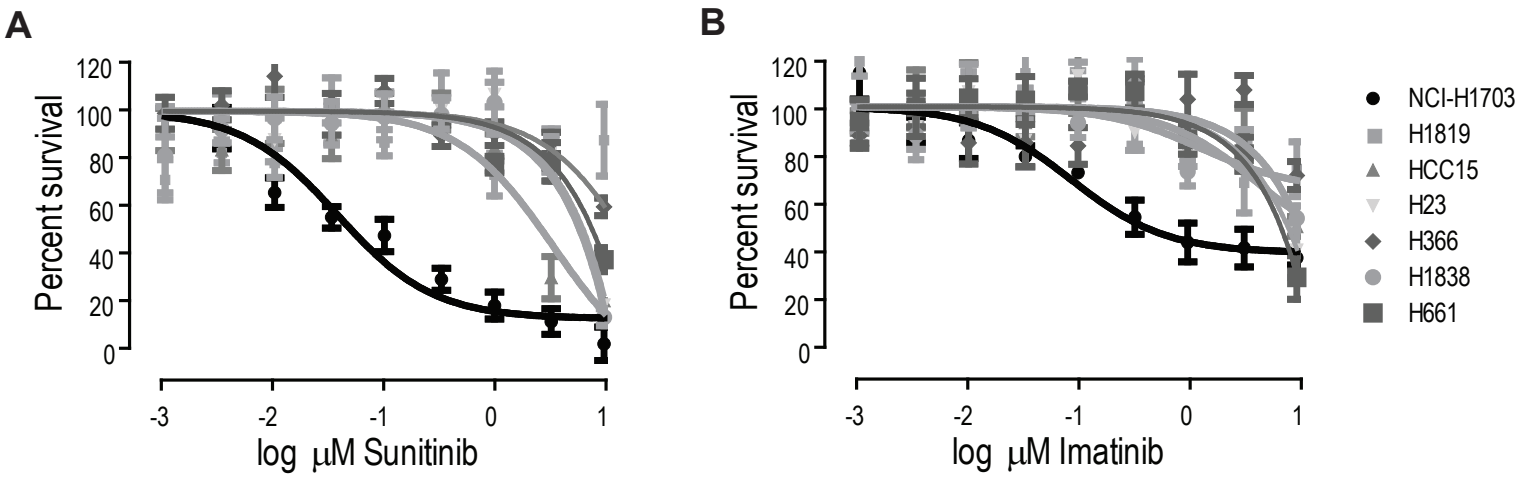
B

PDGFRA Transcript Fold Change (qRT-PCR)	358.14	0.61	0.44	0.10	0.16	0.16	0.017
PDGFRA Genomic Copy Number (qPCR)	24.22	2.00	5.58	13.20	8.85	7.98	2.71
4q12 Amplicon Length (Mb)	1.28	n/a	10.35	15.75	25.9	13.65	48.4
Number of Genes in 4q12 amplicon	4	n/a	35	48	97	54	207

Supplementary Figure S4.



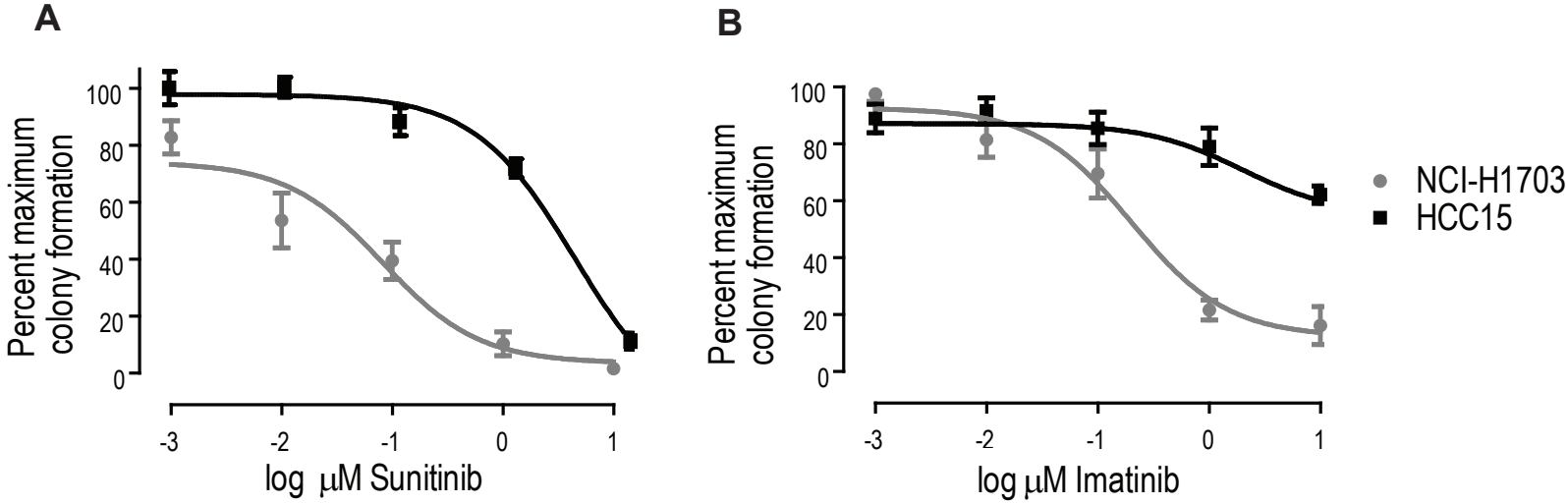
Supplementary Figure S5.



	NCI-H1703	H1819	HCC15	H23	H366	H1838	H661
Sunitinib IC50 (nM)	37	>10000	3481	>10000	>10000	>10000	>10000

	NCI-H1703	H1819	HCC15	H23	H366	H1838	H661
Imatinib IC50 (nM)	87	1333	>10000	>10000	>10000	4495	>10000

Supplementary Figure S6.



	NCI-H1703	HCC15
IC50 (nM)	81	3200

	NCI-H1703	HCC15
IC50 (nM)	20	1900