COMPREHENSIVE STUDY OF GENETIC ALTERATIONS IN GASTRIC CANCER

DENG NIANTAO

(B.Sc), SJTU

(M.Sc), NUS

A THESIS SUBMITTED FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

NUS GRADUATE SCHOOL FOR INTEGRATIVE SCIENCES AND ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2013

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

DENG NIANTAO

22 Nov 2013

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to my supervisor Professor Patrick Tan and co-supervisor Dr. Goh Liang Kee for their invaluable advice and guidance through the four years of my PhD study. Without their support and encouragement, this work would not have been accomplished. I have learnt a lot from working and countless discussion with them, and it was great and memorable experience for me. I feel extremely grateful for having this opportunity to study under their supervision.

I would also like to thank everyone in the Centre for Computational Biology at Duke NUS Graduate Medical School, who has constantly answered my questions and helped me to solve different kinds of problems. I am thankful for their time and help, and I would also like to offer my time if they need me to help.

I would also like to thank many people who had helped in this project. They are Hannah Wang, Kakoli Das, Tao Jiong, Shenli Zhang, Ming Hui Lee, Jeanie Wu, Glenn Goh, Angie Tan, Dianne Poh, who has spent a significant amount of time to help in doing experiments and generating data. And thanks also goes to Iain Tan, Zhengdeng Lei, Jun Hao Koo, Zhu Feng, who had help in part of the data preparation, and many collaborators from local and overseas institutions who provided reagent and data for this project.

The work was supported by grants NMRC grant TCR/001/2007, BMRC 10/1/24/19/655 and core grants from Duke-National University of Singapore and Cancer Sciences Institute of Singapore. This work was also supported by an ASCO grant, Singhealth Talent Development grant, and Khoo Discovery Award (KDP/2008/0002 and KDP/2009/0006). I wish to acknowledge NUS Graduate School for Integrative Sciences and Engineering in National University of Singapore and Duke NUS Graduate Medical School, for providing various supports from course education to the management of administrative matters.

I would also like to acknowledge my great gratitude to my parents, who are always giving me their full support throughout the journey.

TABLE OF CONTENTS

SUMMARY

Gastric cancer (GC) is a major gastrointestinal malignancy for which targeted therapies are emerging as treatment options. GC patients with *ERBB2*-amplified tumors can clinically benefit from *ERBB2*-targeted therapies. Similar to *ERBB2*, several other molecularly-targeted therapies are currently being evaluated in GC. Little is known regarding which molecular targets are concurrently expressed in the same gastric tumors, or independently in different tumors. We sought to identify the most prevalent molecular targets in GC and to elucidate systematic patterns of exclusivity and co-occurrence among these targets, through comprehensive genomic analysis of a large panel of GCs.

Using high-resolution single nucleotide polymorphism (SNP) arrays, we profiled copy number alterations in a panel of 233 gastric cancers (193 primary tumors, 40 cell lines) and 98 primary matched gastric non-malignant samples. For selected alterations, we evaluated their impact on gene expression and clinical outcome.

We identified 22 recurrent focal alterations (13 amplifications and 9 deletions). These included both known targets (*FGFR2*, *ERBB2*), and also novel genes in GC (*KLF5*, *GATA6*). We found that receptor tyrosine kinase (RTK)/RAS alterations are frequent in GC and demonstrate, for the first time, that these alterations occurred in a mutually exclusive fashion, with *KRAS*-gene amplifications highlighting a clinically relevant but previously underappreciated GC subgroup. We also showed that *FGFR2*-amplified GCs are sensitive to dovitinib, an orally bioavailable FGFR/VEGFR

targeting agent, potentially representing a subtype-specific therapy for *FGFR2*-amplified GCs.

Our study demonstrated the existence of five distinct GC patient subgroups, defined by the signature genomic alterations *FGFR2* (9% of tumors), *KRAS* (9%), *EGFR* (8%), *ERBB2* (7%), and *MET* (4%). Collectively, these subgroups suggested that at least 37% of GC patients may be potentially treatable by RTK/RAS directed therapies.

Overall the copy number landscape can provide a useful resource when integrated with other type of data in GC research. We demonstrated in three related studies, the advantages of integrating genomics data in the same cohort of GC samples. The first study integrated CNA with exome sequencing to identify additional genomic deletion in recurrent mutated genes, highlighted the important role of these genes in GC with multiple mechanisms of alterations. The second study utilized copy number landscape with a large panel of gene expression profiling to characterize GC subtypes. The last study reported associations between genome instability with long range epigenetic changes by using both copy number and methylation data.

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Co-amplified (CA)

Confidence Interval (CI)

Comparative Genomic Hybridization (CGH)

Copy Number Alteration (CNA)

Dimension Reduction Permutation (DRP)

Drug Concentration Causing 50% Growth Inhibition (GI50)

False Discovery Rate (FDR)

Fluorescence *in-situ* Hybridization (FISH)

Gastric Cancer (GC)

Gene Expression Omnibus (GEO)

Genomic Identification of Significant Targets in Cancer (GISTIC)

Hazard Ratio (HR)

Immunohistochemistry (IHC)

Inter-Target Relationships (ITRs)

Mutually Exclusive (ME)

National Centre for Biotechnology Information (NCBI)

Quantitative PCR (qPCR)

Receptor Tyrosine Kinase (RTK)

Single Nucleotide Polymorphism (SNP)

Total Growth Inhibition (TGI)

LIST OF PUBLICATIONS

1. A Comprehensive Survey of Genomic Alterations in Gastric Cancer Reveals Systematic Patterns of Molecular Exclusivity and Co-Occurrence among Distinct Therapeutic Targets. **Deng, N***., L. K. Goh*, H. Wang, K. Das, J. Tao, I. B. Tan, S. Zhang, M. Lee, J. Wu, K. H. Lim, Z. Lei, G. Goh, Q. Y. Lim, A. L. Tan, D. Y. Sin Poh, S. Riahi, S. Bell, M. M. Shi, R. Linnartz, F. Zhu, K. G. Yeoh, H. C. Toh, W. P. Yong, H. C. Cheong, S. Y. Rha, A. Boussioutas, H. Grabsch, S. Rozen, and P. Tan. *Gut* 61, no. 5 (2012): 673-84. Epub 2012 Feb 7. (*joint first author)

2. Methylation Subtypes and Large-Scale Epigenetic Alterations in Gastric Cancer. Zouridis, H.*, **N. Deng***, T. Ivanova, Y. Zhu, B. Wong, D. Huang, Y. H. Wu, Y. Wu, I. B. Tan, N. Liem, V. Gopalakrishnan, Q. Luo, J. Wu, M. Lee, W. P. Yong, L. K. Goh, B. T. Teh, S. Rozen, and P. Tan. *Sci Transl Med* 4, no. 156 (2012): 156ra40. (*joint first author)

3. Cd44-Slc1a2 Gene Fusions in Gastric Cancer.
Tao, J., **N. T. Deng**, K. Ramnarayanan, B. Huang, H. K. Oh, S. H. Leong, S. S. Lim, I. B. Tan, C. H. Ooi, J. Wu, M. Lee, S. Zhang, S. Y. Rha, H. C. Chung, D. T. Smoot, H. Ashktorab, O. L. Kon, V. Cacheux, C. Yap, N. Palanisamy, and P. Tan. *Sci Transl Med* 3, no. 77 (2011): 77ra30.

4. Exome Sequencing of Gastric Adenocarcinoma Identifies Recurrent Somatic Mutations in Cell Adhesion and Chromatin Remodeling Genes.
Zang, Z. J., I. Cutcutache, S. L. Poon, S. L. Zhang, J. R. McPherson, J. Tao, V. Rajasegaran, H. L. Heng, **N. Deng**, A. Gan, K. H. Lim, C. K. Ong, D. Huang, S. Y. Chin, I. B. Tan, C. C. Ng, W. Yu, Y. Wu, M. Lee, J. Wu, D. Poh, W. K. Wan, S. Y. Rha, J. So, M. Salto-Tellez, K. G. Yeoh, W. K. Wong, Y. J. Zhu, P. A. Futreal, B. Pang, Y. Ruan, A. M. Hillmer, D. Bertrand, N. Nagarajan, S. Rozen, B. T. Teh, and P. Tan. *Nat Genet* 44, no. 5 (2012): 570-4.

5. Identification of Molecular Subtypes of Gastric Cancer with Different Responses to Pi3-Kinase Inhibitors and 5-Fluorouracil.
Lei, Z., I. B. Tan, K. Das, **N. Deng**, H. Zouridis, S. Pattison, C. Chua, Z. Feng, Y. K. Guan, C. H. Ooi, T. Ivanova, S. Zhang, M. Lee, J. Wu, A. Ngo, S. Manesh, E. Tan, B. T. Teh, J. B. Yan So, L. K. Goh, A. Boussioutas, T. K. Hon Lim, H. Flotow, P. Tan, and S. G. Rozen. *Gastroenterology* (2013). Epub 2013 May 14.

6. Intrinsic Subtypes of Gastric Cancer, Based on Gene Expression Pattern, Predict Survival and Respond Differently to Chemotherapy.
Tan, I. B., T. Ivanova, K. H. Lim, C. W. Ong, **N. Deng**, J. Lee, S. H. Tan, J. Wu, M. H. Lee, C. H. Ooi, S. Y. Rha, W. K. Wong, A. Boussioutas, K. G. Yeoh, J. So, W. P. Yong, A. Tsuburaya, H. Grabsch, H. C. Toh, S. Rozen, J. H. Cheong, S. H. Noh, W. K. Wan, J. A. Ajani, J. S. Lee, M. S. Tellez, and P. Tan. *Gastroenterology* 141, no. 2 (2011): 476-85, 85 e1-11.

7. Genomic Loss of Mir-486 Regulates Tumor Progression and the Olfm4 Antiapoptotic Factor in Gastric Cancer.
Oh, H. K., A. L. Tan, K. Das, C. H. Ooi, **N. T. Deng**, I. B. Tan, E. Beillard, J. Lee, K. Ramnarayanan, S. Y. Rha, N. Palanisamy, P. M. Voorhoeve, and P. Tan. *Clin Cancer Res* 17, no. 9 (2011): 2657-67.

8. Whole-Genome Reconstruction and Mutational Signatures in Gastric Cancer.
Nagarajan, N., D. Bertrand, A. M. Hillmer, Z. J. Zang, F. Yao, P. E. Jacques, A. S. Teo, I. Cutcutache, Z. Zhang, W. H. Lee, Y. Y. Sia, S. Gao, P. N. Ariyaratne, A. Ho, X. Y. Woo, L. Veeravali, C. K. Ong, **N. Deng**, K. V. Desai, C. C. Khor, M. L. Hibberd, A. Shahab, J. Rao, M. Wu, M. Teh, F. Zhu, S. Y. Chin, B. Pang, J. B. So, G. Bourque, R. Soong, W. K. Sung, B. Tean Teh, S. Rozen, X. Ruan, K. G. Yeoh, P. B. Tan, and Y. Ruan. *Genome Biol* 13, no. 12 (2012): R115.

9. Clinical and Therapeutic Relevance of Pim1 Kinase in Gastric Cancer.
Yan, B., E. X. Yau, S. Samanta, C. W. Ong, K. J. Yong, L. K. Ng, B. Bhattacharya, K. H. Lim, R. Soong, K. G. Yeoh, **N. Deng**, P. Tan, Y. Lam, and M. Salto-Tellez. *Gastric Cancer* 15, no. 2 (2012): 188-97.

Chapter 1 Introduction

1.1 Gastric Cancer

1.1.1 Introduction

Gastric adenocarcinoma, or gastric cancer (GC) is the fourth common cancer and second leading cause of global cancer mortality (Ferlay, Shin et al. 2010). The incidence rate of GC varies across different geographical regions, with predominately high incidence rate in East Asia (16% of total cancer incidence), followed by Central Africa (8%), Central East Europe (7%) ,West Asia (7%), and is less frequently observed in the rest of the world (Fig 1-1, (Ferlay, Shin et al. 2010)). In Singapore, it ranked as 4th most common cancer in males and 6th most common cancer in females(Lim, Wong et al. 2009). Like many other solid tumors, the incidence rate of gastric cancer is positively associated with age and the cancer is relatively rare in patients younger than 45 years. Most patients are between 60 and 80 years old at diagnosis. In general, incidence and mortality rates in men are approximately double to those in women (Brenner, Rothenbacher et al. 2009).

Risk factors that can contribute to GC are *Helicobactor pylori* (*H.pylori*) infection, dietary factors (high intake of salt-preserved food and less intake of fruit and vegetables), a family history of GC and smoking(Brenner, Rothenbacher et al. 2009). These factors could lead to potential approaches for GC prevention and early detection by controlling exposure to risk factors and implementing early screening in high risk group. However, classify patients into high or low risk of GC progressing remains to be challenging as there are no established criteria or clinical guidelines for the disease surveillance (Gonzalez and Agudo 2012). Patients identified with GC

usually has adverse prognosis, with an overall five year survival rate of ~20% (Brenner, Rothenbacher et al. 2009; Hartgrink, Jansen et al. 2009). Challenges remain for GC prediction and treatment to reduce the burden of the disease.

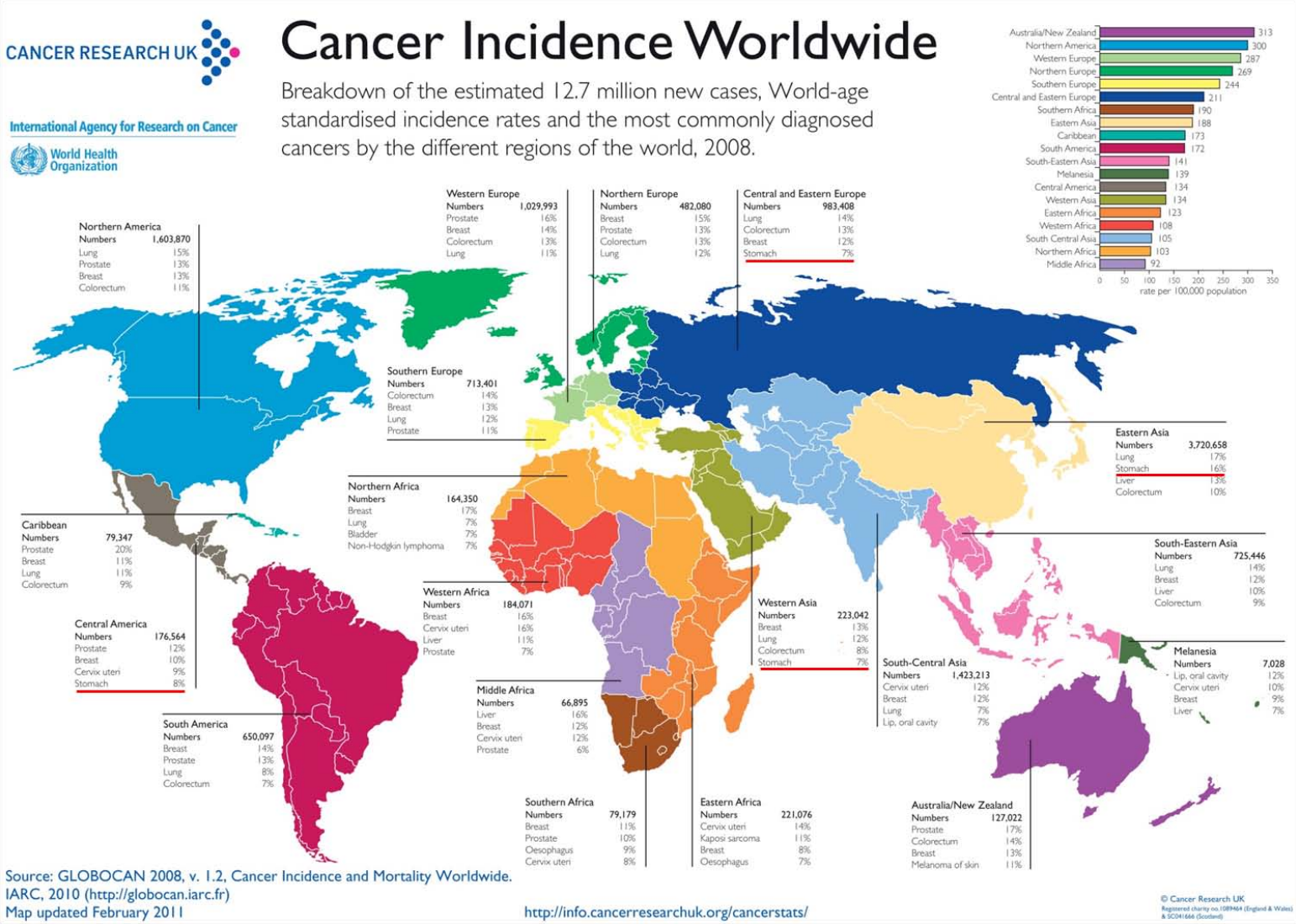


Figure 1-1 Cancer Incidence Worldwide

Most frequent cancers by regions are listed, regions with frequent Gastric Cancer Incidence is underlined in red (updated in Feb 2011). Taken from the Cancer Research UK website: http://publications.cancerresearchuk.org/downloads/Product/worldmap.pdf, reproduced with permission.

1.1.2 Treatment

Current strategies for GC treatment are far from optimal, with conventional surgery and chemotherapy regimens conferring modest survival benefits. Particularly prevalent in many Asian countries (Kamangar, Dores et al. 2006), most GC patients present at advanced disease stages are treated by palliative chemotherapy, with median survival times of 11-12 months (Bang, Van Cutsem et al. 2010).

Surgical removal of part or all of the stomach, as well as regional lymph nodes remains the most common treatment for GC. A few aspects, including the extent of disease, the operative procedure, and patient selection are critical in optimizing patient prognosis. Adjuvant therapy (mainly, chemotherapy ± radiotherapy) still warrants further evaluation for advanced stage GC patients. Neoadjuvant therapy may reduce tumor mass enabling resection with potentially curative intent. For metastatic GC, treatment is exclusively palliative or symptomatic (Catalano, Labianca et al. 2009).

Since GC often reached an advanced stage by the time symptoms occur, where the tumors are unresectable, systematic chemotherapy become the main way of treatment. 5-flurouracil (5-FU) based chemotherapy have been used most frequently for advanced GC treatment after the study of FAM (including 5-flurouracil (5-FU), doxorubicin, and mitomycin C), which reported combination regimen of three drugs for advance GC can improve overall response rate to 42% with median survival time of 5.5 month (MacDonald, Schein et al. 1980). Other chemotherapeutic agents

including cisplatin, epirubicin, capecitabine are usually used in combinations, with an overall response rate between 20% to 60% (Wohrer, Raderer et al. 2004).

In addition to standard cytotoxic regimens, targeted therapies, which are small molecules or antibodies designed to disrupt the activity of specific oncogenic signaling pathways, have recently emerged as a promising therapeutic strategy. In the recent ToGA trial (Bang, Van Cutsem et al. 2010), trastuzumab, an anti-*HER2/ERBB2* targeting antibody, improved the overall survival of patients with *ERBB2*-positive tumors when combined with chemotherapy (Fig 1-2). However, because only 7-17% of GC patients are *ERBB2*-positive (either gene amplification or overexpression) and hence suitable candidates for anti-*ERBB2* therapy (Tanner, Hollmen et al. 2005; Gravalos and Jimeno 2008; Hofmann, Stoss et al. 2008), further research is warranted to increase the population of GC patients for which targeted treatments are clinical options.

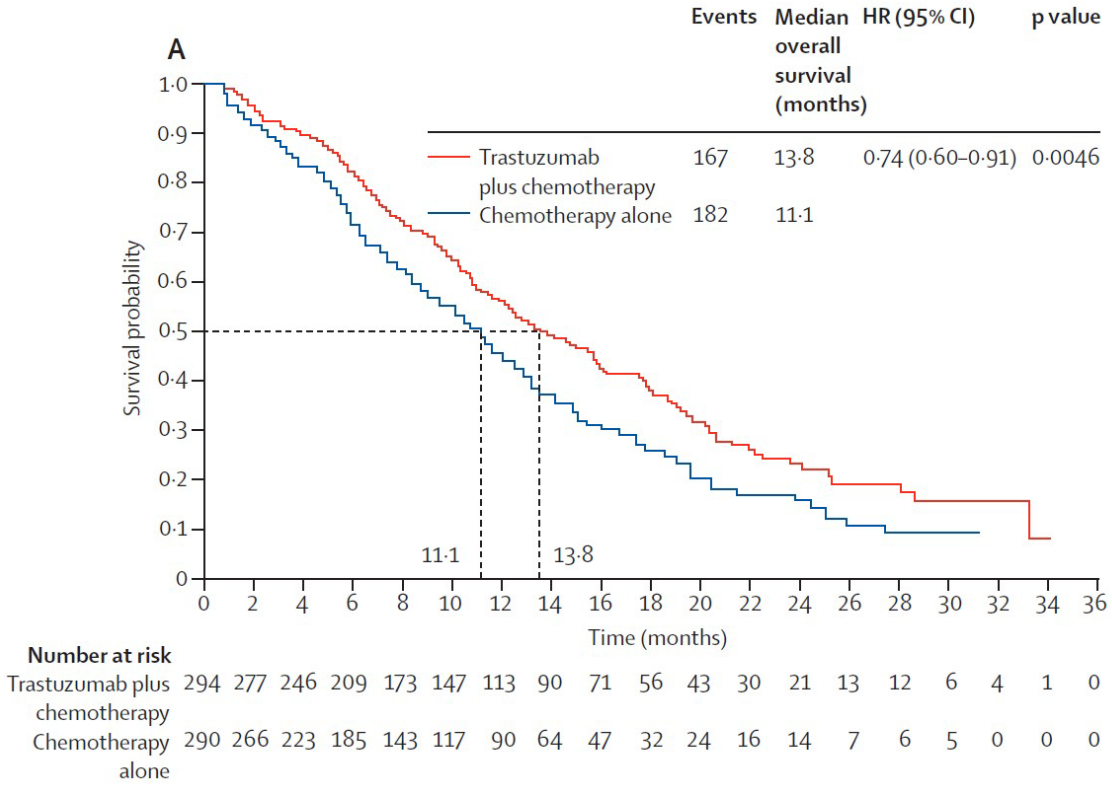


Figure 1-2 ToGA Trial: Overall Survival in the primary analysis population

Median overall survival analysis between trastuzumab treatment (13.8 months) and control group (11.1 months).

Reproduced, with permission from *the Lancet*, Bang, Van Cutsem et al. 2010

1.1.3 Genetic aberrations in Gastric Cancer

Genetic aberrations are underlying hallmarks of cancer. Like many other cancer types, there are multiple mechanisms of genetic aberrations which have been reported in GC, including but not limited to somatic mutations, fusion genes and genomic amplifications or deletions.

Previous studies have reported frequent mutation in *TP53*, with about half of GCs carried a *TP53* mutation(Tamura, Kihana et al. 1991). The tumor protein 53 gene is the most frequently mutated tumor suppressor genes in human cancers and plays many roles in carcinogenesis in response to cellular stresses including apoptosis, cell cycle arrest, senescence, DNA repair, cell metabolism or autophagy by activating specific target genes (Kruse and Gu 2009). With the recent advance in next generation technology, a few studies of GC exome sequencing also reported frequent mutation in *ARID1A* and genes involving chromatin remodeling and cell adhesion. *ARID1A* mutation in GC tends to be associated with microsatellite instability status and exclusive to *TP53* mutation, with an overall percentage of ~20% in all GCs (Wang, Kan et al. 2011; Zang, Cutcutache et al. 2012).

Fusion genes, although a relatively rare kind of genomic rearrangement, can also contribute cancer development and have been reported in other types of cancers. Recently, studies of genomic rearrangement also reported fusion genes of *BRAF* with different partners and *CD44-SLC1A2* fusion in GC, with about 1~5% in all GCs. These fusion genes play a role in cancer development and metabolism, and can provide as potential therapeutic targets (Palanisamy, Ateeq et al. 2010; Tao, Deng et al. 2011).

Studies of genomic instabilities in GC have also reported different scales of genomic structure changes (i.e. amplifications and deletions). For example, cytogenetic studies of karyotypes have identified chromosomal level recurrent genomic amplifications (7q, 8p, and 17q), and deletions (5q, 6p, and 18q) (Kimura, Noguchi et al. 2004). Regions with recurrent aberrations in GC could suggest important genes that of crucial role in gastric carinogenesis, such as *CD44* in recurrent amplifications of 11p13, or *ERBB2* and *TOP2A* in recurrently amplified 17q (Gunther, Schneider-Stock et al. 2000). However, identifying driver oncogenes or tumor suppressor genes in the regions could be challenging due to the large size of the aberration region, usually covering tens to hundreds of genes.

1.2 Copy Number Alterations

1.2.1 Introduction

Definition of Copy Number Alterations

Human genomic variation at single nucleotide level has been observed and since catalogued after the completion of first draft human genome sequence in 2000. Hapmap project, by 2007 has characterized over 3.1 million single nucleotide polymorphisms (SNPs) in a population of 270 individuals (Frazer, Ballinger et al. 2007). Subsequently, numerous genome wide association studies (GWAS) have reported novel SNP association with a range of common disease (Hirschhorn and Daly 2005; 2007; McCarthy, Abecasis et al. 2008). However, detection of large-scale variations and large-scale copy number polymorphisms has shown to some extend that not only SNP but also structural variation in human genome is much greater than previously expected (Iafrate, Feuk et al. 2004; Sebat, Lakshmi et al. 2004).

Structural variations caused by a variable number of copies of a particular DNA segment are referred to as copy number alterations (CNAs). DNA CNAs can be referred to different terms based on their different sizes. CNAs of the entire genetic complements are referred to as polyploidies. CNAs of whole chromosome (34~230 Mbp) are known as aneuploidies. Partial or segmental aneuploidies referred to sub-chromosomal CNAs. Submicroscopic CNAs that are between 1 kbp and 1 Mbp in length are referred to as copy number variations (CNVs). CNAs ranging from 1bp to 1kbp in size are called deletions or insertions depending on whether sequences are deleted or amplified (Feuk, Carson et al. 2006; Lupski 2007; Tang and Amon 2013).

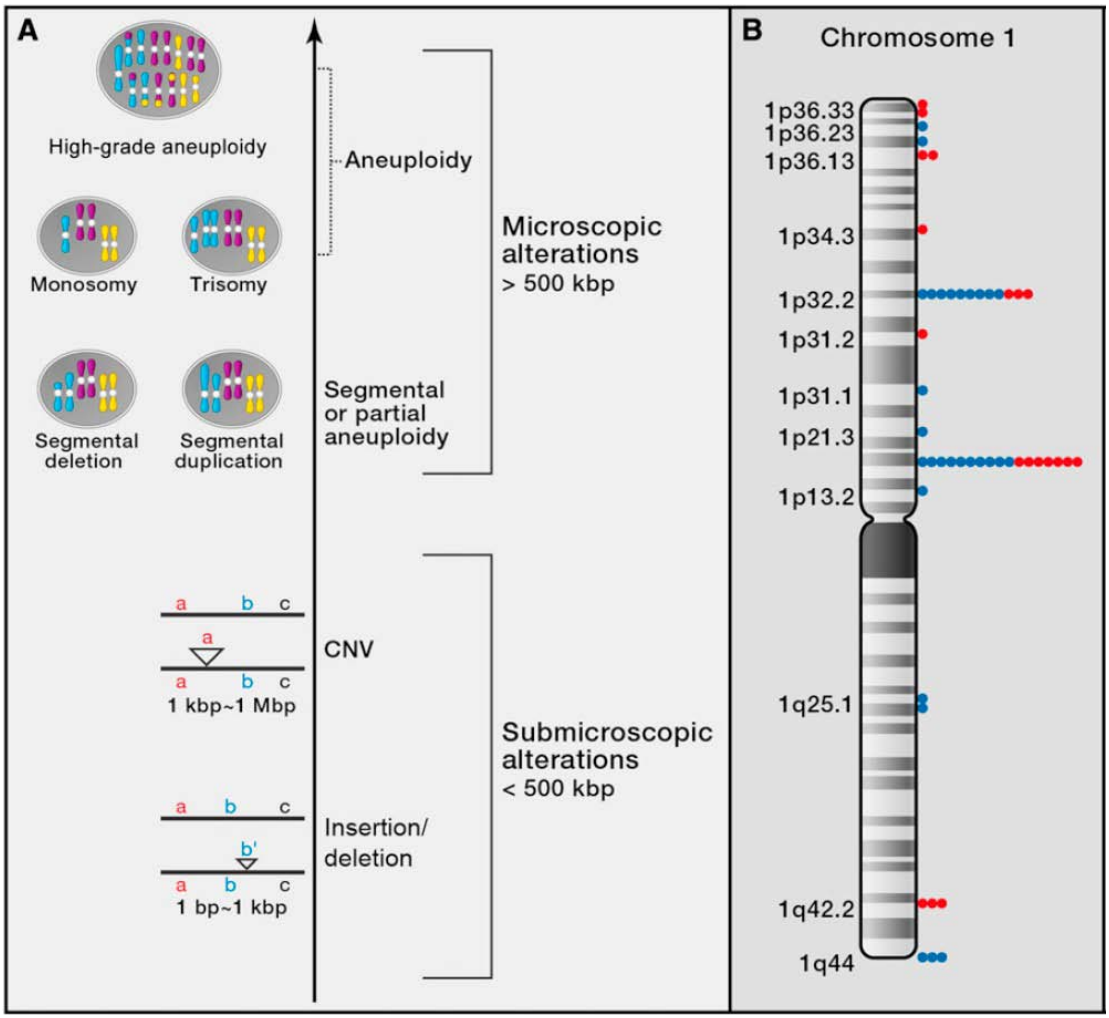


Figure 1-3 Defining Copy Number Alterations (Tang and Amon 2013)

A) DNA copy-number alterations can be categorized into submicroscopic variations, which are smaller than 500kbp, and microscopic alterations, which are greater than 500 kbp. DNA copy number changes between 1bp and 1kbp in size are called insertions or deletions on whether DNA is gained or lost, respectively. Copy number variations (CNVs) vary between 1kbp and 1 Mbp in size. Examples for CNVs are shown for duplication. Microscopically visible karyotype changes are called segmental or partial aneuploidies, when parts or chromosomes are amplified or deleted. whole-chromsome lossed or gains are called aneuploidies. B) CNV distribution on human chromosome 1. Dots show the number of individuals with copy gains (blue) or losses (red) among 39 unrelated, healthy control individuals (Iafrate, Feuk et al. 2004)

Association of Copy Number Alterations with human disease

Genomic CNA at different scales have been long known to be associated with human disease. Down syndrome, for example, is caused by chromosomal level CNA with an extra copy of chromosome 21. Individuals with trisomy 21 exhibit mental retardation and a number of developmental disabilities and are associated with congenital anomalies of the gastrointestinal tract and an increased risk of leukemia (Korenberg, Chen et al. 1994) . Autism spectrum disorder (ASD), which characterized by language impairments, social deficits, and repetitive behaviors, has been found to be associated with copy number duplications or deletions in regions with size from 100bp to 10Mb (Sebat, Lakshmi et al. 2007; Pinto, Pagnamenta et al. 2010). Small-scale CNAs are estimated to be responsible for at least 15% of human neurodevelopmental defects and are associated with psychiatric disorders and kidney and heart defects (Girirajan, Campbell et al. 2011).

Specific CNA associated with particular oncogenes or tumor suppressor genes can contribute to various types of cancer development, such as duplications of epidermal growth factor receptor (*EGFR*) in non small cell lung cancer (Cappuzzo, Hirsch et al. 2005) and deletions of phosphatase and tensin homolog (*PTEN*) in breast cancer (Alimonti, Carracedo et al. 2010). With increasing interests of CNA in human disease, few large-scale genomic studies recently have identified role of CNA in the core pathways of glioblastoma, lung cancer and multiple types of cancers (Weir, Woo et al. 2007; Chaudhuri, Handcock et al. 2008; Beroukhim, Mermel et al. 2010).

1.2.2 Detection of Copy Number Alterations

Numerous cytogenetic techniques can help to identify CNA, such as virtual karyotyping, fluorescent in situ hybridization (FISH), comparative genomic hybridization (CGH), array comparative genomic hybridization (aCGH), and SNP arrays, with the resolution of CNA detection from chromosomal level to a few kbs. Recent advances in DNA sequencing technology have further enabled identification of CNAs to few bases using next-generation sequencing (Korbel, Urban et al. 2007; Mills, Walter et al. 2011) (Fig 1-4).

Fluorescent in situ hybridization (FISH)

FISH is a cytogenetic technique developed by biomedical researchers in the early 1980s that is used to detect and localize the presence or absence of specific DNA sequences on chromosomes (Langer-Safer, Levine et al. 1982), and now has been widely applied to the detection of specific normal and aberrant DNA sequences, as well as been successfully applied in a number of diseases diagnosis, such as pediatric neoplastic diseases (Raimondi 2000).

Comparative genomic hybridization (CGH) or array comparative genomic hybridization (aCGH)

CGH is a molecular cytogenetic method for analyzing CNA in the DNA of a test sample compared to a reference sample. This technique was originally developed to evaluate the changes between the chromosomal complements of solid tumor and matched normal tissue (Kallioniemi, Kallioniemi et al. 1992). Together through the

use of microarrays in conjunction with CGH techniques, also referred as aCGH, can identify CNA at a higher resolution as low as 200bp (Urban, Korbel et al. 2006).

Single nucleotide polymorphism (SNP) array

SNP array is a type of DNA microarray which is used to detect polymorphisms within a population. It can be used to detect SNPs, loss of heterozygosity, and genomic amplification or deletions (LaFramboise 2009). High density SNP array can help to identify patterns of allelic imbalance and small scale CNAs, and has been used in Hapmap project and a few other large-scale genomic studies (Frazer, Ballinger et al. 2007; Weir, Woo et al. 2007; 2008; Beroukhim, Mermel et al. 2010).

Affymetrix SNP 6.0 array

The Genome-Wide Human SNP Array 6.0 features more than 1.8 million markers of genetic variation, including SNPs, as well as probes for the detection of copy number variation, with the median inter-marker distance over all 1.8 million SNP and copy number markers combined is less than 700 bases (McCarroll, Kuruvilla et al. 2008) (Fig 1-5).

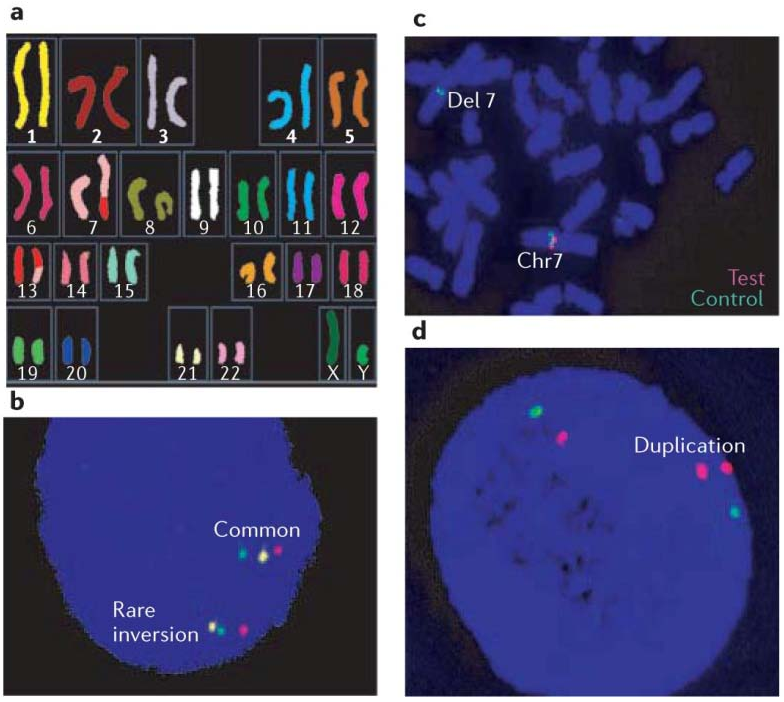


Figure 1-4 Cytogenetic detection and confirmation of structural variants

a. Spectral karyotyping is ideal for the identification of rearrangements that involve the exchange of DNA between chromosomes. Differentially labelled DNA probes for all chromosomes are used, making it possible to identify every chromosome in a single hybridization. The example shows the detection of a t(7;13) translocation.**b.** Three-colour FISH that was carried out using interphase nuclei shows that a 700-kb micro-inversion at 7p22 is olymorphic, as indicated by the change in order of BAC clones (which are labelled in different colours) between the two copies of the same chromosome that are present in the nucleus. **c.** the green control probe is present in two copies on chromosome 7 (chr7), whereas the red probe shows a signal on only one of the homologous chromosome 7 copies. **d.** Two-colour FISH reveals a large-scale copy-number variant, in this case a duplication. (Feuk, Carson et al. 2006)

Reproduced, with permission from *Nature Reviews Genetics*, Fuek, Carson et al 2006

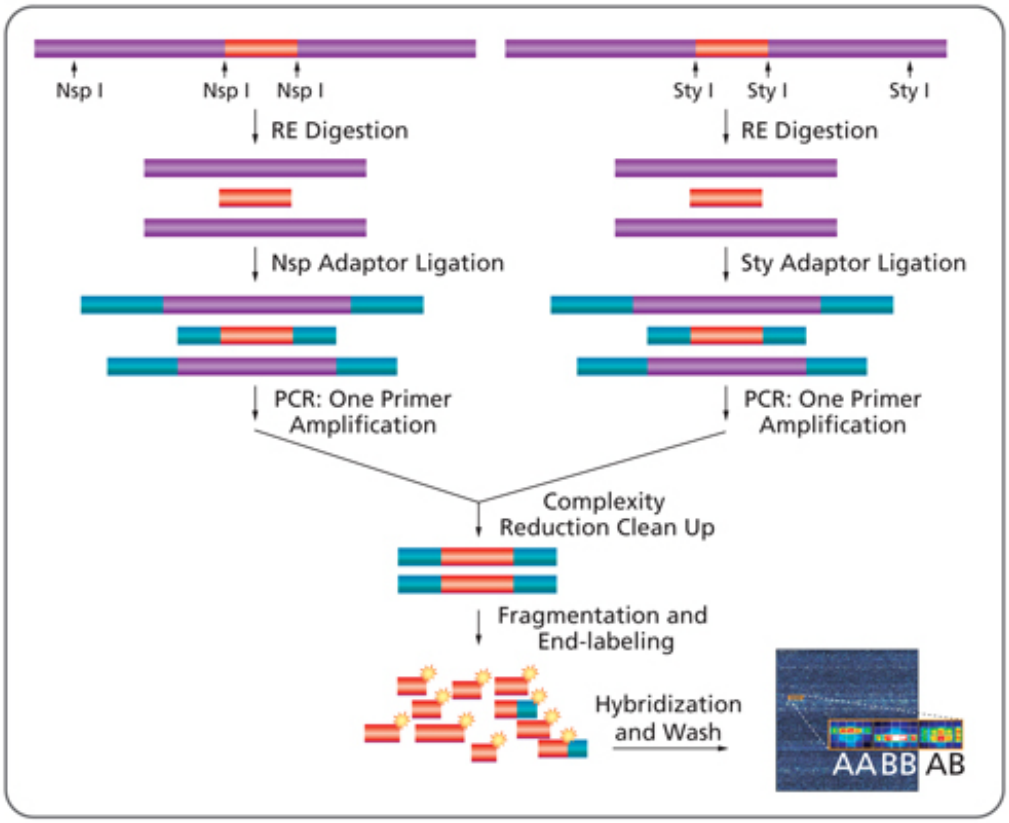


Figure 1-5 Overview of the Genome-Wide Human SNP Assay 6.0

Total genomic DNA (500 ng) is digested with Nsp I and Sty I restriction enzymes and ligated to adaptors that recognize the cohesive 4 bp overhangs. All fragments resulting from restriction enzyme digestion, regardless of size, are substrates for adaptor ligation. A generic primer that recognizes the adaptor sequence is used to amplify adaptor-ligated DNA fragments. PCR conditions have been optimized to preferentially amplify fragments in the 200 to 1,100 bp size range. PCR amplification products for each restriction enzyme digest are combined and purified using polystyrene beads. The amplified DNA is then fragmented, labeled, and hybridized to a SNP Array 6.0.
(from http://www.affymetrix.com affymetrix snp6.0 datasheet)

1.2.3 Previous studies of CNA in GC

Currently there are a few studies on copy number in GC: *CD44* has been reported to be recurrently amplified in 11p region using CGH in a panel of 25 GC cell lines(Fukuda, Kurihara et al. 2000); amplification of 19q12, which covered gene *CCNE1* was studied in a cohort of 126 gastric primary samples with CGH at a ~1.5Mb resolution (Leung, Ho et al. 2006); (Kimura, Noguchi et al. 2004) reported a study of genetic alterations in 102 Gastric Tumors using chromosomal level CGH, and identified recurrent gain of 20q and loss of 18q were associated with tumor progression; (Myllykangas, Junnila et al. 2008; Tsukamoto, Uchida et al. 2008; Nakamura, Migita et al. 2009) also profiled genetic copy number using aCGH in about 30 ~ 50 gastric tumor samples and reported recurrent amplifications in 17q and 20q; (Tada, Kanai et al. 2010) applied SNP array with about 50K probes in 30 ~ 40 cell lines and gastric primary tumors, suggesting a list of candidate genes, including *MYC* and a few other genes.

In general, these studies have largely suffered from various weaknesses such as relatively small patient populations, focus on in vitro cell lines, and use of low-resolution technologies (e.g. chromosomal CGH) (summary in Table 1.1 and Table 1.2).

Table 1-1 List of previous copy number studies in Gastric Cancer

Study	Sample Size	platform
Tada M et al, 2010	34CL + 42T	Affymetrix 50k SNP array
Tsukamoto Y et al, 2008	30T	Array CGH
Nakamura Y et al, 2009	50T	Array CGH
Myllykangas S et al, 2008	38T + 8N	Array CGH
Kimura Y et al, 2004	102T	CGH

Table 1-2 Identified chromosomal level copy number alterations of Gastric Cancer in literatures

Amplification	Deletion	Reference
7q, 8q, and 20q	3p, 4q, 9p, and 18q	Tada M et al, 2010, Cancer Science
1q, 3q, 5p, 6p, 7p, 8q, 13q, 17q, 19q, 20q, 20p,	3p, 4p, 4q, 5q, 9p, 10q, 12q, 16q, 17p, 18q, 21q	Tsukamoto Y et al, 2008, J Pathology
5p, 7q, 8q, 13q, 17q, 20p, 20q	4p, 4q, 5q, 15q, and 17p, 18q, 19p, 21q	Kimura Y et al, 2004, Mod Pathology

1.3 Rationale of this study

Reflecting this urgency to identify more therapeutic targets and increase GC population which can benefit from targeted treatment, several other targeted therapies are currently undergoing preclinical and clinical testing in GC, directed against diverse oncogenic proteins including signaling receptors, histone deacetylases and cellular proteins (Weichert, Roske et al. 2008; Moser, Lang et al. 2009; Yap, Olmos et al. 2011). However, because most of these targeted therapies were originally designed against proteins expressed or discovered in other cancers (eg trastuzumab for breast cancer), in many cases surprisingly little is actually known either regarding the true prevalence of their oncogenic targets in primary GCs, or if expression of these oncogenic targets is correlated with key clinico-pathologic parameters such as patient outcome. As one example, the *FGFR2* receptor tyrosine kinase (RTK) has been previously proposed as a potential therapeutic target in GC (Turner and Grose 2010). However, most *FGFR2*-related studies in GC have been primarily restricted to *in vitro* cultured cell lines (Takeda, Arao et al. 2007; Kunii, Davis et al. 2008), and little data is available regarding the true prevalence of *FGFR2* gene amplification in primary GCs particularly at the high-resolution genomic level. As such, a comprehensive and unbiased survey to identify the most prevalent molecular targets in GC could facilitate many aspects of GC translational research, for example in focusing clinical trials efforts on those therapies that might benefit the greatest numbers of GC patients.

Besides identifying the most prevalent targets, recent findings have also highlighted the importance of determining if certain combinations of targets are expressed either

independently from one another (i.e. mutual exclusivity), or co-occurring in the same tumor. Knowledge of such "inter-target relationships" (ITRs) can shed critical insights into the signaling networks of a cancer cell, case examples being the mutual exclusivity of *KRAS* and *BRAF* activating mutations in colorectal cancer, and the exclusivity of *EGFR* and *KRAS* mutations in lung cancer (Rajagopalan, Bardelli et al. 2002; Ding, Getz et al. 2008). Identifying ITRs may also highlight promising drug combinations for combination therapy, and suggest rational molecular criteria for patient inclusion and exclusion in clinical trials. Recent studies exemplifying both the basic and clinical importance of ITRs include *ERBB2* and *PIK3CA*, where co-occurring *PIK3CA* mutations in *ERBB2*-positive breast cancers can modulate clinical responses to trastuzumab (Berns, Horlings et al. 2007), and *EGFR* and *MET*, where clinical resistance to gefitinib in *EGFR*-mutated lung cancers can be caused by co-existing *MET* gene amplifications (Bean, Brennan et al. 2007).

In this study, we sought to identify the most prevalent molecular targets in GC and to elucidate systematic patterns of exclusivity and co-occurrence among these targets, through comprehensive genomic analysis of a large panel of GCs.

Chapter 2 Material and Methods

2.1 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. "Normal" (i.e. non-malignant) samples used in this study refer to samples harvested from the stomach, from sites distant from the tumor and exhibiting no visible evidence of tumor or intestinal metaplasia/dysplasia upon surgical assessment. 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. Clinicopathological information of these patients including age, disease stage, histological subtype, treatment and anatomical location, are included in Table 2-1. Only three patients received neo-adjuvant or pre-operative chemotherapy prior to surgery. 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).

Table 2-1 Clinical Characteristic 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 neo-adjuvant therapy, and of 131 patients where subsequent treatment information was available, 28 patients received 5-FU chemo-radiation 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.

2.2 DNA and RNA Extraction

Genomic DNA was extracted from flash-frozen tissues and cells using a Qiagen genomic DNA extraction kit and profiled on Affymetrix SNP 6.0 arrays according to the manufacturer's specifications.

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).

2.3 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.

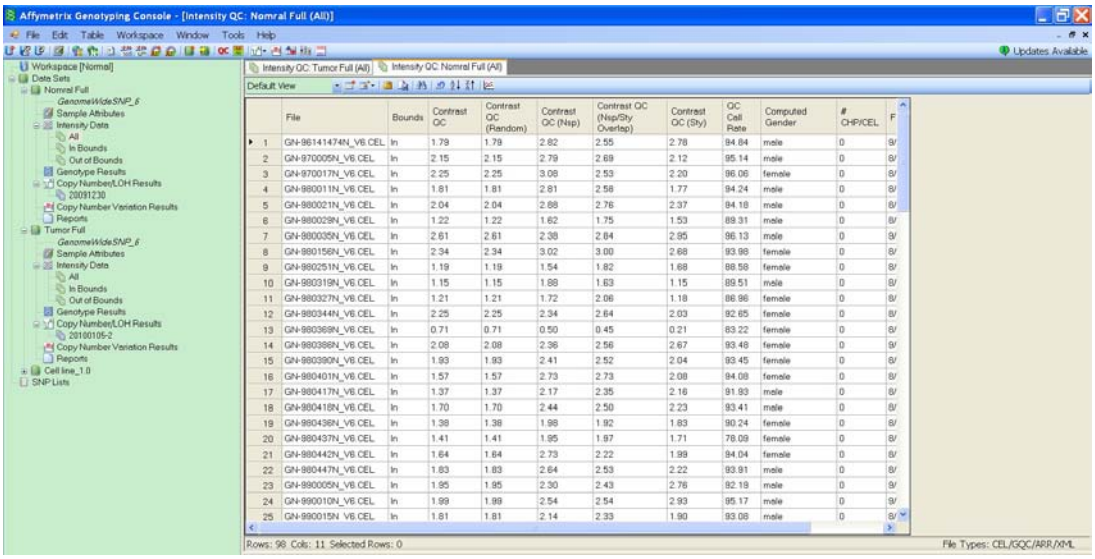


Figure 2-1 Illustration of Affymetrix GTC working panel for data preprocessing

Step 2) Segmentation: Copy number segmentation data was produced using the Circular Binary Segmentation (CBS) algorithm using the R package *DNAcopy* (Olshen, Venkatraman et al. 2004) 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.

A brief description of CBS:

Suppose c = change point if $X_1,\ldots X_c$ *has distribution F and* $X_{c+1},\ldots,$ *has distribution G,*

X_j *is log-ratio intensities, indexed by marker location,*

$S_i = X_1+\ldots+X_i$, $i= 1,\ldots,n$ *is the partial sum,*

$Z_i = [1/i+1/(n-i)]^{-1} [S_i/i – (S_n - S_i)/(n-i)]$,

CBS is based on the likelihood ratio test statistic for:

H_0*: no change point vs* H_1*: exactly one change point at an unknown location i*

$Z_B = max_{1<=i<n}|Z_i|$

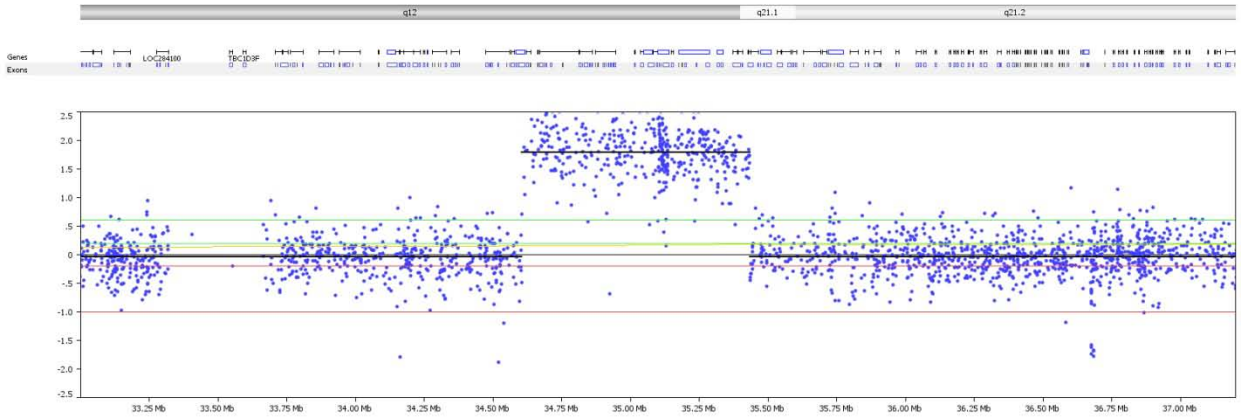


Figure 2-2 An example of break points identified by Circular Binary Segmentation algorithm.

Each dot represent a Affymetrix SNP6 probe, x-axis refers to genomic coordinates, y-axis indicates log-ratio copy number data. Black horizontal line specify boundary of segmented data, indicating the average of log-ratio for probes contained in the segmented region.

3) GISTIC Analysis: The GISTIC algorithm (Beroukhim, Getz et al. 2007) 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 false discovery q-value, which is obtained after 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. Tumor-specific genomic alterations were identified by normalizing the primary GC profiles against

the primary matched gastric normal samples.

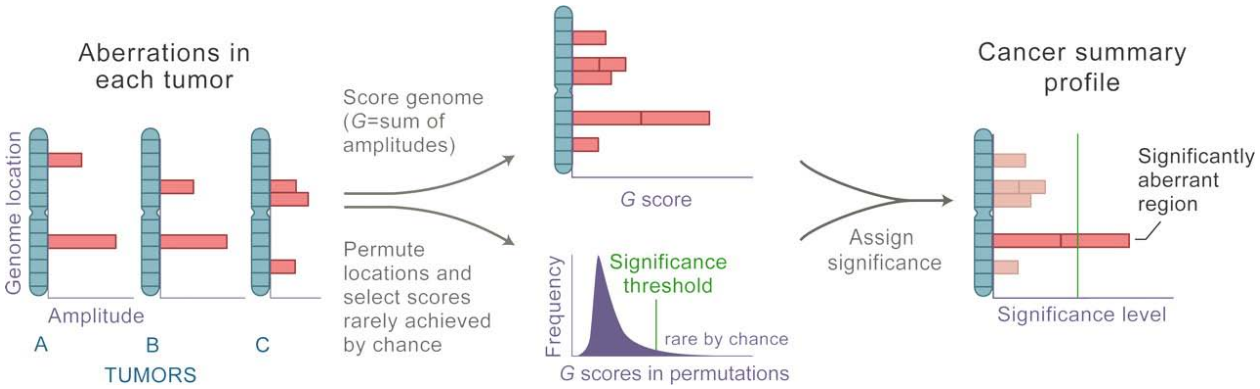


Figure 2-3 Overview of GISTIC method

After identifying the locations and, in the case of copy-number alterations, magnitudes (as log2 signal intensity ratios) of chromosomal aberrations in multiple tumors (Left), GISTIC scores each genomic marker with a G-score that is proportional to the total magnitude of aberrations at each location (Upper Center). In addition, by permuting the locations in each tumor, GISTIC determines the frequency with which a given score would be attained if the events were due to chance and therefore randomly distributed (Lower Center). A significance threshold (green line) is determined such that significant scores are unlikely to occur by chance alone. Alterations are deemed significant if they occur in regions that surpass this threshold (Right). Reproduced, with permission from (Beroukhim, Getz et al. 2007)

The SNP6 copy number data has been deposited into the National Centre for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) website, with series accession number GSE31168.

2.4 DRP: Identification of Mutually Exclusive and Co-Altered CNAs

To identify significant relationships between regions of frequent CNAs, we implemented a dimension reduction permutation (DRP) statistical algorithm adapted from a previous study analyzing patterns of somatic DNA mutations in tumor (Ding, Getz et al. 2008). 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 below, and the DRP script is attached in appendix.

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 (Ding, Getz et al. 2008). 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.

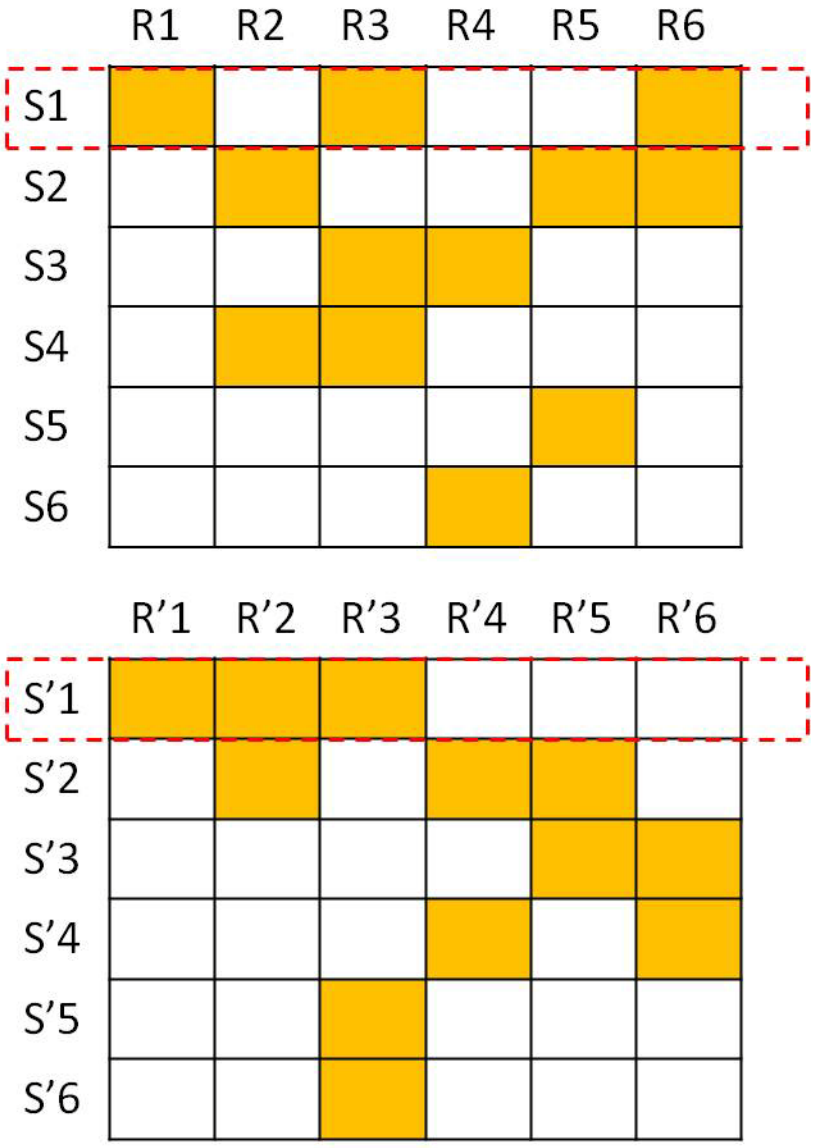


Figure 2-4 Illustration of Permutations by DRP

(above) This figure illustrates one situation of CNA events in multiple samples, rows indicate samples, columns indicate CNA regions, and presence of a CNA event in a particular sample is highlighted in orange. (below) This figure illustrates one example of a possible permutation of CNA events by DRP. Note that numbers of CNA events by samples or by regions are fixed.

2.5 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 (Hofmann, Stoss et al. 2008). *CSMD1* immunohistochemistry was performed on full sections as described in (Kamal, Shaaban et al. 2010). 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).

2.6 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.

2.7 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 3-15). Additional details of the gene expression data set are presented in (Ooi, Ivanova et al. 2009) and are publicly available at GEO under accession number GSE15460. To analyze *FGFR2* mRNA survival associations in Figure 3-19, we analyzed a combined GC gene expression data set of 398 tumors. The 156 patients analyzed in Figure 3-15 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 3-8. Briefly, individual arrays were normalized using the MAS5 algorithm, and batch effects removed using the COMBAT algorithm (Johnson, Li et al. 2007).

2.8 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 $p<0.05$.

2.9 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.

2.10 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.

2.11 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.

2.12 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.

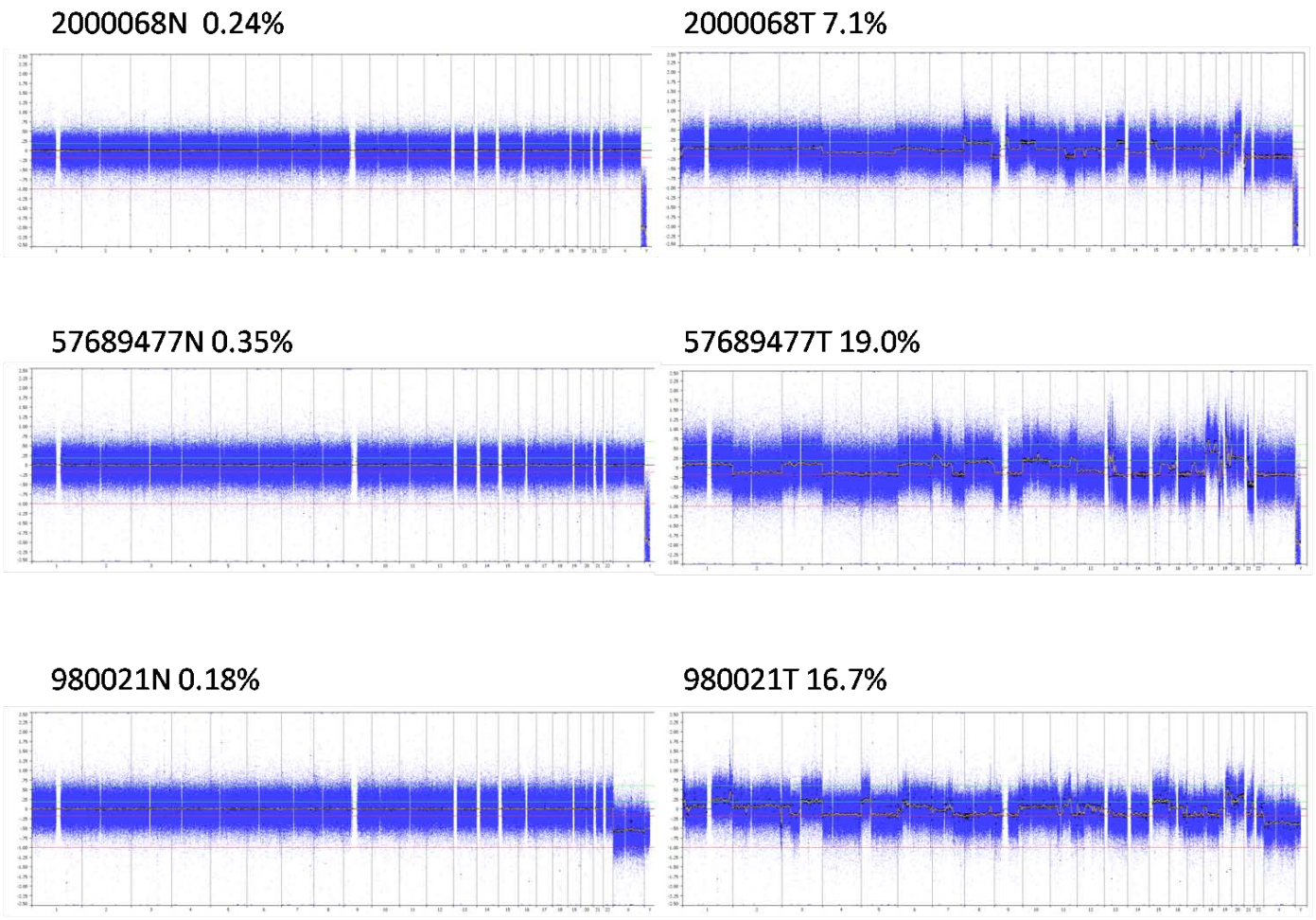
Chapter 3 Results

3.1 Genomic Landscape of CNAs in GC

We profiled genomic DNA samples from 193 primary GCs, 98 primary matched gastric normal samples, and 40 GC cell lines on Affymetrix SNP6 microarrays containing ~1.8 million probes with a median inter-probe spacing of 680bp. To identify tumor-specific genomic alterations and exclude regions of potential germ line copy number variation, we normalized the GC profiles against the matched gastric normal samples (Figure 3-1 for representative profiles). On average, we observed ~150 genomic aberrations per GC, comprising a mixture of broad and focally altered regions. Frequently amplified broad chromosomal regions included 1q, 3q, 5p, 6p, 7pq, 8q, 12pq, 13q, 18pq, 19p, 20pq and 21p (frequencies 9.8% to 33.7%), and frequently deleted chromosomal regions included 3p, 4pq, 5q, 6q, 8p, 9p, 9q, 11q, 12p, 14q, 16q, 17p, 18p, 18q, 19p 21q and 22q (frequencies 7.8% to 13.0%) (Figure 3-2). These results are highly concordant with previous comparative genomic hybridization (CGH/aCGH) studies of GC (Peng, Sugihara et al. 2003; Tay, Leong et al. 2003; Kimura, Noguchi et al. 2004; Tsukamoto, Uchida et al. 2008; Tada, Kanai et al. 2010; Rossi, Klersy et al. 2011).

Figure 3-1 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.



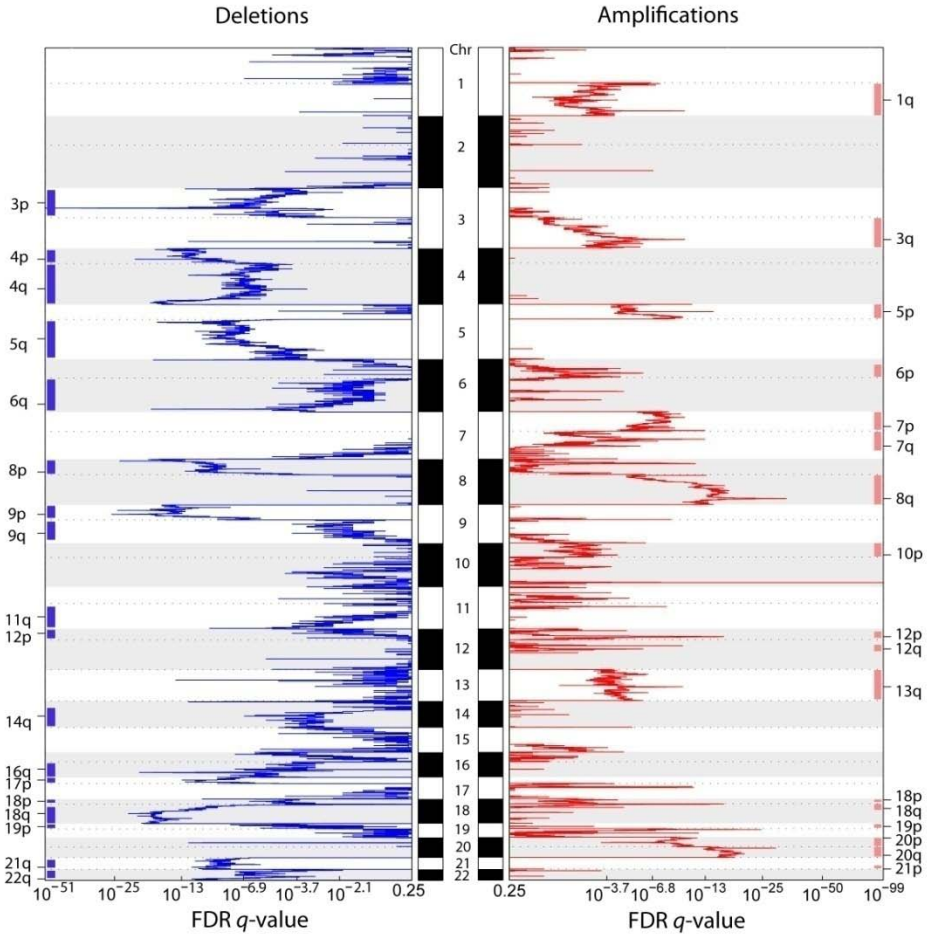


Figure 3-2 Broad Genomic Alterations in GC

Large-scale copy number alterations (CNA). The diagram shows a CNA plot where chromosomal regions of the 22 autosomes are represented on the y-axis, and GISTIC computed FDR q- values are on the x-axis. Chromosomal deletions are on the left (blue) and amplifications are on the right (red). Significantly altered regions of broad CNA are highlighted at the sides, as blue and red bars (GISTIC q value<0.25).

3.2 Focal Genomic Alterations Highlight 22 Potential Targets in GC

We identified 22 focal genomic alterations, defined as narrow regions (typically <100 kb) exhibiting high levels of copy number gain or loss (Table 3-1). Among the amplified genes were several oncogenes previously known to be amplified in GC including *EGFR*, *ERBB2/HER2* and *CCND1* (Figure 3-3) (Hirono, Tsugawa et al. 1995; Tanner, Hollmen et al. 2005; Bizari, Borim et al. 2006). Among the focally deleted genes in GC, we re-identified *FHIT*, *RB1*, *CDKN2A/B*, and *WWOX*, also previously known to be deleted in GC (Schneider, Pulitzer et al. 1995; He, Su et al. 2001; Lee, Leung et al. 2002; Aqeilan, Kuroki et al. 2004; Xiao, Wu et al. 2006). The re-discovery of these classical oncogenes and tumor suppressor genes supports the accuracy of the SNP6 array data. To further validate the array data, we performed *ERBB2* immunohistochemistry (IHC) on 146 of the 193 cases (Figure 3-4), and confirmed a significant association between *ERBB2* copy number gain and *ERBB2* protein expression ($p<0.01$, Fisher's exact test, Table 3-2).

Besides known genes, the analysis also revealed novel genes not previously reported in GC. These included genomic amplification of the transcription factors *GATA6* and *KLF5*, and somatic deletions in *PARK2*, *PDE4D*, *CSMD1* and *GMDS*. Recent data suggests that GATA factors in particular may play an oncogenic role in certain gastrointestinal cancers, for example *GATA6* has been shown to be amplified in pancreatic cancer (Kwei, Bashyam et al. 2008). *PARK2* and *PDE4D* deletions have also been recently observed in glioblastoma and lung adenocarcinomas (Weir, Woo et al. 2007; Chaudhuri, Handcock et al. 2008). Using IHC, we confirmed that one of these novel deleted genes, *CSMD1*, was downregulated or absent in ~40% of primary

GCs at the protein level, but highly expressed in normal gastric epithelium (n=42; Figure A1).

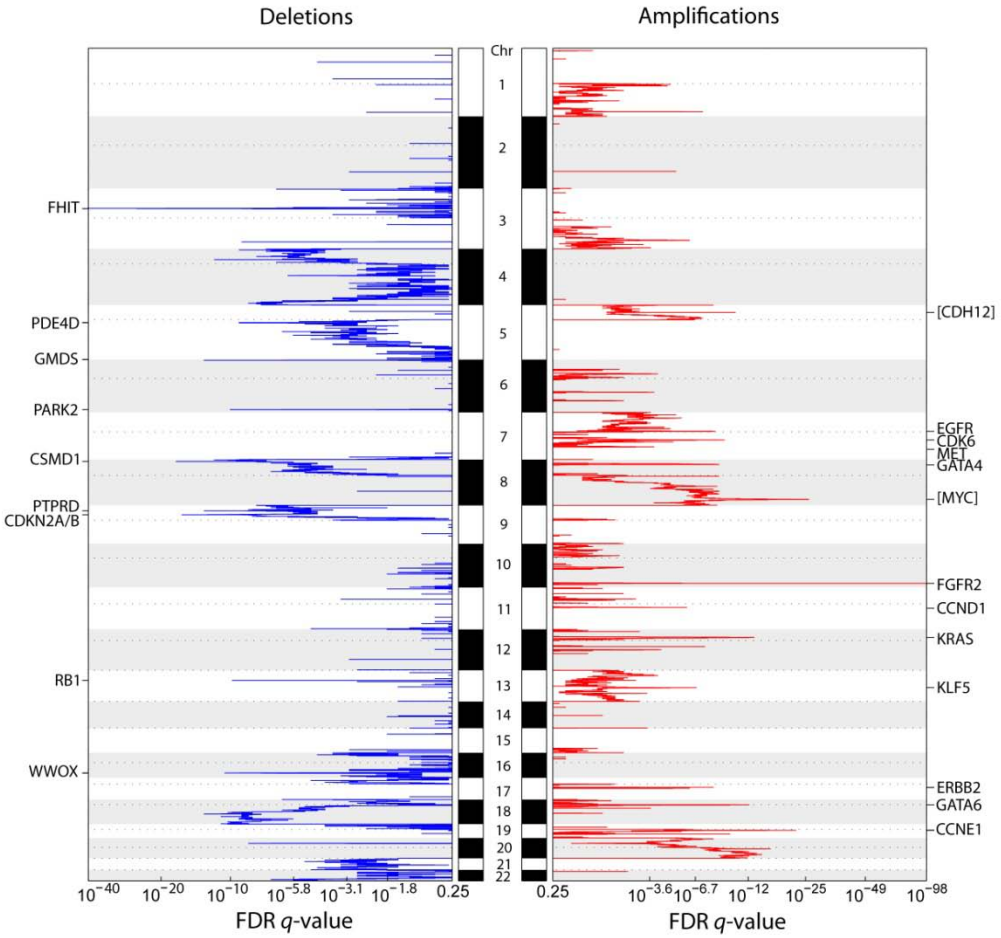


Figure 3-3 Focal Genomic Alterations in GC

Focal alterations. Genes localized within the peaks of the focally altered regions are specified. Genes in square brackets are genes that lie immediately adjacent to the alteration peak (e.g. *MYC*). Significantly altered focal events (GISTIC q-value <0.001) are highlighted at the sides and summarized in Table 3-1.

Table 3-1 Focal regions of CNA regions in GC

Focal recurrent CNAs (amplifications and deletions) identified by GISTIC. Genes previously reported as oncogenes or tumor suppressor genes are highlighted in bold. Start and End indicates the boundary of the region identified. Length indicates size of each region identified. Q value represents the significance of the recurrent CNA region across all the gastric tumors. Genes in peak - genes covered by the corresponding region, a square bracket indicates that the gene lies immediately adjacent to the peak.

CNAs	Chr	Start	End	Length (kb)	Cytoband	Q value	Genes in peak
Amplification							
1	10	123336181	123337713	1.5	10q26.13	3.9561E-99	***FGFR2***
2	8	128628340	128670251	41.9	8q24.21	7.984E-27	***[MYC]***
3	19	34982652	35002397	19.7	19q12	3.1439E-23	***CCNE1***
4	12	25213920	25336398	122.5	12p12.1	1.5713E-14	***KRAS****,CASC1,LYRM5*
5	18	17947474	18040783	93.3	18q11.2	1.0616E-13	*GATA6*
6	5	21377838	21406308	28.5	5p14.3	9.501E-12	***[CDH12]***
7	7	91921079	92111471	190.4	7q21.2	2.0612E-10	***CDK6****,PEX1,GATAD1,DKFZP564O0523,FAM133B*
8	8	11346688	11659701	313.0	8p23.1	9.0544E-10	*BLK,GATA4,C8orf13*
9	7	55237447	55373693	136.2	7p11.2	2.4109E-09	***EGFR***
10	17	35102118	35136335	34.2	17q12	3.8268E-09	***ERBB2***
11	13	72528937	72770614	241.7	13q22.1	1.4729E-07	*KLF5*

12	11	69161019	69306967	145.9	11q13.2	9.1737E-07	***CCND1***,*FGF4,FGF19,ORAOV1*
13	7	115987034	116178774	191.7	7q31.2	0.00012527	*CAV1*,***MET***
Deletion							
1	3	60447451	60472964	25.5	3p14.2	3.4002E-41	***FHIT***
2	8	4182635	4182916	0.3	8p23.2	1.0797E-18	*CSMD1*
3	9	21953419	21995192	41.8	9p21.3	1.0299E-17	***CDKN2A,CDKN2B***
4	6	2019538	2068880	49.3	6p25.3	1.7756E-14	*GMDS*
5	16	77269209	77293232	24.0	16q23.1	5.4871E-12	*WWOX*
6	6	162551244	162610874	59.6	6q26	2.1056E-11	*PARK2*
7	13	47806677	47809375	2.7	13q14.2	3.3682E-11	***RB1***
8	5	58436441	58569237	132.8	5q11.2	1.6661E-10	*PDE4D*
9	9	9524063	9675303	151.2	9p23	1.2287E-09	*PTPRD*

Figure 3-4 *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.

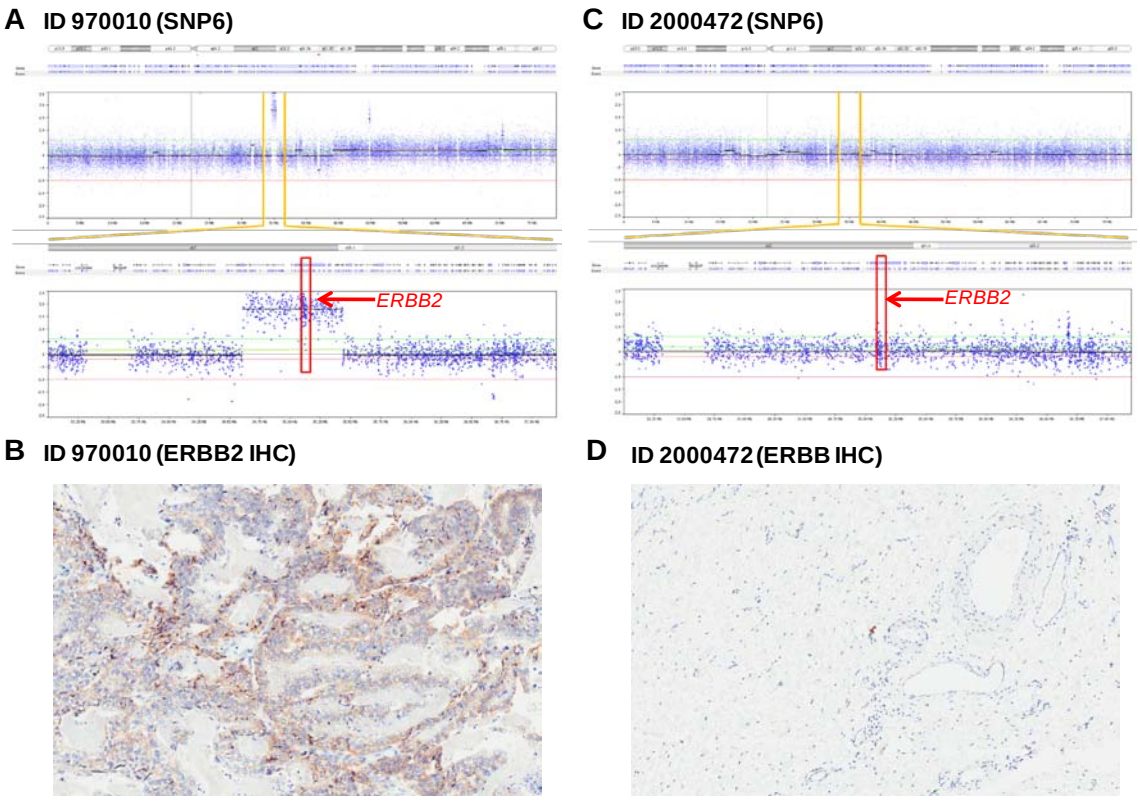


Table 3-2 Concordance Table between *ERBB2* SNP6 and *ERBB2* IHC

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

3.3 A Network of Non-Random ITRs Define Relationships between GC Targets

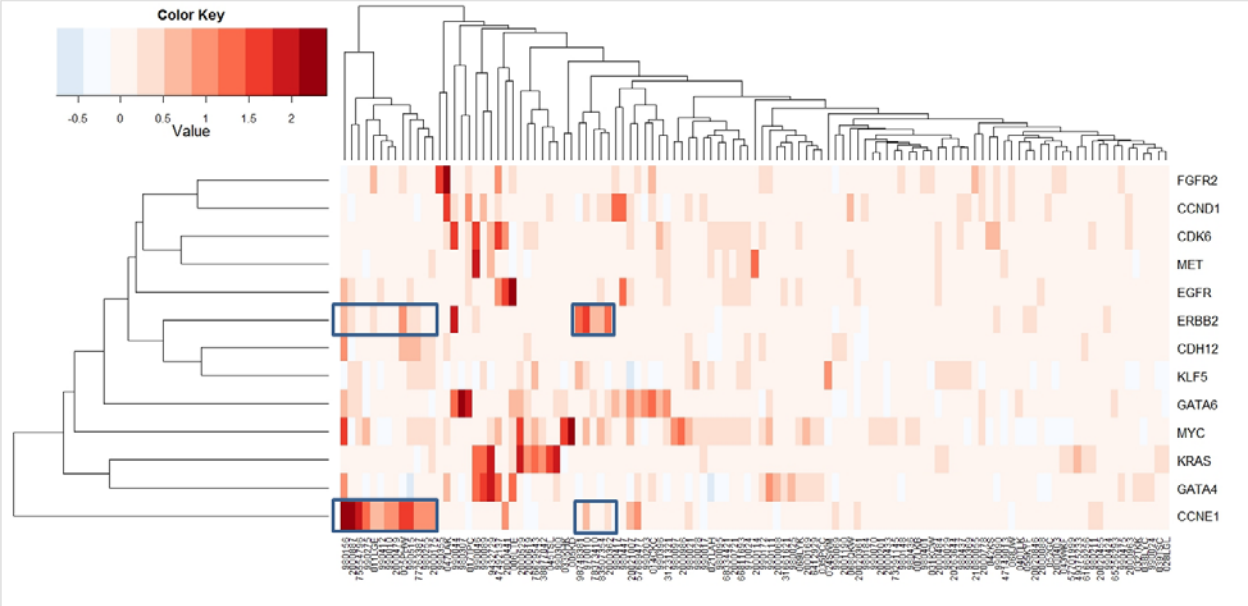
A major goal of our study was to identify non-coincidental ITRs between the 22 GC targets in a systematic, unbiased, and statistically rigorous manner. We developed a statistical method called Dimension Reduction Permutation (DRP) for this purpose (see Methods section 2.4). Briefly, DRP identifies non-random ITRs between targets by comparing the numbers of tumor samples exhibiting a particular ITR (associations between distinct alterations) against a null distribution of background ITRs generated through random permutation. Compared to other methods such as hierarchical clustering and correlation tests, DRP provides additional sensitivity in identifying ITRs, without requiring *a priori* knowledge of specific gene functions (Figure 3-5).

We uncovered several significant ITRs associated with the 22 GC targets. These target pairs were either amplified in a mutually exclusive (ME) manner in different tumors, or co-amplified (CA) in the same tumor (Figure 3-3 and Table A1). Functionally, the GC ITRs tended to involve two specific target classes– a) genes related to RTK/RAS signaling, including *KRAS*, *FGFR2*, *ERBB2*, *EGFR*, and *MET*, and b) genes related to transcription factor biology (*MYC*, *GATA4*, *GATA6*, and *KLF5*). For example, tumors exhibiting *KRAS* amplifications were largely distinct from tumors exhibiting *ERBB2* or *FGFR2* amplification (p=0.02 and p=0.005 for *KRAS/ERBB2* and *KRAS/FGFR2* respectively), while tumors exhibiting *MET* amplifications were distinct from tumors with *FGFR2* amplifications (p=0.03; Figure 3-6 and Table A1). Likewise, *GATA4*, *GATA6* and *KLF5* were significantly co-amplified with *MYC* (*KLF5*: p=0.0005; *GATA4*: p=0.008; *GATA6*: p=0.01), while *KLF5* and *GATA4* amplifications were mutually exclusive to one another (p=0.01).

Other notable ITRs included a significant coamplification interaction between *EGFR* and *MYC* ($p = 0.002$), and between *ERBB2* and *CCNE1* ($p=0.05$) (Figure 3-7), a co-amplification pattern recently linked to trastuzumab resistance in breast cancer (Scaltriti, Eichhorn et al. 2011). Taken collectively, these results support the existence of a complex functional network of ITRs in GC. They provide evidence that as opposed to each target behaving independently from one another, the presence of one target in a GC is likely to exert a profound influence on the repertoire of other targets expressed in that same tumor.

Figure 3-5 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.



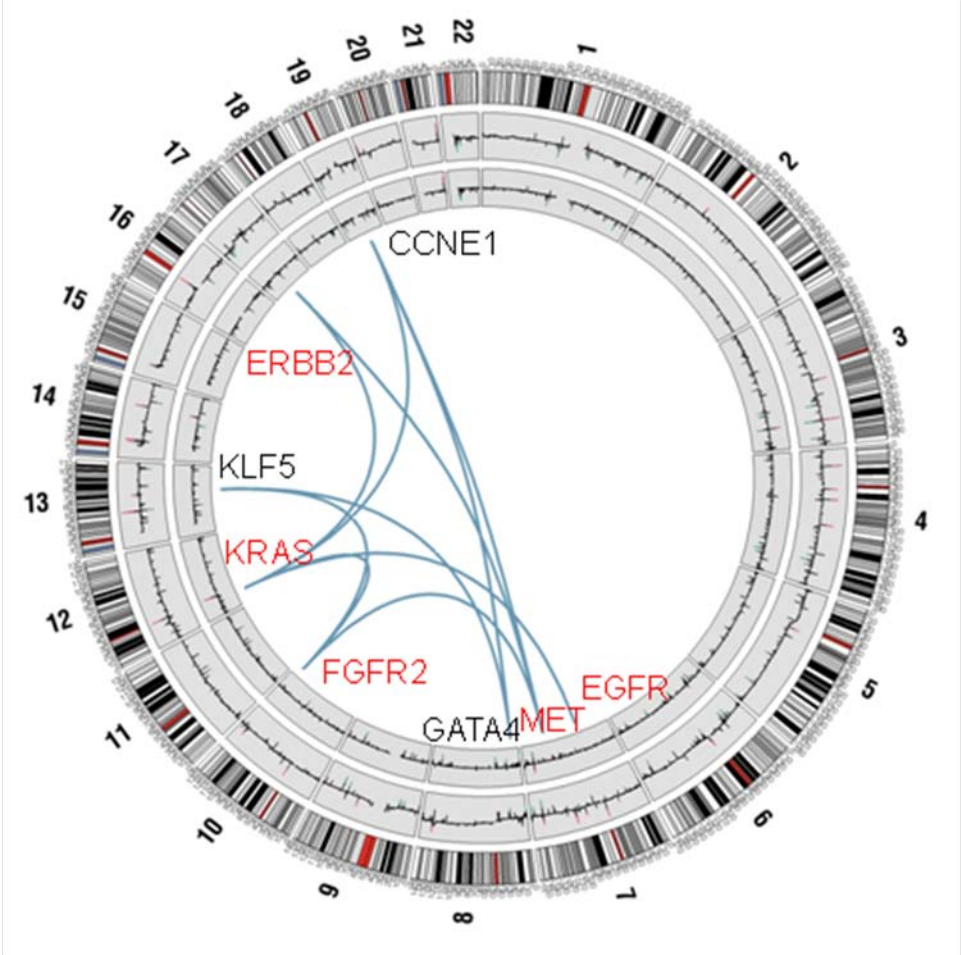


Figure 3-6 Mutually Exclusive Genomc Alterations

Focal regions exhibiting mutually exclusive patterns of genome amplification. Chromosomal diagrams were created using Circos software (Krzywinski, Schein et al. 2009). Circular tracks from outside to in: Genomic positions by chromosomes (black lines are cytobands, red lines are centromeres); summarized CNA values in gastric tumors, summarized CNA values in normal gastric samples. Blue lines indicate pairs of focal regions (genes) exhibiting significant patterns of mutually exclusive genomic amplification identified by DRP analysis ($p<0.05$; *EGFR/KRAS* – $p=0.05$). Genes involved in RTK/RAS signaling are highlighted in red.

Table A1 provides a complete list of significant mutually exclusive relationships for amplifications.

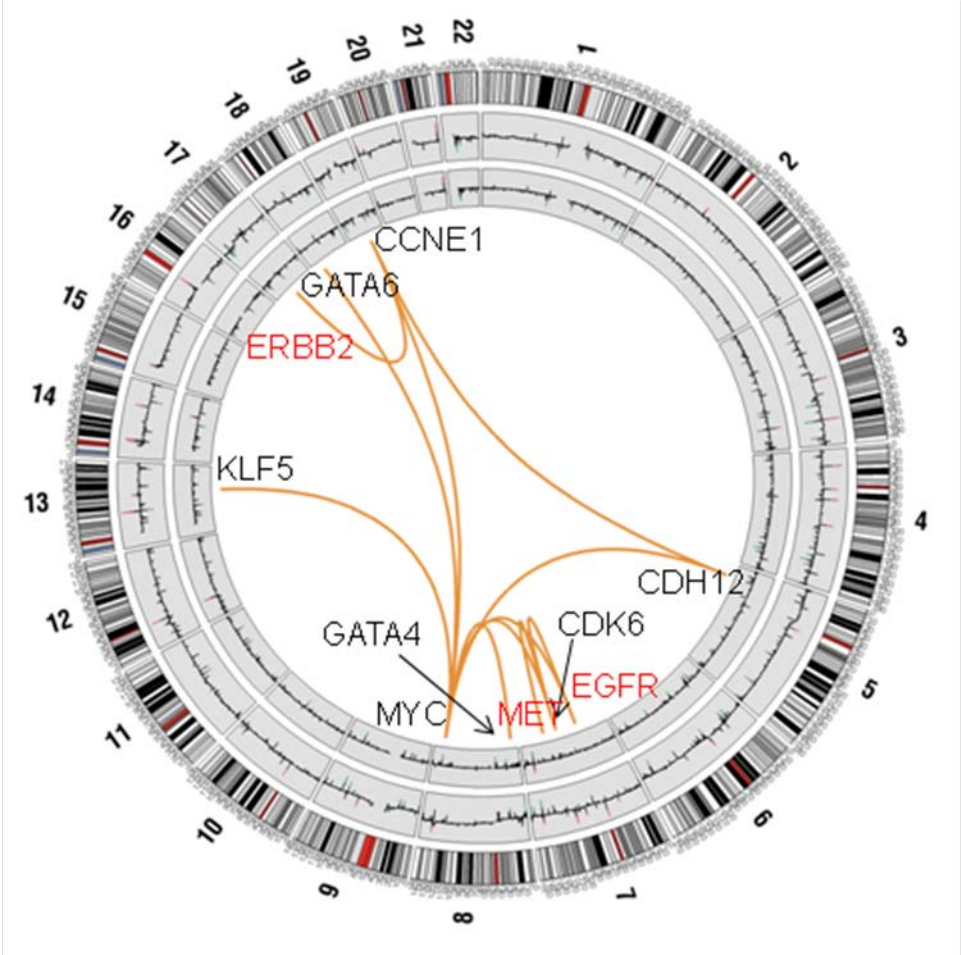


Figure 3-7 Co-Amplified Genomic Alterations

Focal regions exhibiting patterns of genomic co-amplification. Orange lines indicate pairs of focal regions (genes) exhibiting significant patterns of genomic co-amplification identified by DRP analysis ($p<0.05$). Genes involved in RTK/RAS signaling are highlighted in red.

Table A1 provides a complete list of significant co-alteration relationships for amplifications.

3.4 Genomic Alterations in RTK Signaling Genes - Frequent, Mutually Exclusive, and Associated with Patient Survival in GC

Motivated by the clinical success of trastuzumab and the availability of other RTK-targeting drugs in the GC translational pipeline (Asaoka, Ikenoue et al. 2011), we decided to characterize the RTK genomic alterations and their impacts on patient outcome. A heat-map representation of the SNP array data confirmed that the four amplified RTKs (*FGFR2*, *ERBB2*, *EGFR* and *MET*) were mutually exclusive to one another (Figure 3-8A). In addition, *KRAS* genomic amplifications were also mutually exclusive to the other RTKs (Figure 3-8A), suggesting these five components may activate the same downstream pathway in GC (Figure 3-9). The *KRAS* amplifications are examined in more detail in the next section.

Taken collectively, RTK/RAS genomic amplifications occurred in approximately 37% of the entire GC cohort (Figure 3-8B). The most frequently amplified RTK/RAS component was *FGFR2* (9.3%), followed by *KRAS* (8.8%), *EGFR* (7.7%) and *ERBB2* (7.2%). Of 72 tumors exhibiting amplification in at least one RTK/RAS component, 73.6% (53/72) exhibited amplification of only one component, and 26.4% (19/72) tumors exhibited high level amplification of one component with low level amplification of another. Only two tumors exhibited high level amplification of two RTK/RAS components (black arrows in Figure 3-8A). Taken collectively, these results suggest that 37% of the GC population is thus potentially targetable by a RTK/RAS directed therapy.

To assess the prognostic impact of RTK amplifications in GC, we performed a survival analysis comparing the clinical outcome of patients bearing tumors with RTK amplifications compared to patients with tumors lacking RTK amplification. In a

univariate analysis, patients with RTK amplified tumors (*FGFR2*, *ERBB2*, *EGFR*, *MET*) experienced poor survival outcome compared to patients with RTK amplification-negative cancers (p=0.01, Hazard Ratio (HR) 1.636, 95% Confidence Interval (CI) (1.101, 2.432); Figure 3-10). Moreover, in multivariate Cox regression models including RTK amplification status, stage, grade and treatment status (surgery alone or 5-FU adjuvant chemoradiation), RTK amplification status was shown to be an independent prognosis predictor (p = 0.01, HR 1.966, 95% CI (1.180, 3.279)) (Table 3-3). The adverse prognosis of RTK-amplified GCs was also largely independent from chromosomal instability (p=0.07), indicating that it is not a mere consequence of increased aneuploidy (Table 3-3) (Sanchez-Perez, Garcia Alonso et al. 2009).

To evaluate individual RTKs, we performed a follow-up univariate Cox model analysis considering the four different amplified RTKs (*FGFR2*, *ERBB2*, *EGFR* and *MET*) as independent factors. Patients with *ERBB2*-amplified tumors and *MET*-amplified tumors were found to exhibit the worst prognosis (*ERBB2*: p=0.0006, HR 2.824, 95% CI (1.558, 5.119); *MET*: p=0.002, HR 2.744, 95% CI (1.190, 6.327)) (Table 3-4). The adverse prognostic impact of *ERBB2* amplification was also observed in a multivariate Cox model with adjustment for tumor stage and grade (Table 3-5) (Tanner, Hollmen et al. 2005; Gravalos and Jimeno 2008). Thus, among the four different RTKs, *ERBB2* amplifications appear to exert the strongest prognostic impact in GC.

Figure 3-8 Genomic Alterations of RTK/RAS Signaling Components in GC

A) Mutually exclusive amplification patterns of RTK/RAS signaling components. In the heatmap, each row represents a different RTK/RAS signaling component. Each column represents an individual tumor exhibiting RTK/RAS amplification (72 tumors). The red color gradient (top right) highlights the degree of copy number amplification. Black arrows highlight two tumors exhibiting high level amplifications in two RTK/RAS components.

B) Overall frequency of RTK/RAS Genomic Alterations in GC. The pie chart displays the different GC subgroups exhibiting RTK/RAS amplification. GCs exhibiting at least one RTK/RAS amplification event comprise a collective 37% of the GC cohort analyzed.

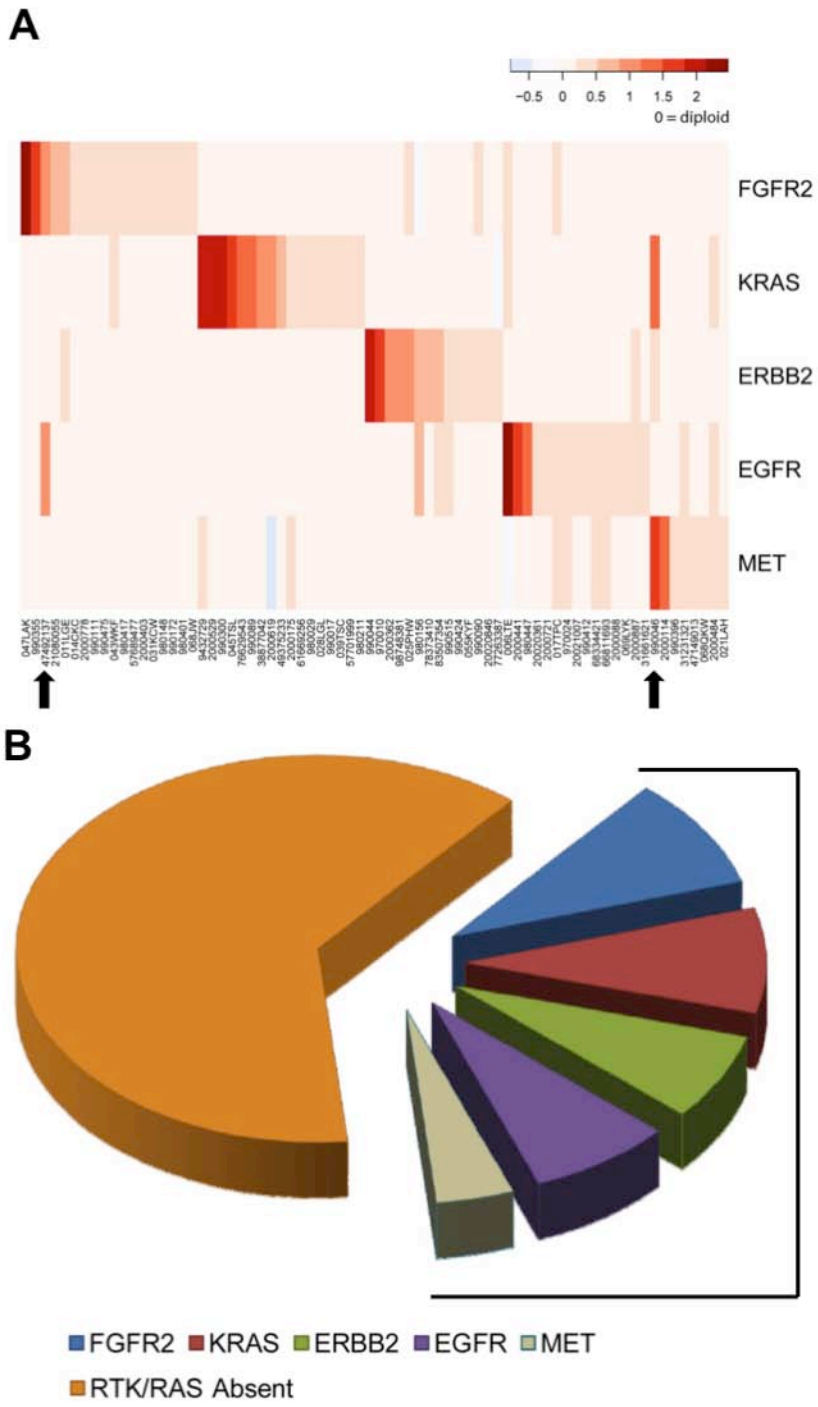

A
−0.5 0 0.5 1 1.5 2
0 = diploid
FGFR2
KRAS
ERBB2
EGFR
MET
B
FGFR2 KRAS ERBB2 EGFR MET
RTK/RAS Absent

Figure 3-9 Network Diagram Showing Relationship of RTK Signaling to RAS

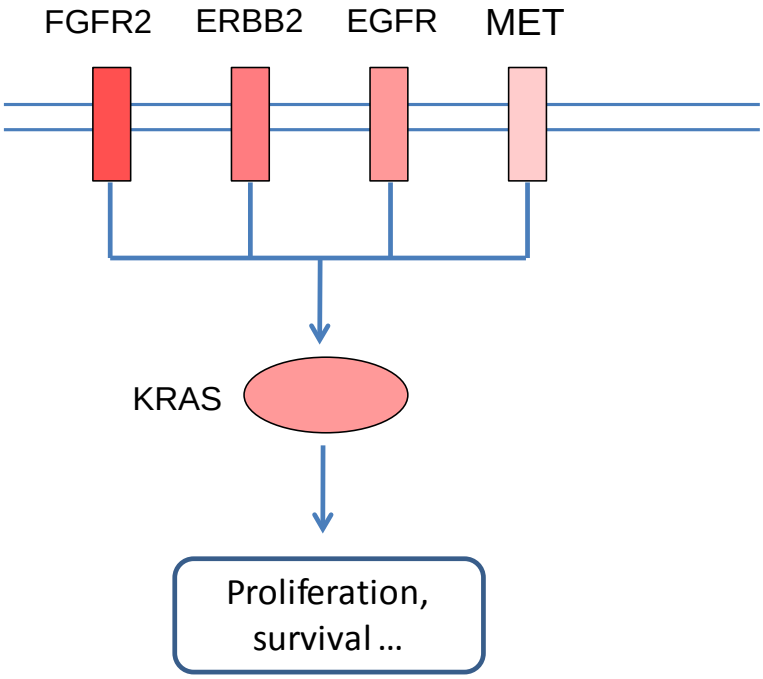


Table 3-3 Multivariate analysis comparing RTK amplification status with tumor stage, grade, adjuvant treatment and genome instability

(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 (see Methods section 2.3).

Table 3-4 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 3-5 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.

3.5 *KRAS*-genomic Amplifications Highlight a Previously Underappreciated GC Subgroup

KRAS amplifications were frequently observed in our series, occurring in 9% of patients. This finding is of interest, since canonical activating mutations in *KRAS* at codons 12 and 13 are strikingly infrequent in GC, unlike other GI cancers (e.g. colorectal and pancreatic cancer (Lievre, Bachet et al. 2006; Mita, Toyota et al. 2009)). Confirming these earlier studies (Mita, Toyota et al. 2009), the *KRAS* mutation rate in our own series was extremely low - among 139 GCs genotyped for *KRAS* codon 12 and 13 mutations, only one tumor exhibited a *KRAS* mutation (G13D in 069LYK). We thus hypothesized that *KRAS* genome amplification, rather than mutation, may represent a predominant mechanism for *KRAS* activation in GC.

To obtain additional evidence that *KRAS* genomic amplifications represent a distinct GC molecular subgroup, we performed a Kaplan Meier survival analyses comparing outcomes of patients with *KRAS* amplified samples versus patients with tumors lacking RTK or *KRAS* amplification. Patients with *KRAS*-amplified tumors exhibited significantly poorer prognosis (p=0.01, HR 2.158, 95% CI (1.172, 3.971) (Figure 3-10). Supporting the robustness of this survival association, similarly significant associations were observed when patients with *KRAS*-amplified tumors were compared against patients lacking *KRAS*-amplification but irrespective of RTK amplification, or when the copy number threshold defining *KRAS* amplification was relaxed (p=0.06, HR 1.744, 95% CI (0.973, 3.127); p=0.01, HR 1.665, 95% CI (1.114, 2.488); Figure 3-11).

To benchmark the prognostic effect of *KRAS* amplification against other RTKs, we applied a univariate Cox regression model consisting of all five genes. Similar to *ERBB2*

and *MET* amplifications, GC patients with *KRAS* amplifications also exhibited significantly worse prognosis relative to patients with tumors lacking either RTK or *KRAS* amplifications (p=0.02, HR 2.116, 95% CI (1.155, 6.940)) (Table 3-6), however this association may be related to tumor stage (p=0.2, HR 1.455, 95% CI (0.790, 2.682)) (Table 3-7).

Finally, to provide functional evidence that *KRAS* genomic amplification represents an important 'driver' event in *KRAS*-amplified GCs, we performed genetic knockdown experiments. siRNA-mediated knockdown of *KRAS* in *KRAS* amplified and *KRAS*-mutated GC cell lines caused significant reductions in proliferation but not in *KRAS*-wild-type lines, supporting an earlier report (Mita, Toyota et al. 2009) (Figure A2). These results suggest that *KRAS* amplification in GC likely define a specific subgroup of poor prognosis patients for which *KRAS* signaling in tumors is critical.

Figure 3-10 Genomic Alterations of RTK/RAS Signaling Components in GC (cont'd)

(above) Kaplan-Meier survival analysis comparing outcomes of patients with tumors exhibiting RTK amplification (either *FGFR2*, *ERBB2*, *EGFR*, or *MET*) amplification to patients with tumors lacking RTK amplification. Patients with tumors exhibiting focal *KRAS* amplifications were included in analysis, and fall into the RTK low/no CNA group. Overall survival was used as the outcome metric.

(bottom) Kaplan-Meier survival analysis comparing outcomes of patients with tumors exhibiting *KRAS* amplification (15 patients) to patients with non-RTK/*KRAS* amplified tumors. Overall survival was used as the outcome metric. The inset photo displays a patient tumor (ID 49375233) with *KRAS* amplification confirmed by FISH analysis (blue – DAPI nuclear stain, green – *KRAS* FISH probe, red = centromere 12 probe).

FISH experiment was performed by Kakoli Das and was necessarily included to present the whole story.

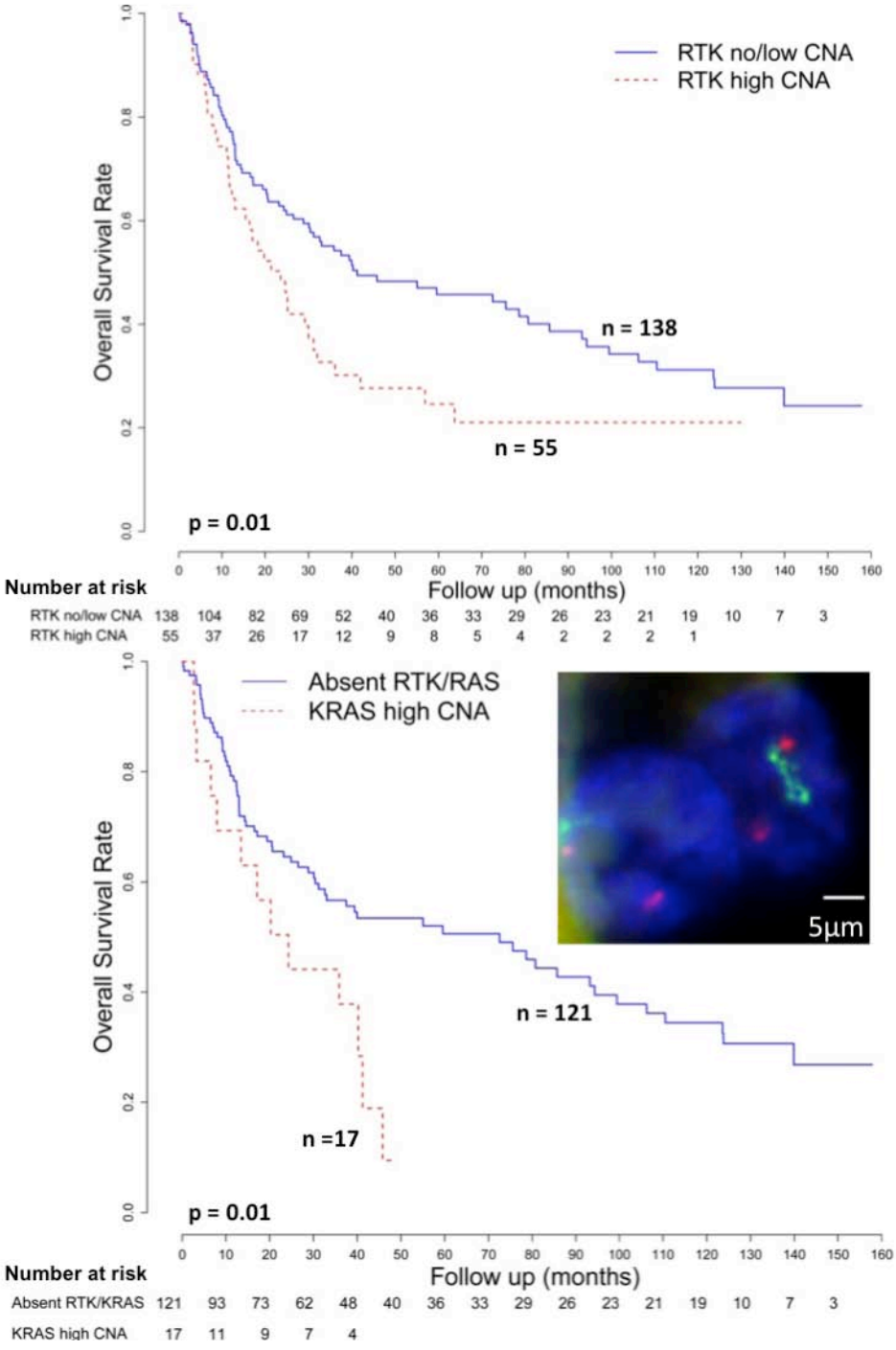
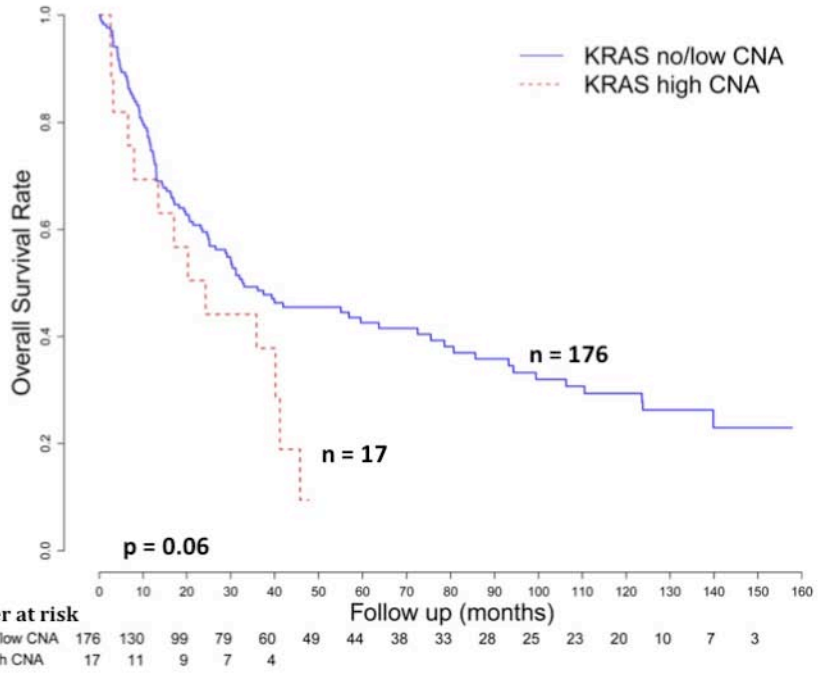

RTK no/low CNA
RTK high CNA
Overall Survival Rate
n = 138
n = 55
p = 0.01
Follow up (months)
Number at risk
RTK no/low CNA 138 104 82 69 52 40 36 33 29 26 23 21 19 10 7 3
RTK high CNA 55 37 26 17 12 9 8 5 4 2 2 2 1
Absent RTK/RAS
KRAS high CNA
5µm
n = 121
n =17
p = 0.01
Follow up (months)
Number at risk
Absent RTK/KRAS 121 93 73 62 48 40 36 33 29 26 23 21 19 10 7 3
KRAS high CNA 17 11 9 7 4

Figure 3-11 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 3-8A 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).

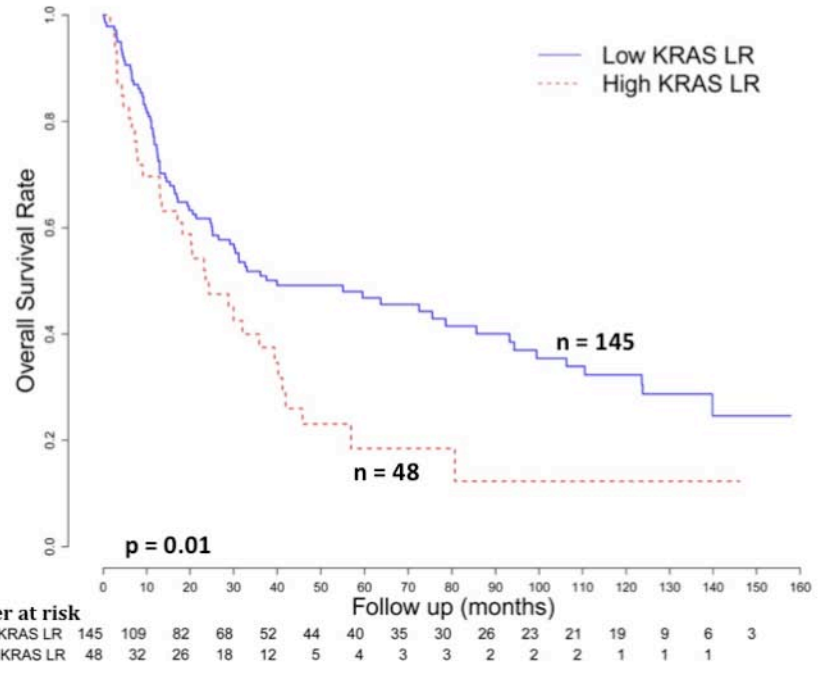
A



Number at risk

KRAS no/low CNA	176	130	99	79	60	49	44	38	33	28	25	23	20	10	7	3
KRAS high CNA	17	11	9	7	4											

B



Number at risk

Low KRAS LR	145	109	82	68	52	44	40	35	30	26	23	21	19	9	6	3
High KRAS LR	48	32	26	18	12	5	4	3	3	2	2	2	1	1	1	

Table 3-6 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 3-7 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.

3.6 *FGFR2* Amplifications in GC: Relationships to Gene Expression, Clinical Outcome, and Drug Sensitivity

FGFR2 was being amplified in 9-10% of GCs in our series (Table 3-1). Consistent with *FGFR2* being the main driver of amplification in this locus, intersection of the amplification regions across twenty *FGFR2*-amplified tumors confirmed that *FGFR2* was the sole gene in this region exhibiting common copy-number gain (Figure 3-12). Validating the SNP data, a quantitative PCR (qPCR) analysis using primers directed towards *FGFR2* confirmed that samples with high *FGFR2* qPCR values were associated with *FGFR2* amplification. ($p = 0.0006$, Fisher test) (Figure 3-13). FISH analysis using BAC probes targeting *FGFR2* also confirmed *FGFR2* gene amplification in patient tumors and cell lines, relative to a centromere 10 probe (Figure 3-14).

FGFR2 has been previously proposed as a potential therapeutic target in GC (Asaoka, Ikenoue et al. 2011), but little is known regarding the impact of *FGFR2* amplification on gene expression and other clinicopathologic parameters. To investigate relationships between *FGFR2* gene amplification and *FGFR2* gene expression, we analyzed gene expression profile data for 156 of the 193 GCs analyzed by SNP arrays in this study, which we have described in an earlier report (Ooi, Ivanova et al. 2009). *FGFR2*-amplified GCs indeed exhibited significantly increased *FGFR2* gene expression levels (Figures 3-15 and 3-16), when compared against a reference set of 100 normal gastric samples, or to non-*FGFR2* amplified tumors (Kruskal-Wallis test : p=6.7e-9, Wilcoxon test : p=1.7e-7 (vs normal) and p=1.9e-5 (vs non-FGFR2 amplified GCs). In comparison, *ATE1* and *BRWD2*, two genes located adjacent to *FGFR2* exhibited less significant levels

of copy number/gene expression correlation (p=0.004-0.3, relative to normals) (Figure 3-17), further supporting *FGFR2* as the major driver gene in this region.

Examining clinicopathologic variables, *FGFR2*-amplified GCs did not exhibit any significant associations with histology (Lauren's; p=0.8, Grade; p=0.8; or Tumor Stage; p = 0.9) or patient survival (p= 0.8, Table 3-4). However, in an expanded gene expression data set of 398 gastric tumors derived from four distinct cohorts of which the previous 156 GCs form a subset (see Methods section 2.7 and Table 3-8), high *FGFR2* expression (compared to normals, Figure 3-18) was associated with poor survival outcome in a univariate analysis (p=0.01, HR 1.492, 95% CI (1.094, 2.035)) (Figure 3-19). In a multivariate Cox regression model, samples with *FGFR2* high expression tended to exhibit borderline significance after adjusting for stage and grade (p=0.08, HR 1.321, 95% CI (0.966, 1.807)) (Table 3-9). This result suggests that *FGFR2* overexpression in GC may be of prognostic relevance.

Dovitinib (TKI258) is an investigational multi-targeting oral tyrosine kinase inhibitor with potent inhibitory activity against bFGF receptors-1,2,3, VEGF receptors-1,2,3, PDGFR and *c-KIT* (Lee, Lopes de Menezes et al. 2005; Trudel, Li et al. 2005). In pre-clinical models, dovitinib has exhibited anti-tumor activity in *FGFR1* amplified breast cancer (Dey, Bianchi et al. 2010), and in several phase one clinical trials has shown good therapeutic profiles in human patients (Ocio, Mateos et al. 2008; Sarker, Molife et al. 2008). To test the potential efficacy of dovitinib in *FGFR2*-amplified GC, we treated *FGFR2*-amplified and non-amplified GC lines (Figure A3-A) with increasing dosages of dovitinib, to determine the GI50 concentration (the drug concentration required to cause

50% growth inhibition). We observed potent growth inhibitory activity of dovitinib specifically in *FGFR2* amplified gastric cancer cell lines with GI50 dosages in the sub micromolar range (KATO-III : 0.12uM; SNU-16 : 0.17uM, Figure A3-B). Decreased phosphorylation of *FGFR2*, ERK and AKT was also observed after 1 hour of dovitinib treatment (Figure A3-C). Besides inhibiting cell proliferation, dovitinib treatment also induced a significant decrease in soft-agar colony formation in *FGFR2*-amplified lines (KATO III: p=0.002; SNU16 : p=0.05; Figure A3-D and Figure A4). In a cell death assay, dovitinib treatment induced apoptosis, measured by caspase 3/7 activation, in SNU-16 cells after 24 hours of treatment, but not in KATO III cells (Figure A3-E). These results suggest that dovitinib treatment can inhibit several pro-oncogenic traits in *FGFR2*-amplified lines, but additional factors may be required for *FGFR2*-amplified cells to undergo apoptosis upon dovitinib treatment.

To evaluate the efficacy of dovitinib in an *in vivo* model, we performed drug treatment experiments using an *FGFR2*-amplified primary human gastric cancer xenograft model, comparing dovitinib responses to the positive control drug 5-FU. Mean tumor sizes of vehicle treated mice reached 1163 mm^3 at day 25 post treatment, while treatment with 5-FU at 20 mg/kg (Qd x 5/wk x 2 wks, i.p.) produced a reduced mean tumor size of 518 mm^3 (Total Growth Inhibition (TGI) = 63%, p = 0.08) after the same period. Importantly, treatment with dovitinib at 30 mg/kg and 50 mg/kg (Qd x 25 days, *p.o.*) significantly inhibited tumor growth compared to vehicle treated tumors (p = 0.006 and 0.002, respectively), with final tumor sizes of 194 and 53 mm^3, respectively, at day 25 post treatment (Figure A3-F). Dovitinib may thus represent a promising subtype-specific therapy for *FGFR2*-amplified GCs.

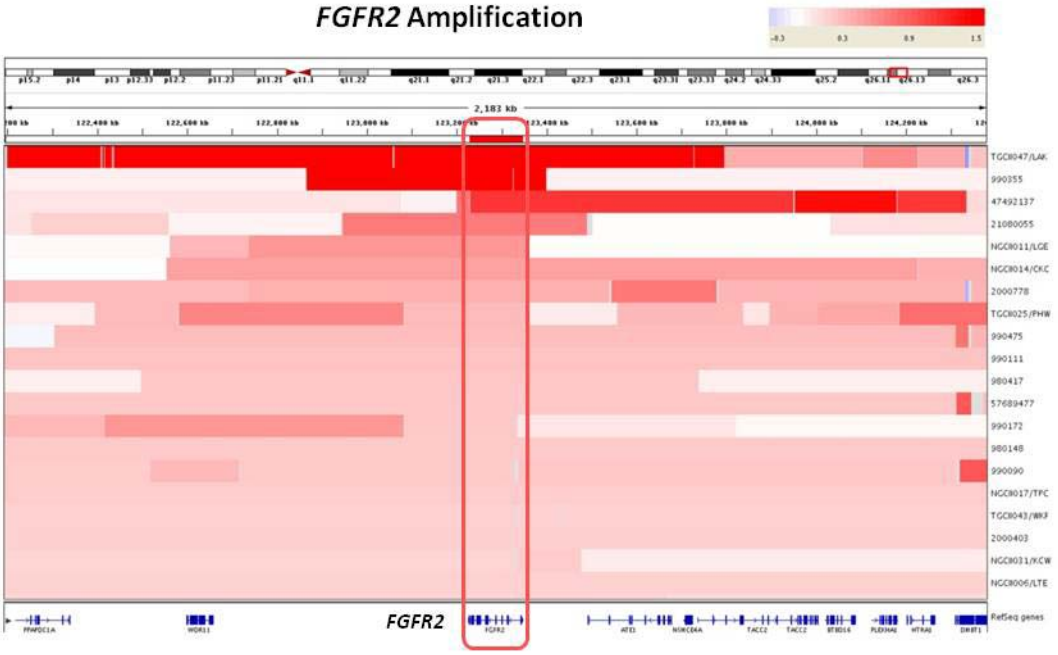
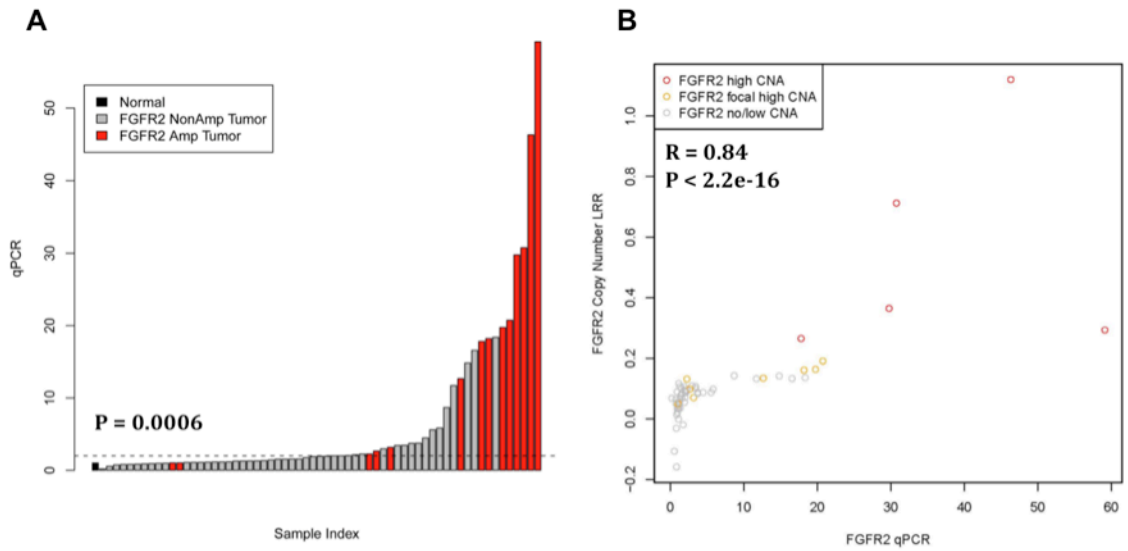


Figure 3-12 *FGFR2* Gene Amplification in GC

Heatmap showing the *FGFR2* gene amplification region in individual GC samples (20 tumors). Each row indicates one GC sample with the amplified region in red. Intensity of the red bar indicates the level of copy number amplification. Genes located in this region are shown at the bottom. The intersection of these amplified regions covers only the *FGFR2* gene (red box, gene outlined at bottom).

Figure 3-13 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 3-8A) 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 3-8A), 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)



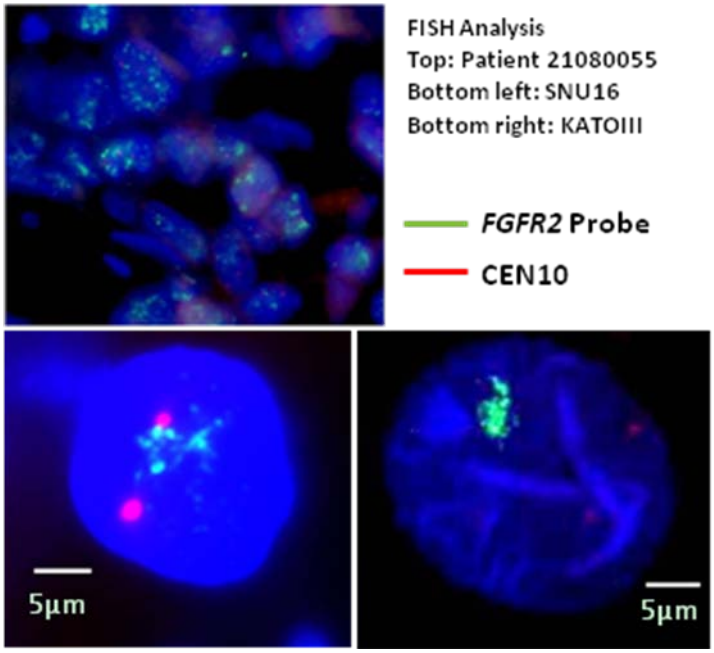


Figure 3-14 *FGFR2* genomic amplification confirmed by FISH.

The photo displays a patient tumor (ID 21080055) with *FGFR2* amplification and two *FGFR2*-amplified cell lines KATO-III and SNU16 confirmed by FISH analysis. Green signals indicate the *FGFR2* FISH probe, red signals probes to centremore 10.

FISH experiment was performed by Kakoli Das and was necessarily included to present the whole story.

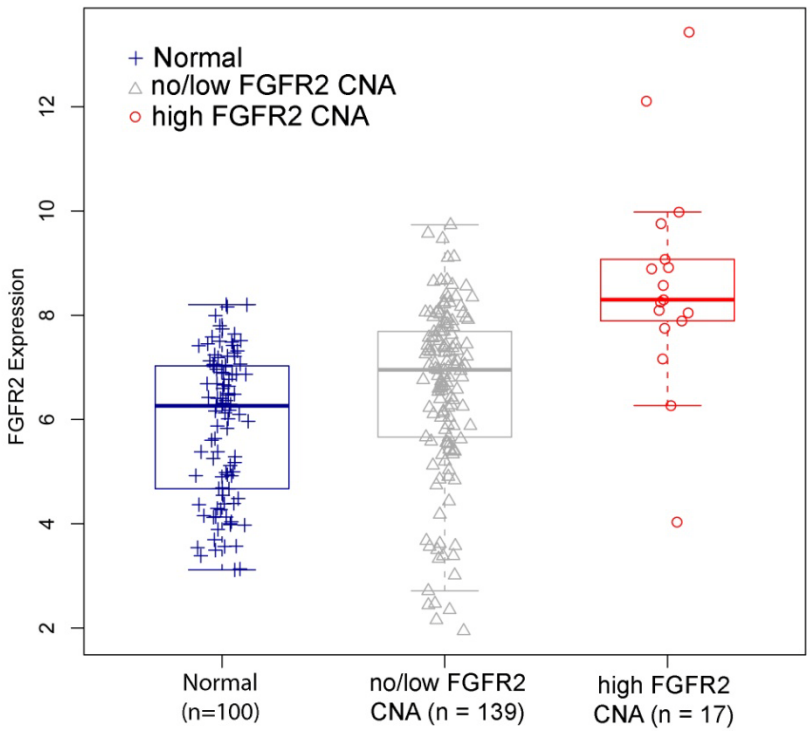


Figure 3-15 *FGFR2* gene expression in clinical specimens.

FGFR2 gene expression was compared across three categories, each represented by a box-plot: non-malignant gastric tissues (normal) (n = 100), tumors exhibiting no/low *FGFR2* CNA (n = 139), and tumors exhibiting high *FGFR2* CNA (n = 17). mRNA comparisons were based on 156 GCs where gene expression data was available, representing a subset of the 193 GCs analyzed by SNP arrays. *FGFR2* gene expression was inferred from Affymetrix microarrays (*FGFR2* probe 211401_s_at). *FGFR2* mRNA levels are significantly higher in samples with *FGFR2* high CNAs compared to the other two categories (p = 6.7e-9, Kruskal Wallis test). Tumors exhibiting *FGFR2* amplification exhibit significantly increased *FGFR2* gene expression compared to tumors exhibiting no/low *FGFR2* CNA or non-malignant samples (p = 1.9e-5 and 1.7e-7, Wilcoxon test).

Figure 3-16 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.

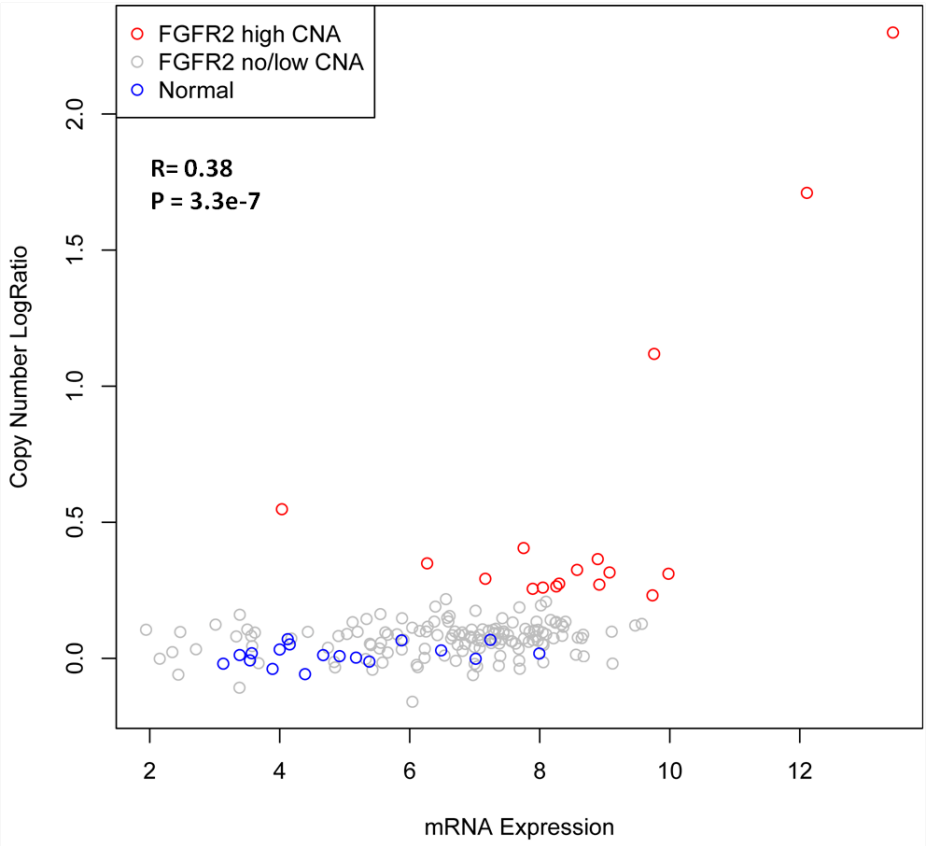


Figure 3-17 Relationship between Copy Number and Gene Expression for *ATE1* and *BRWD2*, Genes Adjacent to *FGFR2*

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.

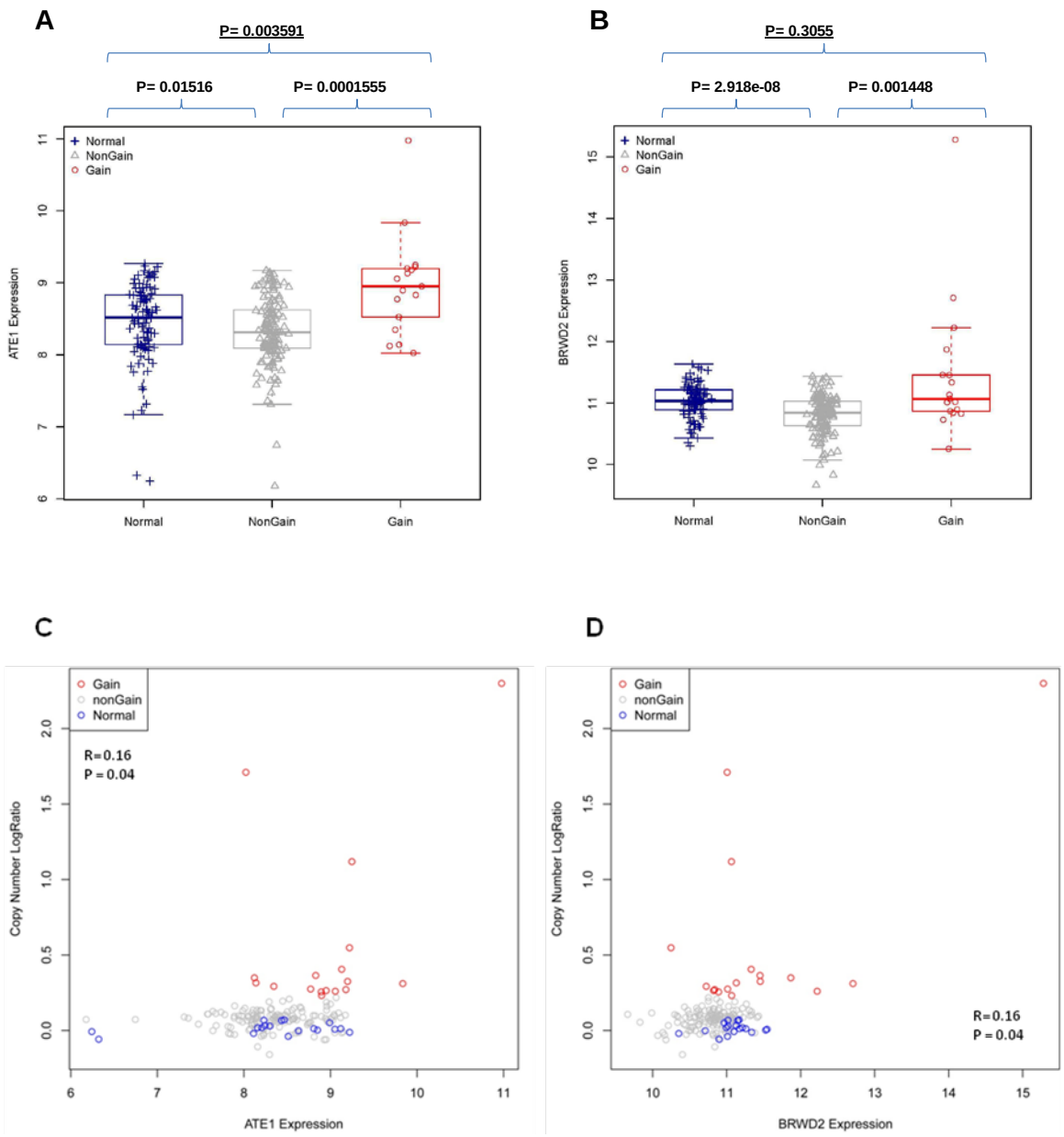
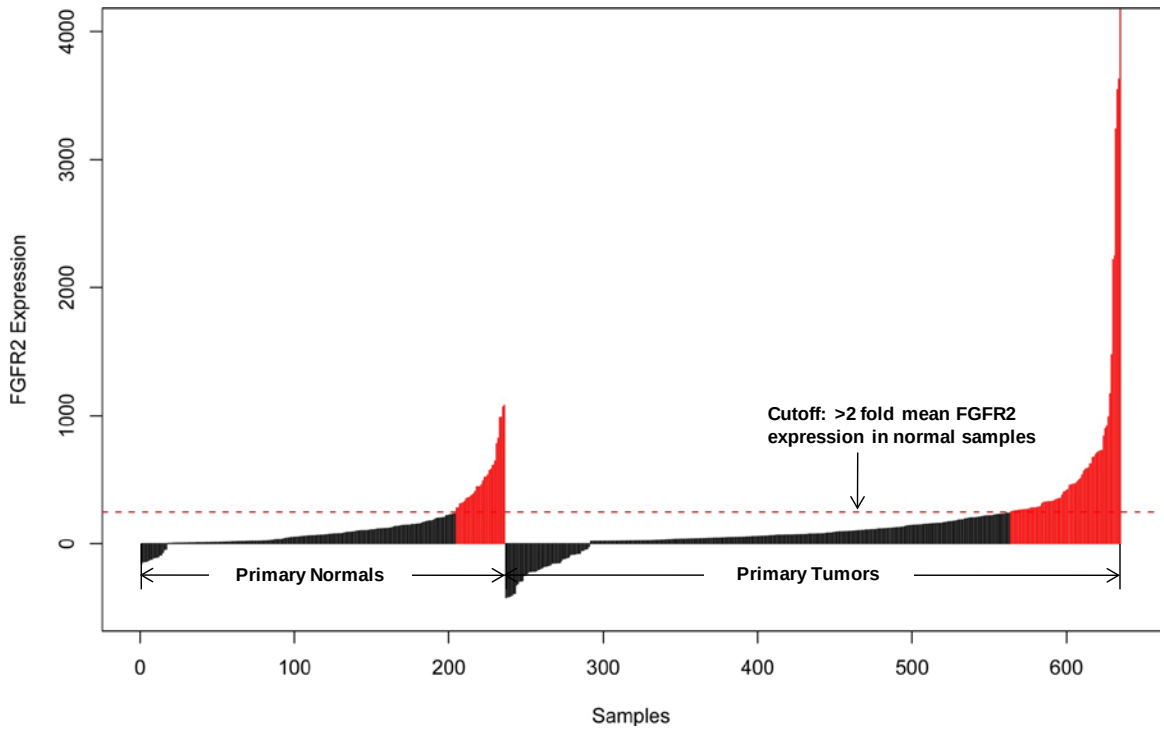

A
P= 0.003591
P= 0.01516
P= 0.0001555
Normal
NonGain
Gain
ATE1 Expression
Normal
NonGain
Gain
B
P= 0.3055
P= 2.918e-08
P= 0.001448
Normal
NonGain
Gain
BRWD2 Expression
Normal
NonGain
Gain
C
Gain
nonGain
Normal
R=0.16
P = 0.04
Copy Number LogRatio
ATE1 Expression
D
Gain
nonGain
Normal
R=0.16
P = 0.04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

Table 3-8 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

Figure 3-18 *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).



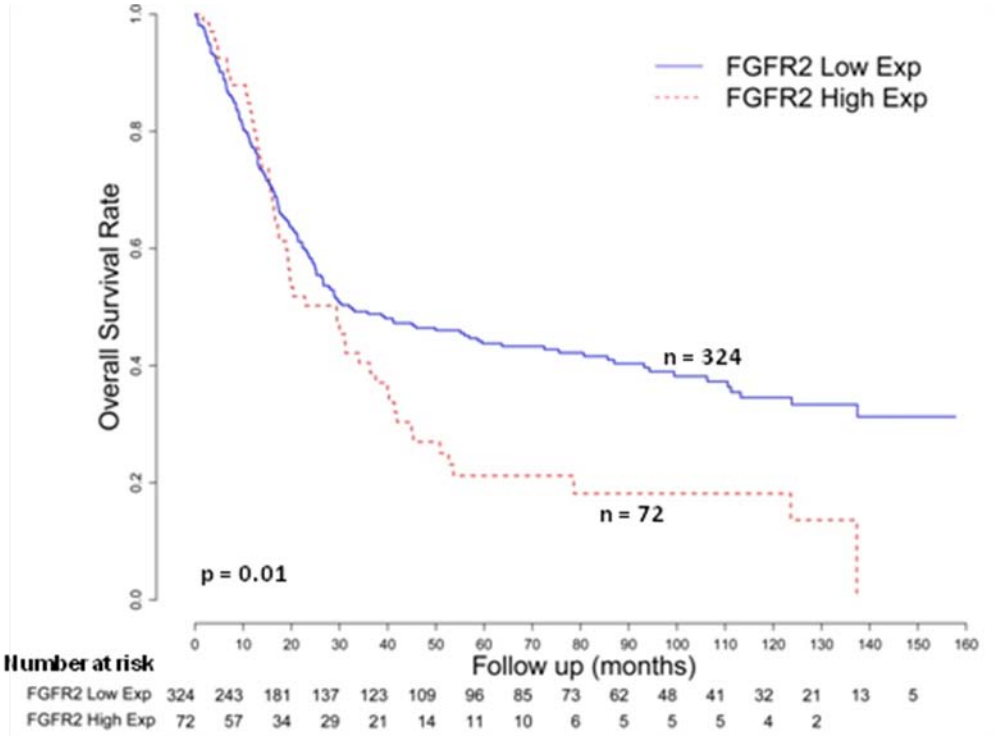


Figure 3-19 Kaplan-Meier survival analysis of GC patients by *FGFR2* status

Kaplan-Meier survival analysis comparing patients with tumors exhibiting high *FGFR2* gene expression, defined as 2 fold higher than the average *FGFR2* gene expression level in normal samples (72 tumors) to patients with tumors exhibiting low *FGFR2* gene expression (total = 398 patients, the 156 patients analyzed in Figure 3-15 are a subset of these 398 patients). Overall survival was used as the outcome metric.

Table 3-9 Multivariate analysis comparing high *FGFR2* gene expression (>2-fold mean level in normal gastric tissues) with tumor stage and grade

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.

3.7 GATA Factors and *KLF5* are Candidate Lineage-Specific Oncogenes in GC

Finally, in addition to classical oncogenes, we analyzed the SNP array data set to identify novel genes not previously reported to be amplified in GC. We were particularly interested in "lineage-specific" oncogenes, a recently-described class of genes that contribute to tumorigenesis in specific developmental lineages, and which tend to be transcription factors (Garraway and Sellers 2006). To identify candidate lineage-specific factors in GC, we filtered the GC SNP data against lists of genes highlighted as amplified in comparable large scale copy number studies from glioblastoma, lung adenocarcinoma and a study covering multiple cancer types (Weir, Woo et al. 2007; Chaudhuri, Handcock et al. 2008; Beroukhim, Mermel et al. 2010). Our analysis revealed three transcription factors, *GATA4*, *GATA6* and *KLF5* to be specifically amplified in GC. Interestingly, recent data has implicated GATA factors, and particularly *GATA6*, in gastrointestinal cancer. For example, *GATA6* has been shown to be amplified in pancreatic cancer, and may also exhibit oncogenic properties in oesophageal cancer (Kwei, Bashyam et al. 2008; Alvarez, Opalinska et al. 2011).

KLF5 amplification was observed in 11% of GCs (22 tumors). Intersection of the minimal common amplified region in these twenty-two samples highlighted only the *KLF5* gene (Figure 3-20A), and *KLF5* amplifications were confirmed in both cell lines and primary tumors (Figure 3-20B). *KLF5* amplifications were also significantly associated with increased *KLF5* gene expression ($p=0.00086$, Wilcoxon test), consistent with *KLF5* behaving as the driver gene in this region (Figure 3-21). To define prevalent biological themes associated with *KLF5* amplification, we generated a *KLF5* expression signature of 425 genes (175 upregulated, 250 downregulated, FDR<0.001) by comparing

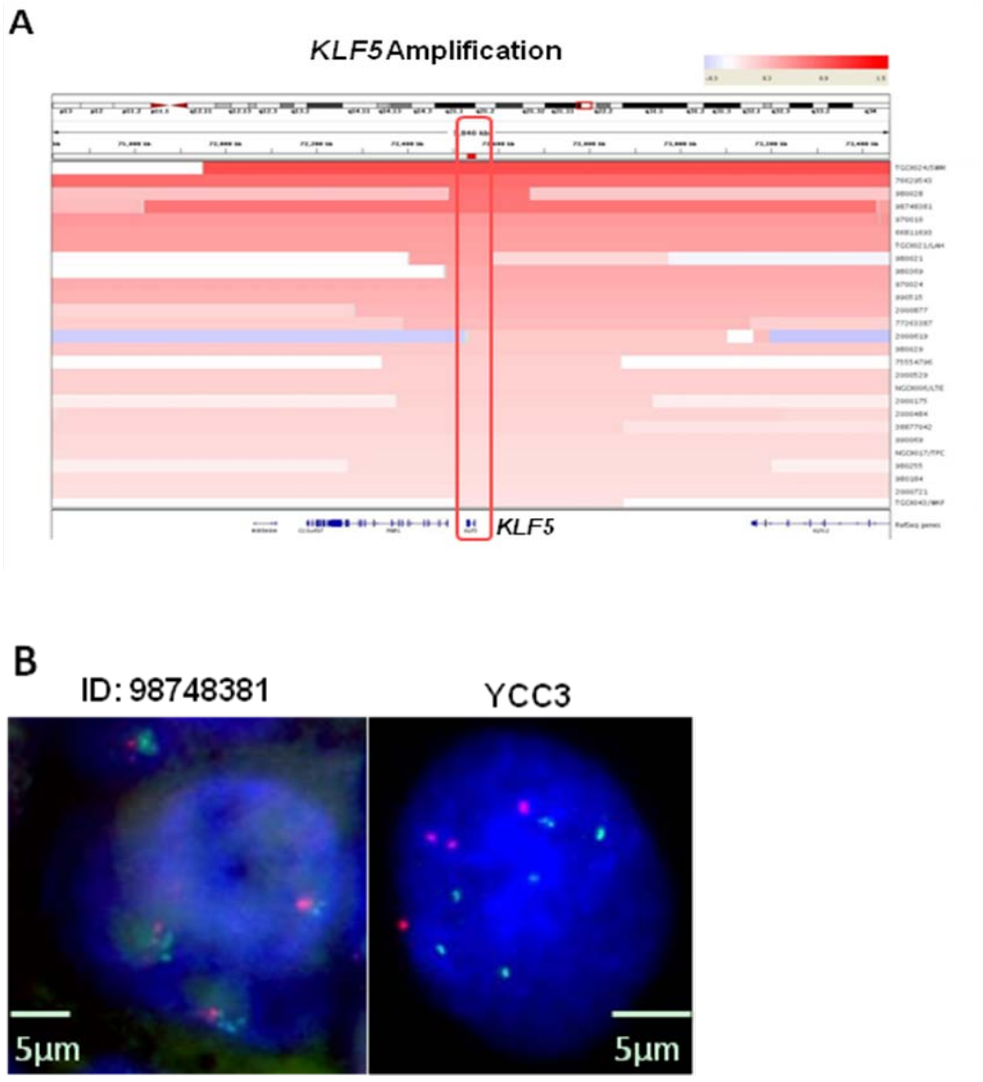
the mRNA expression profiles of the top and bottom 10% of *KLF5*-expressing tumors. Gene set enrichment analysis (Mootha, Lindgren et al. 2003; Subramanian, Tamayo et al. 2005) of genes upregulated in *KLF5* expressing tumors revealed significant enrichments in genesets related to early cancer (gastric, ovarian), epithelial-mesenchymal transition, *MYC* targets, and embryonic stem cell networks (p-values ranging from 0 to $2x10^{-7;}$). These results support a plausible role for *KLF5* in GC development and progression.

To establish a functional role for *KLF5* in GC, we performed Western blotting to screen *KLF5* protein levels in GC cell cells. SNU5 and YCC3 cells, confirmed to exhibit *KLF5* genomic amplifications by both SNP arrays and FISH analysis also exhibited high *KLF5* protein expression (Figure A5-A). Conversely, SNU1, a non-*KLF5* amplified line, expressed relatively low *KLF5* protein levels and normal *KLF5* copy number levels. To investigate the functional consequences of *KLF5* in GC cells, we treated SNU5 and YCC3 with *KLF5* siRNAs, and confirmed successful *KLF5* knockdown at the protein level (Figure A5-B, top Western lots). Both SNU5 and YCC3 cells silenced with *KLF5* siRNAs resulted in a significant reduction in cell proliferation capacity compared to cells treated with scrambled siRNAs ($p = 0.006$ and $p=0.011$ respectively). Similar results were observed when these experiments were repeated using multiple independent and non-overlapping *KLF5* siRNAs, indicating that this growth inhibition is likely not an off-target effect. Besides growth proliferation, *KLF5* siRNA treatment also reduced soft-agar colony formation in YCC3 cells ($p=0.01$, experimental results not shown). Notably, a similar experiment could not be performed for SNU5 cells as they do not form colonies in soft agar. To address the consequences of *KLF5* overexpression, we then ectopically expressed *KLF5* in SNU1 cells, which normally express low *KLF5* levels. Compared to

control transfected cells, *KLF5* over-expressing SNU1 cells exhibited enhanced cell proliferation ($p<0.0001$) and soft-agar colony formation ($p=0.04$; experimental results not shown). These results suggest that *KLF5* may be important for various cancer cell traits in GC.

Figure 3-20 *KLF5* amplifcation in GC

A) Heatmap showing the *KLF5* gene amplification region in 22 GC samples. Each row indicates one GC sample with the amplified region in red. Intensity of the red bar indicates the level of copy number amplification. Genes located in this region are shown at the bottom. The intersection of these amplified regions covers only the *KLF5* gene (red box). B) *KLF5* FISH analysis. *KLF5* gene amplifications were confirmed in a primary tumor (ID 98748381) and cell lines (YCC3). Green signals represent the *KLF5* FISH probe. Red signals in the primary tumor represent hybridaztion of a control LSI D13S319 probe, which binds close to the Chromosome 13 centromere. Red signals on right represent a Centromere 13/21 probe, confirming two Chromosome 13s and two Chromosome 21s.



FISH experiment was performed by Kakoli Das and was necessarily included to present the whole story.

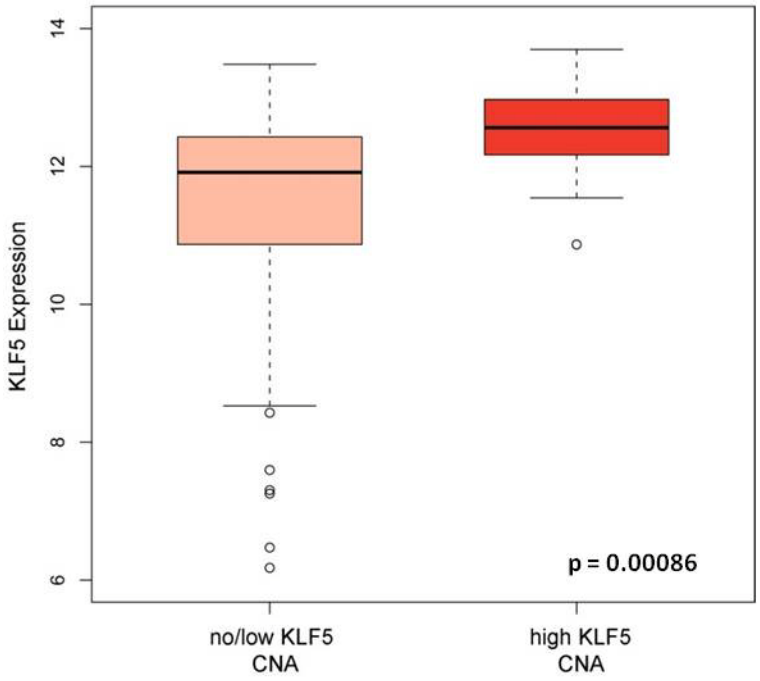


Figure 3-21 *KLF5* Gene Expression is associated with copy number status

KLF5 gene expression in clinical specimens. *KLF5* mRNA expression levels were compared between samples with high *KLF5* copy number (high CNA) versus samples with normal or low-level amplification (no/low CNA). *KLF5* gene expression was inferred from Affymetrix microarrays (*KLF5* probe 209212_s_at). P-values (Wilcoxon test) indicate that *KLF5* mRNA levels are significantly higher in samples with *KLF5* copy number gains.

Chapter 4 Application of CNA landscape in other associated GC genomics studies

Cancer is a complex disease, multi-level cancer genomics data has to be integrated for a better understanding of the cancer development. For this purpose, a few large scale projects, such as TCGA studies in various types of cancers (2008; 2011; 2012; 2013; Kandoth, Schultz et al. 2013) were designed for integrative genomics analyses involving multiple genomics datasets including gene expression, DNA copy number, epigenetics and exome or whole genome sequencing. In this study, the genomic DNA copy number profiling is by far the largest sample size for Asian cohort using a high resolution SNP array, providing a useful resource for integrative studies with other type of data for GC research.

There are many advantages of integrating data from different platforms in cancer research. Firstly, cancer driver genes usually undergo various mechanisms of genetic aberrations. Oncogene activation can be triggered through mutation, gene amplification or structural rearrangement (Tabin, Bradley et al. 1982), while a tumor suppressor gene loss of function can be activated through a classic "two-hit" hypothesis, by mutation, LOH, or genomic deletion (Lamlum, Ilyas et al. 1999). Combining data from several platforms may highlight a larger patient population associated with the driver gene. For example, *PIK3CA* has a 3.6% mutation and about 18% amplification in lung cancer (Okudela, Suzuki et al. 2007), these together suggest a higher percentage of lung cancer patients who can benefit from *PIK3CA* targeted treatment. Secondly, using data from more than one platform may help to enhance understanding of cancer subtypes. For example, integrative study of glioblastoma identified four subtypes, each subtype was

characterized by combined gene expression, somatic mutation and DNA CNAs information in a number of key genes: *EGFR*, *NF1*, and *PDGFRA*/*IDH1* (Verhaak, Hoadley et al. 2010). Thirdly, integrative analysis can help to identify novel cancer associated genes. For examples, integration of somatic mutation or DNA copy number data with mRNA level gene expression may help to identify potential candidate genes with DNA associated mRNA level alterations. In this chapter we will demonstrate these advantages by applying the landscape of CNAs to integrate with other types of data in GC study.

4.1 Integration of CNA with Next Generation Sequencing can identify cancer driver genes

Zang, Cutcutache et al. 2012 identified recurrent key genes' somatic mutations in cell adhesion and chromatin remodeling pathway in GC by exome sequencing. One of the cell adhesion genes was found to be mutation in about 5% of total GC cases is *FAT4*. *FAT4* belongs to the E-cadherin family, a gene family that has been implicated in gastric cancer (Berx and van Roy 2009) and that may regulate non canonical Wnt/planar cell polarity signaling (Wang 2009). *FAT4* is widely expressed in many tissue types and may have a key role in gastrointestinal tract development (Saburi, Hester et al. 2008; Mao, Mulvaney et al. 2011). Mutation pattern of *FAT4* suggested it might play a tumor suppressor role in GC. In order to identify further evidence, we selected matched tumor normal pairs profiled by SNP6 array for copy number analysis. Indeed, we observed about 4% somatic deletions in the matched samples (83 pairs) (Fig 4-1). In addition, with the matched mRNA gene expression data is available, we found that samples with *FAT4* gene deletion are associated with extremely low *FAT4* expression ($p = 0.012$, Fig 4-2). Another critical recurrently mutated gene reported in the paper is *ARID1A*, with ~8% prevalence in 110 GC samples, which involves chromatin remodeling process. It's important to study the functional effect of *ARID1A* to better understand its role in GC. However, it's difficult to identify appropriate cell line model with *ARID1A* mutations. Instead, by using the copy number SNP array for GC cell lines, we observed two cell lines model with *ARID1A* deletions (Fig 4-1). Re-expression of *ARID1A* in these two cell lines repressed cell proliferation, suggesting its anti-oncogenic role. *FAT4* and *ARID1A* are just two examples to show their genomic abnormality by multiple mechanisms, i.e., somatic mutation and genomic CNAs. There may be more genes of interest in GC identified

through other mechanisms; the data in this copy number study could serve as a useful resource for comprehensive study of genomic alterations at individual gene level

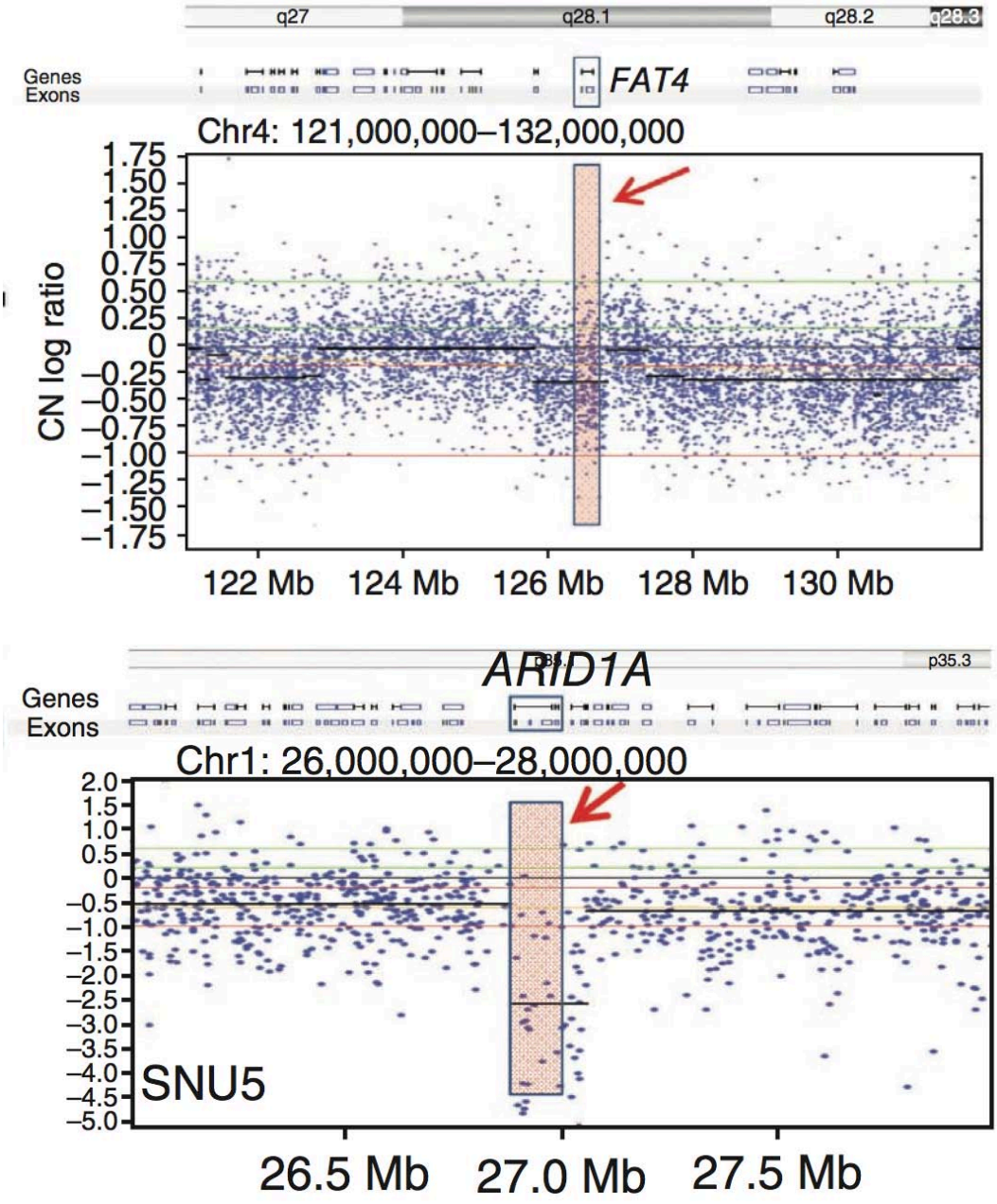


Figure 4-1 SNP array identifies *FAT4* and *ARID1A* deletion in GC

(above)A representative FAT4 genomic deletion (red arrow) in gastric cancer based on Affymetrix SNP array. Blue dots represent probes from SNP array, black lines are CBS segmented copy number logratio.

(bottom) *ARID1A* genomic deletion (red arrow) in gastric cancer cell lines based on Affymetrix SNP arrays. Blue dots represent probes from SNP array, black lines are CBS segmented copy number logratio

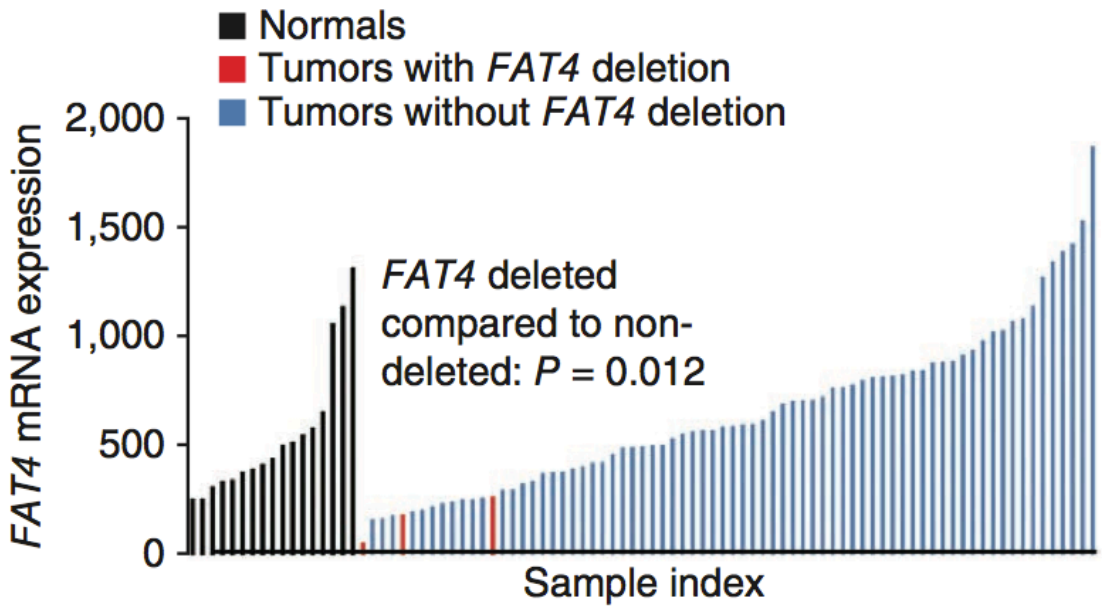


Figure 4-2 *FAT4* deletion in GC is associated with low mRNA expression

Gastric cancers with *FAT4* deletions (red lines) have significantly lower *FAT4* expression compared to tumors without *FAT4* deletions (blue lines). Black lines indicate normal gastric samples.

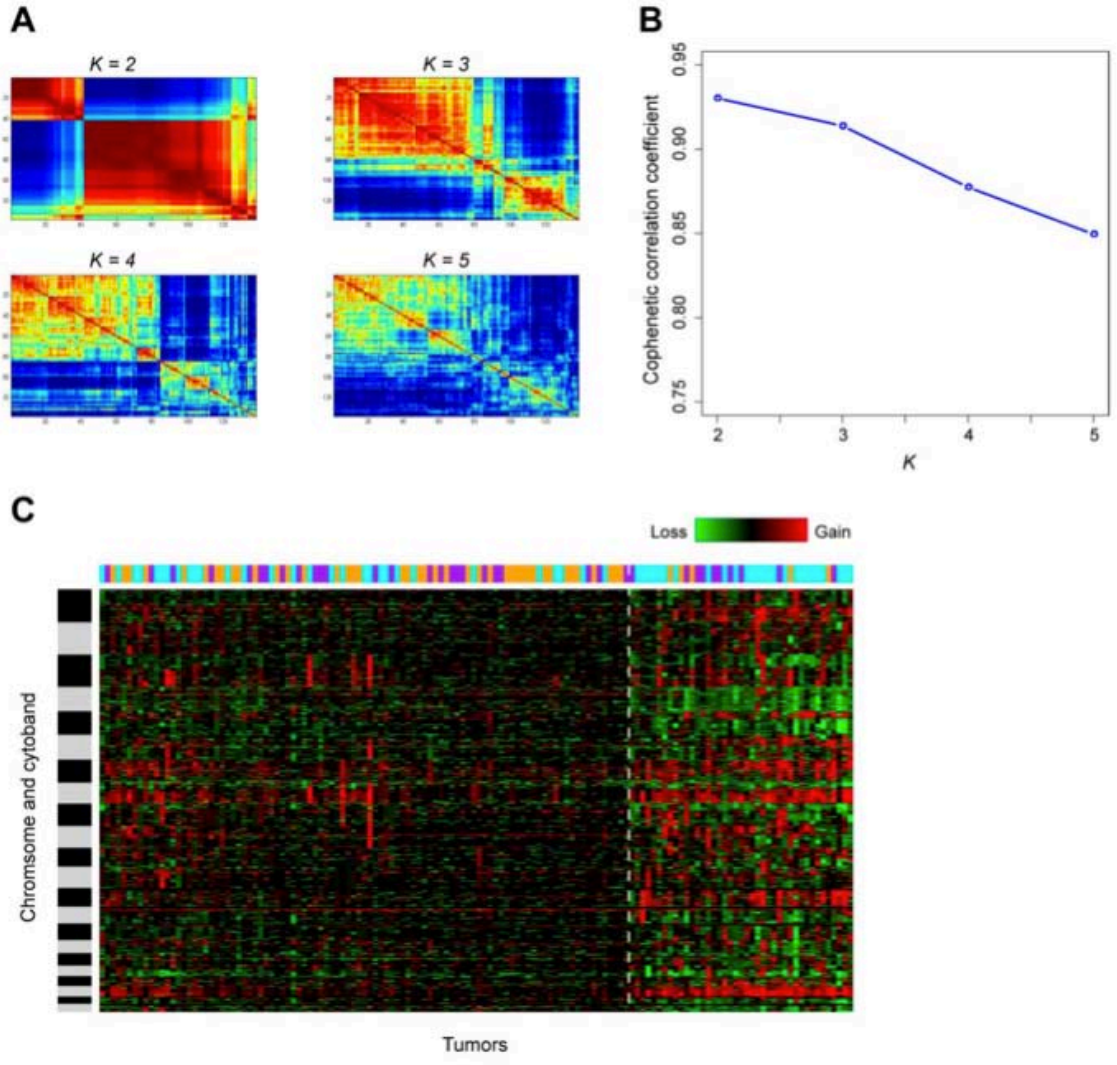
4.2 Integration of Copy Number Landscape with large scale of gene expression data can characterize cancer subtypes

GC, like many other caner types, is a heterogeneous disease. Understanding and characterizing subtypes of GC is important for targeted therapy. In some cases, cancer subtypes could be characterized by a number of few classic oncogenes, such as RTKs, which could be therapeutic targets; in some other cases, cancer subtypes are characterized by a group of genes, ranging from tens to hundreds of genes. In GC, a number of studies have been performed to investigate molecular subtypes. Based on gene expression in a panel of GC cell lines, which mostly consist of cancer cells, GC has been categorized into two subtypes (Tan, Ivanova et al. 2011). This study provides clear clinical value as one of the two subtypes identified is associated with poor prognosis and more sensitive to 5-FU treatment. Recently, another study based on primary gastric samples identified three subtypes from consensus clustering of gene expression profiling (Lei, Tan et al. 2013). The three subtypes are messenchymal, proliferative, metabolic subtypes, named by their representative gene ontologies. In order to understand the genomic different among the three subtypes, we took the 138 overlapped samples with SNP array profiling and performed clustering using non-negative matrix factorization (Brunet, Tamayo et al. 2004; Carrasco, Tonon et al. 2006). Two groups of copy number profiles were identified; with one group has more CNAs at cytobands level and with more extreme CNAs (Fig 4-3). Enrichment analysis suggested that the proliferative harbored more CNAs compared to other two subtypes, while mesenchymal subtype is significant associated with low CNAs (Fig 4-4). We are also interested to know how the three subtypes are associated with focal CNA regions identified from Table 3-1. Consistent with its high level of CNAs, proliferative subtypes is associated with gains of *KRAS*, *ERBB2*, *MYC* and *CCNE1*, loss

of *PTPRD* and *PDE4D*. Overall the copy number landscape helped to further characterize the GC subtypes, and this may imply specific treatment strategies for patients belong to different subtypes.

Figure 4-3 Two clusters of copy number profiles

(A) Consensus matrices of NMF consensus clustering of 138 samples for various numbers of clusters, *K*, from 2 to 5. Dark red indicates a high consensus index while dark blue indicates low consensus index.
(B) The cophenetic correlation coefficients of corresponding matrices (from *K* = 2 to 5). This suggests two subgroups based on copy number profiling.
(C) Heat map of cytoband copy number. The color in each cell indicates the level of copy number gain or loss for one cytoband in one tumor sample. 811 cytobands on chromosomes 1 through 22 are indicated on the y-axis. On the x-axis are the tumor samples ordered by NMF clustering, with a white dashed line separating the CNA low and CNA high groups.



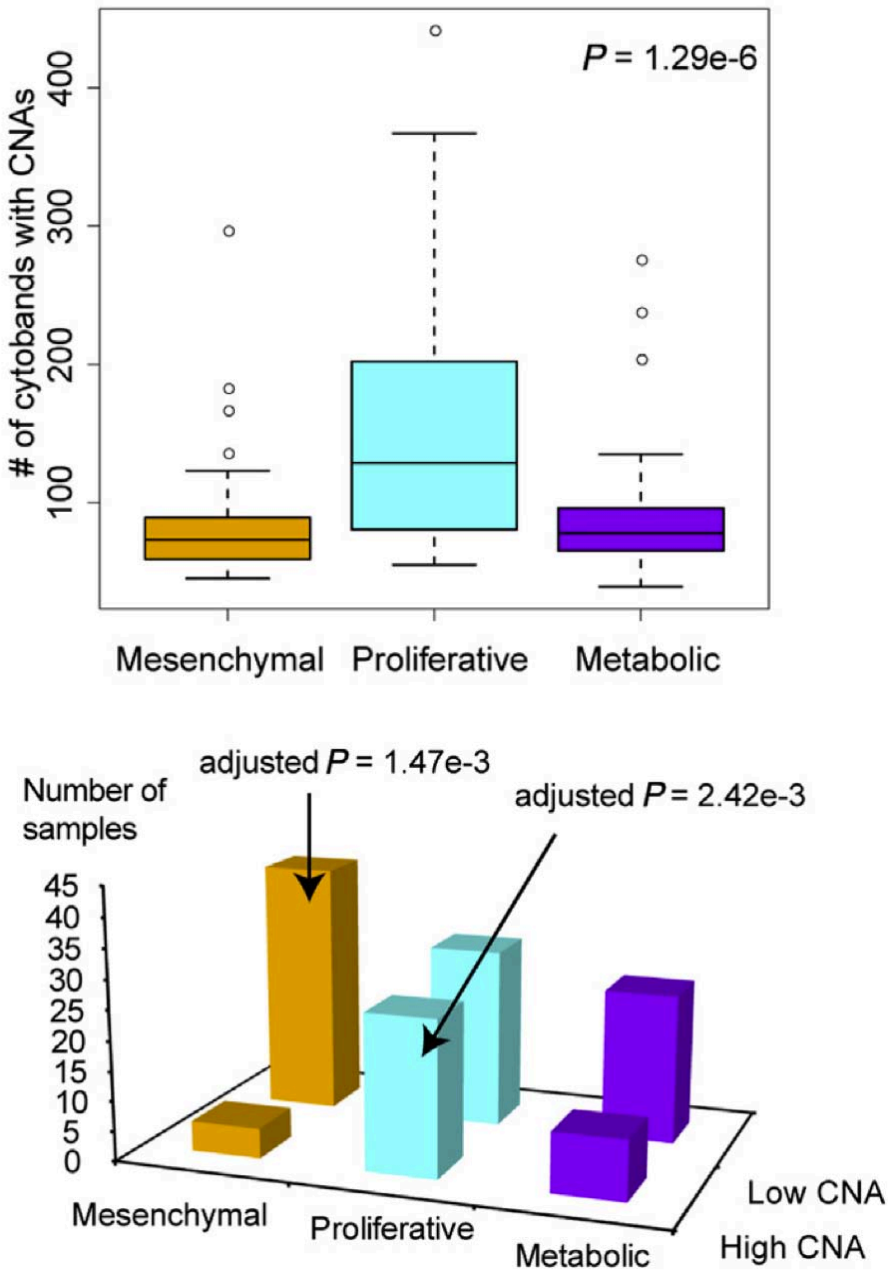


Figure 4-4 Genomic difference among three GC subtypes

(above) The number of cytobands with CNA by subtype (P values by Kruskal–Wallis test).

(bottom) The number of tumors in each subtype in the low or high CNA group (Bonferroni-adjusted P values by hypergeometric tests).

4.3 Integration of CNA to identify its association with long range epigenetic alterations

Epigenetic changes have been suggested to play an important role in carcinogenesis. CG dinucleotide (CpG) methylation is frequently associated with tumor suppressor gene silencing (Feinberg and Vogelstein 1983; Jones and Baylin 2007), and CpG island (CGI) hyper- methylation in gene promoters has been associated with transcriptional silencing in cancer (Jones and Baylin 2007). By using genome-wide methylation microarray, studies in various cancers performed global methylome analysis to investigate epigenetic subtypes (Figueroa, Lugthart et al. 2010; Noushmehr, Weisenberger et al. 2010), methylation associated "CGI" shores (Irizarry, Ladd-Acosta et al. 2009), or long range epigenetic alterations (LRESs) (Coolen, Stirzaker et al. 2010). Our group recently initiated the effort to characterize methylome in GC by illumina 27k methylation arrays in about 300 gastric cases (Zouridis, Deng et al. 2012). This studies has successfully defined two subtype of GC by global CpG island methyaltion status, with one group exhibits high level of hypermethylation, named CIMP (CpG Island Methylation Phenotype) subtype, and another group with relatively high level of hypomethylation (non-CIMP subtype). CIMP subtype was associated with poor survival outcome and could be suggested for methylation inhibitor treatment.

Hypomethylation can induce chromosome breakage events (Cadieux, Ching et al. 2006). To explore the potential of long range hypomethylation in tumor, this study applied a sliding window method to perform a screen across the whole genome to identify 24 candidate regions of long range hypomethylation (hypo-LRRs). These regions tend to predominantly occurred in non-CIMP tumors. With available high-density copy number

data for this same gastric cohort, we are interested to test for any association between the hypo-LRRs and genome instability. Specifically, we tested if tumors exhibiting specific hypo-LRRs were also prone to genome instability in those same regions. Strikingly, we found that the majority of hypo-LRRs were indeed significantly associated with increased genome instability ($p=6.0x10^{-6}$, Figure 4-5). The association between hypo-LRR presence and genome instability was also observed at the level of individual hypo-LRRs (data not shown). While it is generally accepted that tumors exhibiting hypomethylation are associated with increased genome instability, to our knowledge this is the first time that an association between individual regions of hypomethylation and genome instability in those same regions has been demonstrated.

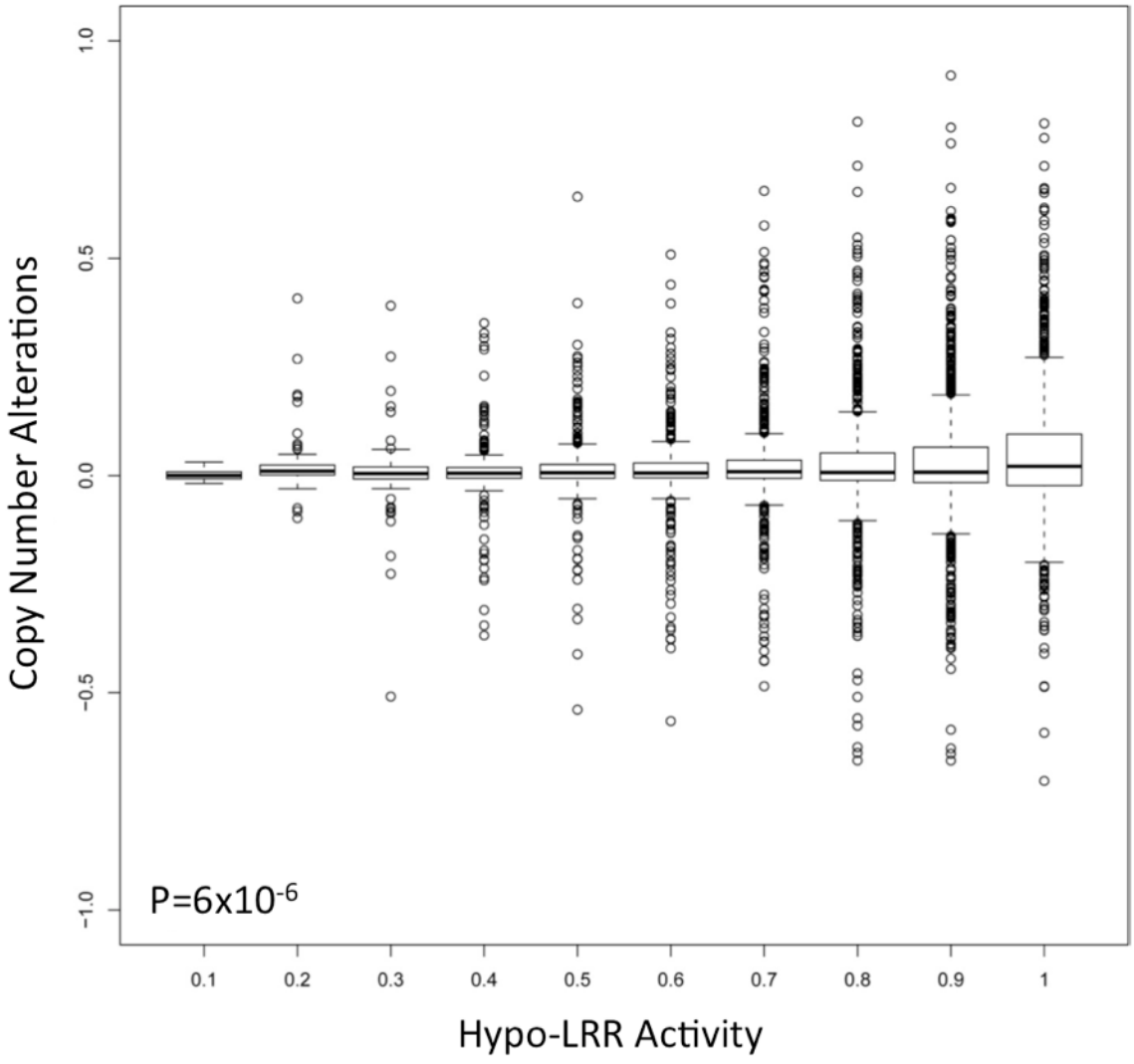


Figure 4-5 CNAs and hypomethylated long-range regions

The boxplot presents data summarized from 24 hypo-LRRs, across 190 gastric cancers with methylation and copy number information. Hypo-LRRs were identified using an "Activity" score similar to that used to define LRESs, but analyzing hypomethylated CpGs rather than hypermethylated CpGs. X-axis : Each dot represents one hypo-LRR in one tumor. X-axis : Level of hypo-activity from 0 to 1, marking increasing hypomethylation from left to right. Y-axis : CNA level, represented by the LRR (log relative ratio). P values were calculated using a correlation test of standard deviation between CNA LRR and hypo-LRR activity.

Chapter 5 Discussion

Here, we report a high-resolution genomic analysis of a large cohort of GC primary tumors and cell lines delineating the most prevalent molecular targets in this disease. While earlier reports analyzing GC copy number alterations have largely analyzed small patient populations or used low-resolution technologies (e.g. chromosomal CGH) (Peng, Sugihara et al. 2003; Tay, Leong et al. 2003; Kimura, Noguchi et al. 2004; Tsukamoto, Uchida et al. 2008; Tada, Kanai et al. 2010), these earlier studies were invaluable in benchmarking the reproducibility of our own data. For example, in a recent copy number analysis of 49 GCs using Agilent 44k arrays(Rossi, Klersy et al. 2011), concordant regions commonly identified in this study and ours include the frequent broad amplifications of chromosome 8 and 20, losses of chromosome 16, and amplified genes such as *ERBB2*, *EGFR*, *GATA4*, *MYC*, *KRAS* and *CCNE1*. However, reflecting the increased size (193 vs 49) and resolution (44K vs 1.8 million SNP probes) of our study, we also detected amplifications of chromosome 18 and deletions of chromosome 6q, which were not detected in earlier work (Peng, Sugihara et al. 2003; Tay, Leong et al. 2003; Kimura, Noguchi et al. 2004; Tsukamoto, Uchida et al. 2008; Tada, Kanai et al. 2010; Rossi, Klersy et al. 2011).

Using GISTIC, we identified 22 recurrently altered regions in GC that are likely to represent the most prevalent molecular targets. For several of these targets, we further confirmed the SNP array results using a variety of orthogonal methodologies, including IHC, FISH, and qPCR. A survey of genes in the 22 altered regions revealed that they could be broadly partitioned into three major functional categories: a) RTK/RAS

signaling (*FGFR2, KRAS, ERBB2, EGFR, MET*), b) transcriptional regulation (*MYC, GATA4, GATA6, KLF5*), and cell cycle control (*CCND1, CCNE1, CDK6, CDKN2A/B, RB*). As expected, many of these genes were already known to be associated with genomic alterations in GC (Hirono, Tsugawa et al. 1995; Tanner, Hollmen et al. 2005; Bizari, Borim et al. 2006; Turner and Grose 2010). Critically however, our analysis also identified several novel genes not previously known to be amplified or deleted in GC. For example, we observed for the first time frequent deletions of *PARK2*, a E3 ubiquitin ligase, in GC (Bedford, Lowe et al. 2011). Mutations in *PARK2* have been associated with early-onset Parkinson's disease (Healy, Falchi et al. 2008), and more recently *PARK2* mutations and deletions have been observed in other cancers (Veeriah, Taylor et al. 2010). Another novel altered GC gene was *CSMD1*, a gene of uncertain function but which has been proposed as a tumor suppressor in breast cancer (Kamal, Shaaban et al. 2010). Using IHC, we confirmed that up to 40% of GCs can exhibit *CSMD1* protein loss or reduced expression. Addressing the functions of these novel altered genes, given their frequency of alteration in GC, will likely be an important goal of future research work. In addition, our study also highlights interesting therapeutic opportunities - for example, the cyclin-dependent kinase *CDK6* was frequently amplified in our series, and small molecule targeted inhibitors of CDKs have been developed (Lapenna and Giordano 2009).

A notable finding in this study was that *GATA4*, *GATA6* and *KLF5* are frequently amplified in GC. Notably, *GATA4* amplifications in GC have also been observed by other groups (Weiss, Kuipers et al. 2004). Intriguingly, when compared against genes identified as amplified in other comparable copy number studies from glioblastoma, lung

cancer and multiple cancer types (Weir, Woo et al. 2007; Chaudhuri, Handcock et al. 2008; Beroukhim, Mermel et al. 2010), it appears that amplification of these three genes appears to be restricted to either GC or to other cancers related of gastrointestinal tract origin. It is possible that these genes may represent “lineage-specific” oncogenes, a recently-described class of cancer genes that enhance oncogenesis by reactivating lineage-specific survival mechanisms normally operative only in early embryonic development (Garraway and Sellers 2006). Examples of lineage survival oncogenes include *MITF* in melanoma, *TITF1/NKX2.1* in lung cancer (Garraway, Widlund et al. 2005; Kwei, Kim et al. 2008), and *SOX2* in esophageal and lung cancers (Bass, Watanabe et al. 2009). Indeed, *GATA6* has recently been proposed to function as an amplified lineage-survival oncogene in pancreatic cancer (Kwei, Bashyam et al. 2008; Alvarez, Opalinska et al. 2011), and *KLF5* has been shown to be expressed during early development in the cardiovascular system and gastrointestinal tract epithelium in the proliferating zone of intestinal crypts (Conkright, Wani et al. 1999; Ohnishi, Laub et al. 2000). These transcription factors may reflect the existence of an underlying transcriptional regulatory program important for the maintenance of the GC phenotype. Interestingly, a recent genomic study from our group reported the discovery of two GC subtypes (G-INT and G-DIF) with distinct gene expression, clinical outcome, and chemotherapy response features (Tan, Ivanova et al. 2011). We have since discovered that G-DIF GCs appears to be significantly enriched in *GATA6* gene amplifications (Fisher exact test, $p = 0.04$), suggesting that *GATA6* may be associated with a specific molecular subtype of GC. From a therapeutic perspective, transcription factors are commonly regarded as "undruggable". It is possible, however, that some of these

transcription factors may regulate the expression of key genes that are pharmacologically targetable. For example, *BCL2* has been described as a target of the *MITF* transcription factor frequently amplified in melanoma (McGill, Horstmann et al. 2002), and *BCL2* inhibitor drugs are available. Such a strategy may represent one method to indirectly target amplified transcription factors.

Of major clinical significance was the observation that genes related to RTK/RAS signaling are a) frequently altered, and b) mutually exclusive to one another in GC. First, because numerous targeted inhibitors directed against various components of the RTK/RAS pathway are already in clinical testing (Bang, Van Cutsem et al. 2010; Yap, Olmos et al. 2011), these results raise the possibility that a substantial proportion (37% of GCs) may be potentially targetable by a RTK/RAS directed therapy. In essence, this finding dramatically increases the population of GC patients for which targeted treatments could be considered. Second, the mutually exclusive nature of these RTK/RAS alterations strongly suggests that the majority of GCs are likely to have only a single RTK/RAS driver oncogene, thereby greatly simplifying the challenge of defining which RTK/RAS targeted inhibitor compound to allocate to which patient population. In terms of clinical trials, the mutually exclusive nature of the RTK/RAS alterations also renders it technically feasible to implement a multi-biomarker based trial (Printz 2010), where multiple targeted compounds are tested in different biomarker-defined populations within a single trial design, as has been recently described for non-small-cell-lung cancer (BATTLE trial, (Kim, Herbst et al. 2011)). Third, these results suggest that a much larger proportions of GCs may be reliant on RTK/RAS signaling than previously appreciated, particularly if one notes that in this study alternative mechanisms of RTK/RAS activation

were not considered, and for certain GCs the presence of non-malignant cells may have reduced the sensitivity of RTK/RAS alteration detection. For example, in a recent kinome sequencing study, kinases related to MAPK signaling, a pathway downstream of *KRAS*, were identified as being the most significantly altered in GC (Zang, Ong et al. 2011). Another alternative mechanism of RTK/RAS activation may also involve gene fusions, where we recently described RAF-related gene rearrangements in GC (Palanisamy, Ateeq et al. 2010). Taken collectively, we believe that our finding that 37% of GCs exhibit a RTK/RAS alteration should best be regarded as a lower limit, and are consistent with the notion that RTK/RAS signaling is a dominant oncogenic pathway in GC.

In our series, *FGFR2* was amplified at frequencies comparable to *ERBB2*, providing one of the first assessments of *FGFR2* gene amplification in primary GCs. Interestingly, the smallest common peak of *FGFR2* amplification in the GCs appears to center around a 1.5 kb region in *FGFR2* intron 2 which overlaps a SNP locus associated with breast cancer susceptibility (Hunter, Kraft et al. 2007). It is intriguing to consider if the process of genomic amplification might also bias the expression of *FGFR2* gene towards transcript isoforms (IIIc) that are pro-oncogenic (Katoh and Katoh 2009). We also found that that in preclinical assays, dovitnib, a *VEGFR/FGFR2* inhibitor, can potently inhibit the growth of *FGFR2*-amplified GC cell lines and xenografts. In breast cancer, dovitinib has been found to exert effects primarily in *FGFR1*-amplified breast cancers, suggesting the importance of *FGFR*-related genome amplification in predicting dovitinib response (Andre, Bachelot et al. 2011). *FGFR2* is thus likely to represent an attractive therapeutic target in GC. However, one question not addressed by our data is whether GCs which

lack *FGFR2* amplification, but nevertheless express *FGFR2*, will also be dovitnib-responsive, since we also observed that a significant number of *FGFR2*-copy neutral tumors also exhibited elevated *FGFR2* expression levels relative to matched normal tissues, indicating that other mechanisms besides gene amplification can also cause *FGFR2* upregulation in tumors. Notably, recent study showed that *FGFR2* inhibition can potentially reverse chemo-resistance in OCUM-2M GC cells, which are also *FGFR2*-copy number amplified (Qiu, Yashiro et al. 2011). We are currently addressing these questions by conducting a biopsy-mandated Phase I/II trial at our center, evaluating the efficacy of dovitinib in *FGFR2*-amplified and *FGFR2* expressing GC samples.

Finally, our results highlight *KRAS* amplification (rather than *KRAS* mutation) as a prevalent event in GC. While *KRAS* amplifications have been reported in other cancers (e.g. lung) (Wagner, Stiedl et al. 2011), these observations have been largely anecdotal, with emphasis directed towards more conventional codon 12 and 13 activating mutations. Consistent with *KRAS* activating as an important driver gene in amplified samples, patients in our series with *KRAS* amplified GCs exhibited poor prognosis, and *in vitro*, *KRAS* amplified GC lines were sensitive to *KRAS* silencing, similar to *KRAS* mutated lines. The high frequency of *KRAS* amplifications in GC is likely a major reason as to why *KRAS* activating mutations are strikingly infrequent in GC (Mita, Toyota et al. 2009). However, the exact mechanisms underlying this striking tissue-specific preference for *KRAS* amplification remain to be elucidated. Nevertheless, given recent data demonstrating that *KRAS*-mutated colon cancers are resistant to anti-*EGFR* therapies (Van Cutsem, Kohne et al. 2009), and that *KRAS*-amplified tumors may be resistant to MEK1/2 inhibitors (Little, Balmanno et al. 2011), our findings strongly suggest that

testing *KRAS* amplification status in tumors should be fully considered in any trials evaluating RTK targeting compounds in GC.

In conclusion, our results provide for the first time a detailed molecular map of genomic alterations in GC, which has revealed several promising targets for subtype-specific therapies. Classifying GC patients by these signature genomic alterations may facilitate patient allocations to the most appropriate clinical trials, thereby maximizing patient participation in combating this lethal disease.

Chapter 6 References

(2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.

(2008). "Comprehensive genomic characterization defines human glioblastoma genes and core pathways." Nature **455**(7216): 1061-1068.

(2011). "Integrated genomic analyses of ovarian carcinoma." Nature **474**(7353): 609-615.

(2012). "Comprehensive molecular characterization of human colon and rectal cancer." Nature **487**(7407): 330-337.

(2013). "Comprehensive molecular characterization of clear cell renal cell carcinoma." Nature **499**(7456): 43-49.

Alimonti, A., A. Carracedo, et al. (2010). "Subtle variations in Pten dose determine cancer susceptibility." Nat Genet **42**(5): 454-458.

Alvarez, H., J. Opalinska, et al. (2011). "Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis." PLoS Genet **7**(3): e1001356.

Andre, F., T. D. Bachelot, et al. (2011). "A multicenter, open-label phase II trial of dovitinib, an FGFR1 inhibitor, in FGFR1 amplified and non-amplified metastatic breast cancer." J Clin Oncol **29**: (suppl; abstr 508).

Aqeilan, R. I., T. Kuroki, et al. (2004). "Loss of WWOX expression in gastric carcinoma." Clin Cancer Res **10**(9): 3053-3058.

Asaoka, Y., T. Ikenoue, et al. (2011). "New targeted therapies for gastric cancer." Expert Opin Investig Drugs **20**(5): 595-604.

Bang, Y. J., E. Van Cutsem, et al. (2010). "Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial." Lancet **376**(9742): 687-697.

Bass, A. J., H. Watanabe, et al. (2009). "SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas." Nat Genet **41**(11): 1238-1242.

Bean, J., C. Brennan, et al. (2007). "MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib." Proc Natl Acad Sci U S A **104**(52): 20932-20937.

Bedford, L., J. Lowe, et al. (2011). "Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets." Nat Rev Drug Discov **10**(1): 29-46.

Berns, K., H. M. Horlings, et al. (2007). "A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer." Cancer Cell **12**(4): 395-402.

Beroukhim, R., G. Getz, et al. (2007). "Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma." Proc Natl Acad Sci U S A **104**(50): 20007-20012.

Beroukhim, R., C. H. Mermel, et al. (2010). "The landscape of somatic copy-number alteration across human cancers." Nature **463**(7283): 899-905.

Berx, G. and F. van Roy (2009). "Involvement of members of the cadherin superfamily in cancer." Cold Spring Harb Perspect Biol **1**(6): a003129.

Bizari, L., A. A. Borim, et al. (2006). "Alterations of the CCND1 and HER-2/neu (ERBB2) proteins in esophageal and gastric cancers." Cancer Genet Cytogenet **165**(1): 41-50.

Brenner, H., D. Rothenbacher, et al. (2009). "Epidemiology of stomach cancer." Methods Mol Biol **472**: 467-477.
Brunet, J. P., P. Tamayo, et al. (2004). "Metagenes and molecular pattern discovery using matrix factorization." Proc Natl Acad Sci U S A **101**(12): 4164-4169.
Cadieux, B., T. T. Ching, et al. (2006). "Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation." Cancer Res **66**(17): 8469-8476.
Cappuzzo, F., F. R. Hirsch, et al. (2005). "Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer." J Natl Cancer Inst **97**(9): 643-655.
Carrasco, D. R., G. Tonon, et al. (2006). "High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients." Cancer Cell **9**(4): 313-325.
Catalano, V., R. Labianca, et al. (2009). "Gastric cancer." Crit Rev Oncol Hematol **71**(2): 127-164.
Chaudhuri, S., M. S. Handcock, et al. (2008). "Generalised Linear Models Incorporating Population Level Information: An Empirical Likelihood Based Approach." J R Stat Soc Series B Stat Methodol **70**(2): 311-328.
Conkright, M. D., M. A. Wani, et al. (1999). "A gene encoding an intestinal-enriched member of the Kruppel-like factor family expressed in intestinal epithelial cells." Nucleic Acids Res **27**(5): 1263-1270.
Coolen, M. W., C. Stirzaker, et al. (2010). "Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity." Nat Cell Biol **12**(3): 235-246.
Dey, J. H., F. Bianchi, et al. (2010). "Targeting fibroblast growth factor receptors blocks PI3K/AKT signaling, induces apoptosis, and impairs mammary tumor outgrowth and metastasis." Cancer Res **70**(10): 4151-4162.
Ding, L., G. Getz, et al. (2008). "Somatic mutations affect key pathways in lung adenocarcinoma." Nature **455**(7216): 1069-1075.
Feinberg, A. P. and B. Vogelstein (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." Nature **301**(5895): 89-92.
Ferlay, J., H. R. Shin, et al. (2010). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008." Int J Cancer **127**(12): 2893-2917.
Feuk, L., A. R. Carson, et al. (2006). "Structural variation in the human genome." Nat Rev Genet **7**(2): 85-97.
Figueroa, M. E., S. Lugthart, et al. (2010). "DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia." Cancer Cell **17**(1): 13-27.
Frazer, K. A., D. G. Ballinger, et al. (2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-861.
Fukuda, Y., N. Kurihara, et al. (2000). "CD44 is a potential target of amplification within the 11p13 amplicon detected in gastric cancer cell lines." Genes Chromosomes Cancer **29**(4): 315-324.
Garraway, L. A. and W. R. Sellers (2006). "Lineage dependency and lineage-survival oncogenes in human cancer." Nat Rev Cancer **6**(8): 593-602.
Garraway, L. A., H. R. Widlund, et al. (2005). "Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma." Nature **436**(7047): 117-122.
Girirajan, S., C. D. Campbell, et al. (2011). "Human copy number variation and complex genetic disease." Annu Rev Genet **45**: 203-226.

Gonzalez, C. A. and A. Agudo (2012). "Carcinogenesis, prevention and early detection of gastric cancer: where we are and where we should go." Int J Cancer **130**(4): 745-753.
Gravalos, C. and A. Jimeno (2008). "HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target." Ann Oncol **19**(9): 1523-1529.
Gunther, T., R. Schneider-Stock, et al. (2000). "Mdm2 gene amplification in gastric cancer correlation with expression of Mdm2 protein and p53 alterations." Mod Pathol **13**(6): 621-626.
Hartgrink, H. H., E. P. Jansen, et al. (2009). "Gastric cancer." Lancet **374**(9688): 477-490.
He, X. S., Q. Su, et al. (2001). "Expression, deletion [was deleton] and mutation of p16 gene in human gastric cancer." World J Gastroenterol **7**(4): 515-521.
Healy, D. G., M. Falchi, et al. (2008). "Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study." Lancet Neurol **7**(7): 583-590.
Hirono, Y., K. Tsugawa, et al. (1995). "Amplification of epidermal growth factor receptor gene and its relationship to survival in human gastric cancer." Oncology **52**(3): 182-188.
Hirschhorn, J. N. and M. J. Daly (2005). "Genome-wide association studies for common diseases and complex traits." Nat Rev Genet **6**(2): 95-108.
Hofmann, M., O. Stoss, et al. (2008). "Assessment of a HER2 scoring system for gastric cancer: results from a validation study." Histopathology **52**(7): 797-805.
Hunter, D. J., P. Kraft, et al. (2007). "A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer." Nat Genet **39**(7): 870-874.
Iafrate, A. J., L. Feuk, et al. (2004). "Detection of large-scale variation in the human genome." Nat Genet **36**(9): 949-951.
Irizarry, R. A., C. Ladd-Acosta, et al. (2009). "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores." Nat Genet **41**(2): 178-186.
Johnson, W. E., C. Li, et al. (2007). "Adjusting batch effects in microarray expression data using empirical Bayes methods." Biostatistics **8**(1): 118-127.
Jones, P. A. and S. B. Baylin (2007). "The epigenomics of cancer." Cell **128**(4): 683-692.
Kallioniemi, A., O. P. Kallioniemi, et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors." Science **258**(5083): 818-821.
Kamal, M., A. M. Shaaban, et al. (2010). "Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma." Breast Cancer Res Treat **121**(3): 555-563.
Kamangar, F., G. M. Dores, et al. (2006). "Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world." J Clin Oncol **24**(14): 2137-2150.
Kandoth, C., N. Schultz, et al. (2013). "Integrated genomic characterization of endometrial carcinoma." Nature **497**(7447): 67-73.
Katoh, Y. and M. Katoh (2009). "FGFR2-related pathogenesis and FGFR2-targeted therapeutics (Review)." Int J Mol Med **23**(3): 307-311.
Kim, E. S., R. S. Herbst, et al. (2011). "The BALLTE Trial: Personalizing Therapy for Lung Cancer." Cancer Discovery **1**(1): 44-53.
Kimura, Y., T. Noguchi, et al. (2004). "Genetic alterations in 102 primary gastric cancers by comparative genomic hybridization: gain of 20q and loss of 18q are associated with tumor progression." Mod Pathol **17**(11): 1328-1337.

Korbel, J. O., A. E. Urban, et al. (2007). "Paired-end mapping reveals extensive structural variation in the human genome." Science **318**(5849): 420-426.
Korenberg, J. R., X. N. Chen, et al. (1994). "Down syndrome phenotypes: the consequences of chromosomal imbalance." Proc Natl Acad Sci U S A **91**(11): 4997-5001.
Kruse, J. P. and W. Gu (2009). "Modes of p53 regulation." Cell **137**(4): 609-622.
Krzywinski, M., J. Schein, et al. (2009). "Circos: an information aesthetic for comparative genomics." Genome Res **19**(9): 1639-1645.
Kunii, K., L. Davis, et al. (2008). "FGFR2-amplified gastric cancer cell lines require FGFR2 and Erbb3 signaling for growth and survival." Cancer Res **68**(7): 2340-2348.
Kwei, K. A., M. D. Bashyam, et al. (2008). "Genomic profiling identifies GATA6 as a candidate oncogene amplified in pancreatobiliary cancer." PLoS Genet **4**(5): e1000081.
Kwei, K. A., Y. H. Kim, et al. (2008). "Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer." Oncogene **27**(25): 3635-3640.
LaFramboise, T. (2009). "Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances." Nucleic Acids Res **37**(13): 4181-4193.
Lamlum, H., M. Ilyas, et al. (1999). "The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis." Nat Med **5**(9): 1071-1075.
Langer-Safer, P. R., M. Levine, et al. (1982). "Immunological method for mapping genes on Drosophila polytene chromosomes." Proc Natl Acad Sci U S A **79**(14): 4381-4385.
Lapenna, S. and A. Giordano (2009). "Cell cycle kinases as therapeutic targets for cancer." Nat Rev Drug Discov **8**(7): 547-566.
Lee, S. H., D. Lopes de Menezes, et al. (2005). "In vivo target modulation and biological activity of CHIR-258, a multitargeted growth factor receptor kinase inhibitor, in colon cancer models." Clin Cancer Res **11**(10): 3633-3641.
Lee, T. L., W. K. Leung, et al. (2002). "Detection of gene promoter hypermethylation in the tumor and serum of patients with gastric carcinoma." Clin Cancer Res **8**(6): 1761-1766.
Lei, Z., I. B. Tan, et al. (2013). "Identification of Molecular Subtypes of Gastric Cancer With Different Responses to PI3-Kinase Inhibitors and 5-Fluorouracil." Gastroenterology.
Leung, S. Y., C. Ho, et al. (2006). "Comprehensive analysis of 19q12 amplicon in human gastric cancers." Mod Pathol **19**(6): 854-863.
Lievre, A., J. B. Bachet, et al. (2006). "KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer." Cancer Res **66**(8): 3992-3995.
Lim, G. H., C. S. Wong, et al. (2009). "Trends in long-term cancer survival in Singapore: 1968-2002." Ann Acad Med Singapore **38**(2): 99-105.
Little, A. S., K. Balmanno, et al. (2011). "Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells." Sci Signal **4**(166): ra17.
Lupski, J. R. (2007). "Genomic rearrangements and sporadic disease." Nat Genet **39**(7 Suppl): S43-47.
MacDonald, J. S., P. S. Schein, et al. (1980). "5-Fluorouracil, doxorubicin, and mitomycin (FAM) combination chemotherapy for advanced gastric cancer." Ann Intern Med **93**(4): 533-536.
Mao, Y., J. Mulvaney, et al. (2011). "Characterization of a Dchs1 mutant mouse reveals requirements for Dchs1-Fat4 signaling during mammalian development." Development **138**(5): 947-957.

McCarroll, S. A., F. G. Kuruvilla, et al. (2008). "Integrated detection and population-genetic analysis of SNPs and copy number variation." Nat Genet **40**(10): 1166-1174.
McCarthy, M. I., G. R. Abecasis, et al. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges." Nat Rev Genet **9**(5): 356-369.
McGill, G. G., M. Horstmann, et al. (2002). "Bcl2 regulation by the melanocyte master regulator Mitf modulates lineage survival and melanoma cell viability." Cell **109**(6): 707-718.
Mills, R. E., K. Walter, et al. (2011). "Mapping copy number variation by population-scale genome sequencing." Nature **470**(7332): 59-65.
Mita, H., M. Toyota, et al. (2009). "A novel method, digital genome scanning detects KRAS gene amplification in gastric cancers: involvement of overexpressed wild-type KRAS in downstream signaling and cancer cell growth." BMC Cancer **9**: 198.
Mootha, V. K., C. M. Lindgren, et al. (2003). "PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes." Nat Genet **34**(3): 267-273.
Moser, C., S. A. Lang, et al. (2009). "Heat-shock protein 90 (Hsp90) as a molecular target for therapy of gastrointestinal cancer." Anticancer Res **29**(6): 2031-2042.
Myllykangas, S., S. Junnila, et al. (2008). "Integrated gene copy number and expression microarray analysis of gastric cancer highlights potential target genes." Int J Cancer **123**(4): 817-825.
Nakamura, Y., T. Migita, et al. (2009). "Kruppel-like factor 12 plays a significant role in poorly differentiated gastric cancer progression." Int J Cancer **125**(8): 1859-1867.
Noushmehr, H., D. J. Weisenberger, et al. (2010). "Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma." Cancer Cell **17**(5): 510-522.
Ocio, E. M., M. V. Mateos, et al. (2008). "New drugs in multiple myeloma: mechanisms of action and phase I/II clinical findings." Lancet Oncol **9**(12): 1157-1165.
Ohnishi, S., F. Laub, et al. (2000). "Developmental expression of the mouse gene coding for the Kruppel-like transcription factor KLF5." Dev Dyn **217**(4): 421-429.
Okudela, K., M. Suzuki, et al. (2007). "PIK3CA mutation and amplification in human lung cancer." Pathol Int **57**(10): 664-671.
Olshen, A. B., E. S. Venkatraman, et al. (2004). "Circular binary segmentation for the analysis of array-based DNA copy number data." Biostatistics **5**(4): 557-572.
Ooi, C. H., T. Ivanova, et al. (2009). "Oncogenic pathway combinations predict clinical prognosis in gastric cancer." PLoS Genet **5**(10): e1000676.
Palanisamy, N., B. Ateeq, et al. (2010). "Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma." Nat Med **16**(7): 793-798.
Peng, D. F., H. Sugihara, et al. (2003). "Alterations of chromosomal copy number during progression of diffuse-type gastric carcinomas: metaphase- and array-based comparative genomic hybridization analyses of multiple samples from individual tumours." J Pathol **201**(3): 439-450.
Pinto, D., A. T. Pagnamenta, et al. (2010). "Functional impact of global rare copy number variation in autism spectrum disorders." Nature **466**(7304): 368-372.
Printz, C. (2010). "BATTLE to personalize lung cancer treatment. Novel clinical trial design and tissue gathering procedures drive biomarker discovery." Cancer **116**(14): 3307-3308.
Qiu, H., M. Yashiro, et al. (2011). "A FGFR2 inhibitor, Ki23057, enhances the chemosensitivity of drug-resistant gastric cancer cells." Cancer Lett **307**(1): 47-52.
Raimondi, S. C. (2000). "Fluorescence in situ hybridization: molecular probes for diagnosis of pediatric neoplastic diseases." Cancer Invest **18**(2): 135-147.

Rajagopalan, H., A. Bardelli, et al. (2002). "Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status." Nature **418**(6901): 934.
Rossi, E., C. Klersy, et al. (2011). "Correlation between genomic alterations assessed by array comparative genomic hybridization, prognostically informative histologic subtype, stage, and patient survival in gastric cancer." Hum Pathol **42**(12): 1937-1945.
Saburi, S., I. Hester, et al. (2008). "Loss of Fat4 disrupts PCP signaling and oriented cell division and leads to cystic kidney disease." Nat Genet **40**(8): 1010-1015.
Sanchez-Perez, I., P. Garcia Alonso, et al. (2009). "Clinical impact of aneuploidy on gastric cancer patients." Clin Transl Oncol **11**(8): 493-498.
Sarker, D., R. Molife, et al. (2008). "A phase I pharmacokinetic and pharmacodynamic study of TKI258, an oral, multitargeted receptor tyrosine kinase inhibitor in patients with advanced solid tumors." Clin Cancer Res **14**(7): 2075-2081.
Scaltriti, M., P. J. Eichhorn, et al. (2011). "Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2+ breast cancer patients." Proc Natl Acad Sci U S A **108**(9): 3761-3766.
Schneider, B. G., D. R. Pulitzer, et al. (1995). "Allelic imbalance in gastric cancer: an affected site on chromosome arm 3p." Genes Chromosomes Cancer **13**(4): 263-271.
Sebat, J., B. Lakshmi, et al. (2007). "Strong association of de novo copy number mutations with autism." Science **316**(5823): 445-449.
Sebat, J., B. Lakshmi, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-528.
Subramanian, A., P. Tamayo, et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proc Natl Acad Sci U S A **102**(43): 15545-15550.
Tabin, C. J., S. M. Bradley, et al. (1982). "Mechanism of activation of a human oncogene." Nature **300**(5888): 143-149.
Tada, M., F. Kanai, et al. (2010). "Prognostic significance of genetic alterations detected by high-density single nucleotide polymorphism array in gastric cancer." Cancer Sci **101**(5): 1261-1269.
Takeda, M., T. Arao, et al. (2007). "AZD2171 shows potent antitumor activity against gastric cancer over-expressing fibroblast growth factor receptor 2/keratinocyte growth factor receptor." Clin Cancer Res **13**(10): 3051-3057.
Tamura, G., T. Kihana, et al. (1991). "Detection of frequent p53 gene mutations in primary gastric cancer by cell sorting and polymerase chain reaction single-strand conformation polymorphism analysis." Cancer Res **51**(11): 3056-3058.
Tan, I. B., T. Ivanova, et al. (2011). "Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy." Gastroenterology **141**(2): 476-485, 485 e471-411.
Tang, Y. C. and A. Amon (2013). "Gene copy-number alterations: a cost-benefit analysis." Cell **152**(3): 394-405.
Tanner, M., M. Hollmen, et al. (2005). "Amplification of HER-2 in gastric carcinoma: association with Topoisomerase IIalpha gene amplification, intestinal type, poor prognosis and sensitivity to trastuzumab." Ann Oncol **16**(2): 273-278.
Tao, J., N. T. Deng, et al. (2011). "CD44-SLC1A2 gene fusions in gastric cancer." Sci Transl Med **3**(77): 77ra30.

Tay, S. T., S. H. Leong, et al. (2003). "A combined comparative genomic hybridization and expression microarray analysis of gastric cancer reveals novel molecular subtypes." Cancer Res **63**(12): 3309-3316.
Trudel, S., Z. H. Li, et al. (2005). "CHIR-258, a novel, multitargeted tyrosine kinase inhibitor for the potential treatment of t(4;14) multiple myeloma." Blood **105**(7): 2941-2948.
Tsukamoto, Y., T. Uchida, et al. (2008). "Genome-wide analysis of DNA copy number alterations and gene expression in gastric cancer." J Pathol **216**(4): 471-482.
Turner, N. and R. Grose (2010). "Fibroblast growth factor signalling: from development to cancer." Nat Rev Cancer **10**(2): 116-129.
Urban, A. E., J. O. Korbel, et al. (2006). "High-resolution mapping of DNA copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays." Proc Natl Acad Sci U S A **103**(12): 4534-4539.
Van Cutsem, E., C. H. Kohne, et al. (2009). "Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer." N Engl J Med **360**(14): 1408-1417.
Veeriah, S., B. S. Taylor, et al. (2010). "Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies." Nat Genet **42**(1): 77-82.
Verhaak, R. G., K. A. Hoadley, et al. (2010). "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1." Cancer Cell **17**(1): 98-110.
Wagner, P. L., A. C. Stiedl, et al. (2011). "Frequency and clinicopathologic correlates of KRAS amplification in non-small cell lung carcinoma." Lung Cancer.
Wang, K., J. Kan, et al. (2011). "Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer." Nat Genet **43**(12): 1219-1223.
Wang, Y. (2009). "Wnt/Planar cell polarity signaling: a new paradigm for cancer therapy." Mol Cancer Ther **8**(8): 2103-2109.
Weichert, W., A. Roske, et al. (2008). "Association of patterns of class I histone deacetylase expression with patient prognosis in gastric cancer: a retrospective analysis." Lancet Oncol **9**(2): 139-148.
Weir, B. A., M. S. Woo, et al. (2007). "Characterizing the cancer genome in lung adenocarcinoma." Nature **450**(7171): 893-898.
Weiss, M. M., E. J. Kuipers, et al. (2004). "Genomic alterations in primary gastric adenocarcinomas correlate with clinicopathological characteristics and survival." Cell Oncol **26**(5-6): 307-317.
Wohrer, S. S., M. Raderer, et al. (2004). "Palliative chemotherapy for advanced gastric cancer." Ann Oncol **15**(11): 1585-1595.
Xiao, Y. P., D. Y. Wu, et al. (2006). "Loss of heterozygosity and microsatellite instabilities of fragile histidine triad gene in gastric carcinoma." World J Gastroenterol **12**(23): 3766-3769.
Yap, T. A., D. Olmos, et al. (2011). "Phase I trial of a selective c-MET inhibitor ARQ 197 incorporating proof of mechanism pharmacodynamic studies." J Clin Oncol **29**(10): 1271-1279.
Zang, Z. J., I. Cutcutache, et al. (2012). "Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes." Nat Genet **44**(5): 570-574.
Zang, Z. J., C. K. Ong, et al. (2011). "Genetic and structural variation in the gastric cancer kinome revealed through targeted deep sequencing." Cancer Res **71**(1): 29-39.

Zouridis, H., N. Deng, et al. (2012). "Methylation subtypes and large-scale epigenetic alterations in gastric cancer." Sci Transl Med **4**(156): 156ra140.

Chapter 7 Appendix

Figure A1: 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.

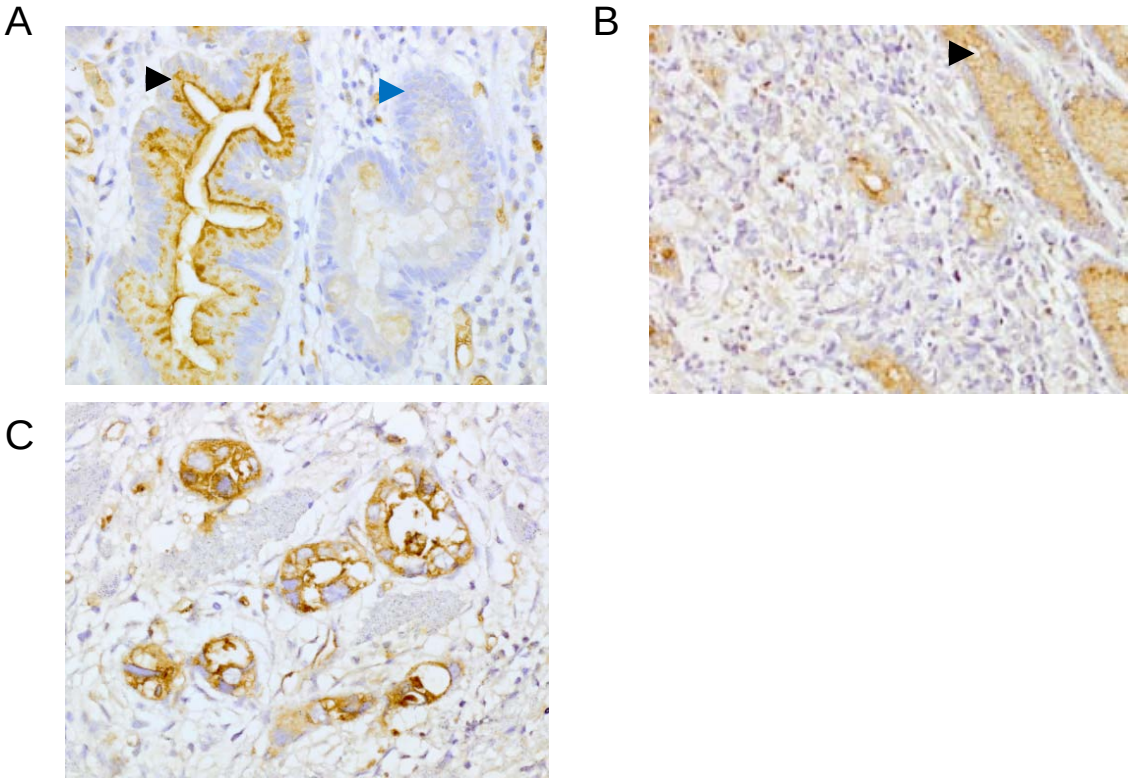


Figure A2: Phenotypic Effects of KRAS siRNA Knockdown in *KRAS*-amplified, Mutated and Wild-type GC Lines

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.

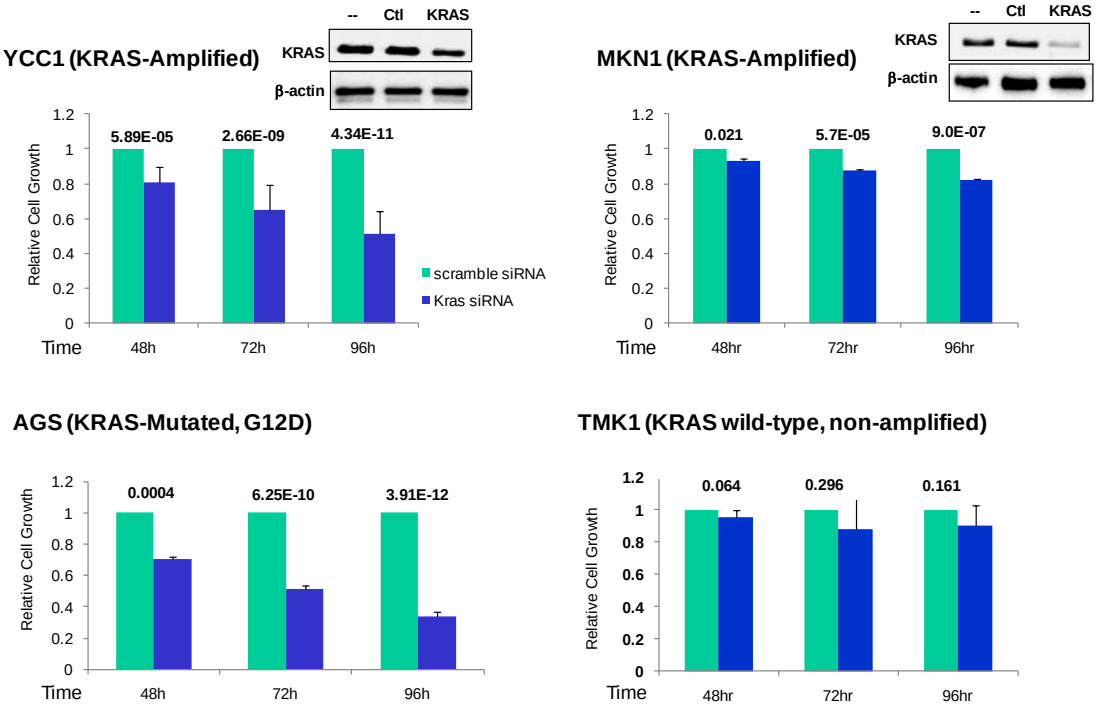


Figure A3. Sensitivity of *FGFR2*-Amplified GC Cell Lines to Dovitinib

A) (top) *FGFR2* RT-PCR analysis of GC cell lines. GAPDH was used as a loading control. (bottom) *FGFR2* protein expression in lines. β-actin was used as a loading control. Cell lines KATOIII and SNU16 are observed to express elevated levels of *FGFR2* mRNA and protein.

B) Cell proliferation effects of dovitinib treatment. Dovitinib GI50 values for FGFR2-amplified and non-amplified cell lines. GI50 – drug concentrated required to cause 50% growth inhibition. GI50 values were calculated after 48 hrs dovitinib treatment. ***** $p<0.05$ compared to non-amplified lines. Results are a mean of three independent experiments.

C) Molecular effects of dovitinib treatment. Cells treated with dovitinib at 50nM, 100nM and 500nM contrations for 1 h. Lysates were immunoprecipitated with *FGFR2* anitbody MAB6841, and probed with 4G10 (phosphotyrosine detection) or MAB6841 for total *FGFR2*. Other antibodies included total and phospho-ERK, and total and phospho-AKT. Experiments were repeated a minimum of three independent times.

D) Dovitinib inhibits soft agar colony formation. *FGFR2*-amplified cells were treated with dovitinib at the GI50 concentration for each cell line (KATO-III: 0.12uM; SNU-16: 0.17uM) for 48 hrs, and soft-agar colony formation monitored over the subsequent 3-4 weeks. Data for KATO-III cells are provided, including representative colony plates. Similar results were observed for SNU16 (Figure A4).

E) Dovitinib induces caspase-3 activation. *FGFR2*-amplified cells were treated with increasing dovitinib concentrations, and apoptosis levels measured after 24hr using Caspase-Glo 3/7 assays. The y-axis represents % of activation normalized against untreated controls. The results are a mean of triplicates $\pm$ standard deviation. Experiments were repeated three independent times.

F) Dovitinib inhibits tumor growth in a human primary GC xenograft model bearing *FGFR2* gene amplification. The mean tumor size of the vehicle treated mice reached 1163 mm^3 at day 25 post treatment. Treatment with the positive control drug 5-FU at 20 mg/kg (Qd x 5/wk x 2 wks, i.p.) produced a mean tumor size of 518 mm^3 (TGI = 63%, p = 0. 08) at the same time. Treatment with dovitinib at 30 mg/kg and 50 mg/kg (Qd x 25 days, *p.o.*) significantly inhibited tumor growth compared to vehicle treated animals, with a mean tumor size of 194 and 53 mm^3, respectively (p = 0.006 and 0.002, respectively at day 25 post treatment.

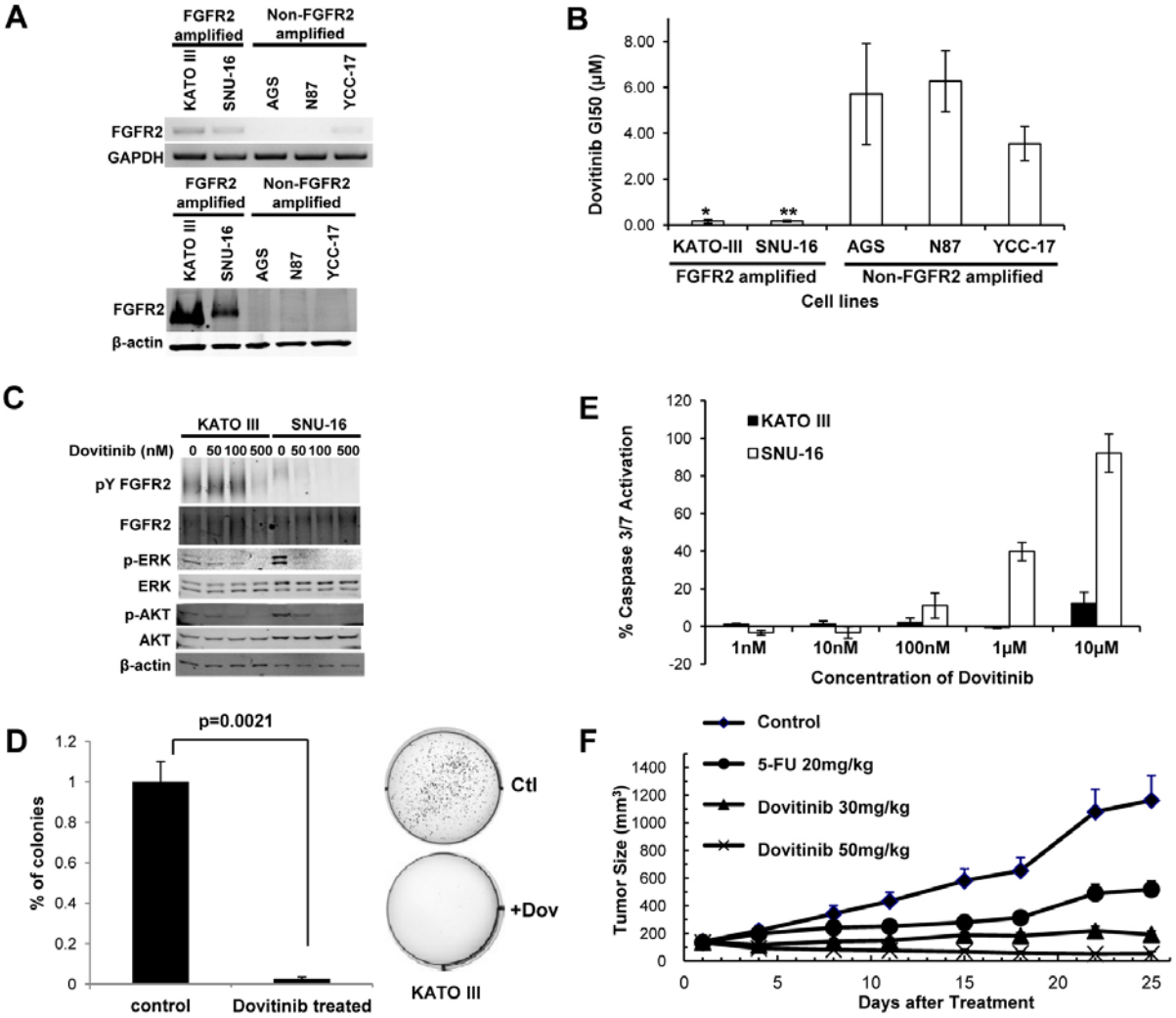

A
FGFR2 amplified
Non-FGFR2 amplified
KATO III
SNU-16
AGS
N87
YCC-17
FGFR2
GAPDH
β-actin
B
Dovitinib GI50 (μM)
8.00
6.00
4.00
2.00
0.00
*
**
KATO-III
SNU-16
AGS
N87
YCC-17
FGFR2 amplified
Non-FGFR2 amplified
Cell lines
C
KATO III
SNU-16
Dovitinib (nM)
0 50 100 500 0 50 100 500
pY FGFR2
FGFR2
p-ERK
ERK
p-AKT
AKT
β-actin
D
p=0.0021
% of colonies
1.2
1
0.8
0.6
0.4
0.2
0
control
Dovitinib treated
Ctl
+Dov
KATO III
E
% Caspase 3/7 Activation
120
100
80
60
40
20
0
-20
KATO III
SNU-16
1nM
10nM
100nM
1μM
10μM
Concentration of Dovitinib
F
Control
5-FU 20mg/kg
Dovitinib 30mg/kg
Dovitinib 50mg/kg
Tumor Size (mm3)
1400
1200
1000
800
600
400
200
0
0
5
10
15
20
25
Days after Treatment

Figure A4: 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.

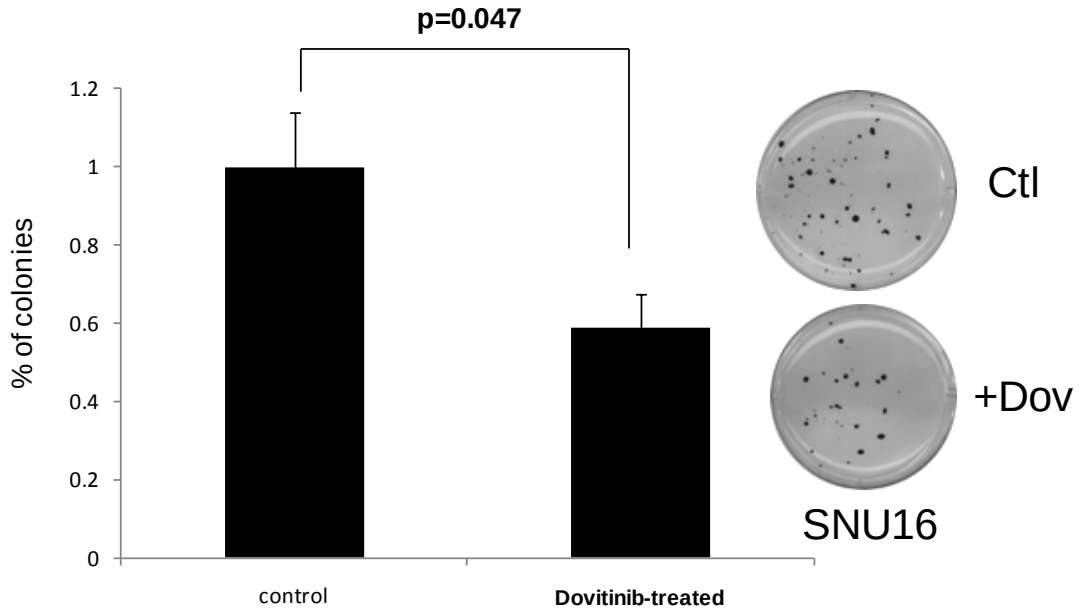


Figure A5: Phenotypic assay of *KLF5* in GC cell lines

A) KLF5 protein expression in GC cell lines. Western blotting was performed using an anti-KLF5 antibody. -actin was used as a loading control.SNU5 and YCC3 exhibit high KLF5 protein expression while SNU1 cells show relatively low KLF5 expression. B) KLF5 silencing and overexpression in SNU5, YCC3 and SNU1 GC cells. (Top Western blots) Confirmation of KLF5 silencing and overexpression. Left, KLF5 silencing by siRNA in SNU5. SNU5 cells silenced with KLF5 siRNAs resulted in a significant reduction in cell proliferation capacity compared to cells treated with scrambled siRNAs. (p=0.006). Center, KLF5 silencing in YCC3. YCC3 cells silenced with KLF5 siRNAs resulted in a significant reduction in cell proliferation capacity compared to cells treated with scrambled siRNAs. (p=0.011). Right, Effects of KLF5 overexpression in SNU1 cells. Compared to cells transfected with vector control, SNU1 cells transfected with a KLF5-expressing vector exhibit enhanced cell proliferation (p<0.0001). All experiments were performed a minimum of three independent times.

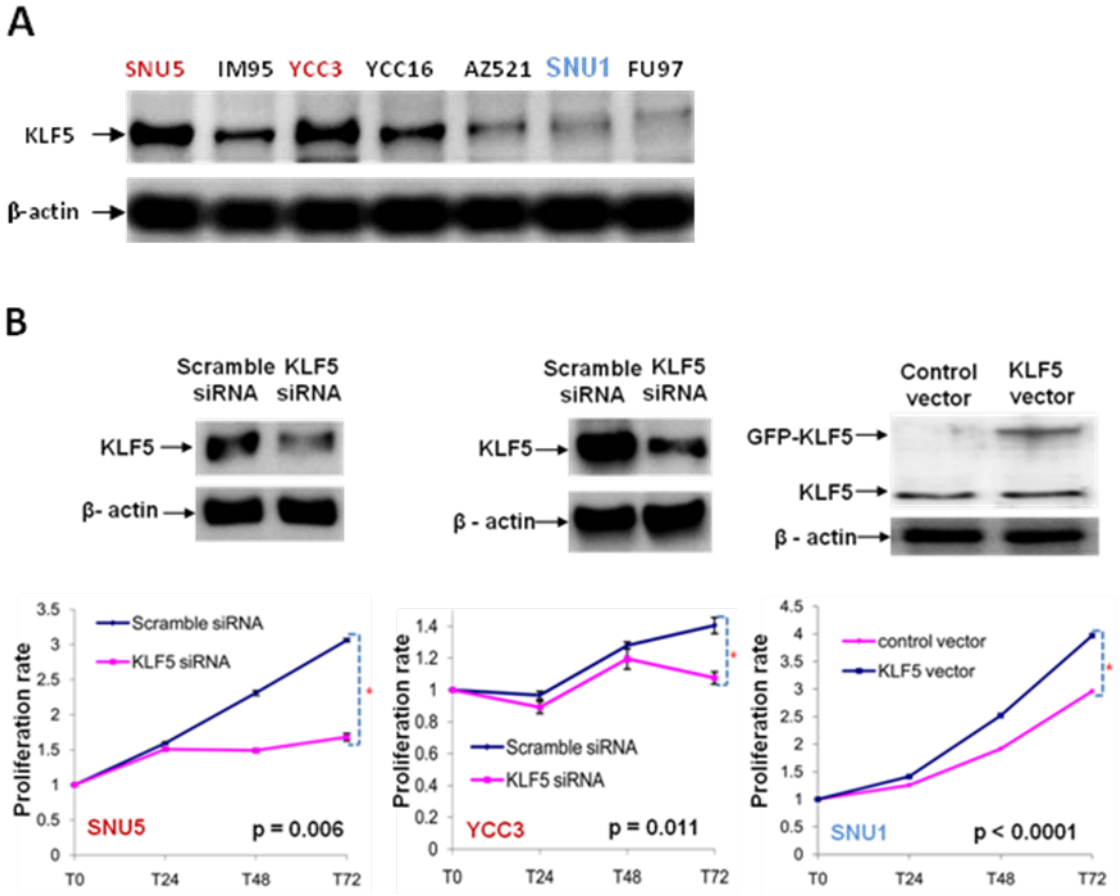


Table A1: DRP Analysis of Mutually Exclusive and Co-Amplification Interactions

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

Software: DRP script

```
DRP <- function(CNA_file, n, type='csv', write=T){
#'CNA_file' is the Copy Number Alterations indicator file;
#n is the required number of permutations
#'type' supports two data file types 'txt' for a tab-delimited text file and 'csv' for an Excel .csv file;
#'write' if 'T' DRP writes permutation results to a file
#%-------------------------
#%Niantao Deng
#%Duke NUS Graduate Medical School
#%niantaodeng@gmail.com
#%-------------------------
        cat('Reading CNA Indicator File\n')
        if(type=='csv'){
                topCNA <- read.csv(CNA_file,header=T)
    rownames(topCNA)<-topCNA[,1];topCNA<-topCNA[,-1] #first colume need to sample ID
                }
    else{
                topCNA <- read.table(CNA_file,header=T,comment.char='',fill=T,sep='\t')
    rownames(topCNA)<-topCNA[,1];topCNA<-topCNA[,-1] #first colume need to sample ID
                }
        ## Check for missing values
        NAs = any(is.na(topCNA))
        if(NAs){cat(c('Found',sum(is.na(topCNA)),'Missing Data Values\n'),sep=' ')}
        ## Check format
  if(is.matrix(topCNA) ==F){topCNA<-as.matrix(topCNA)}

        topCNA<-topCNA[order(rowSums(topCNA),decreasing=T),] # sort by number of events for each sample
        topCNA<-topCNA[,order(colSums(topCNA),decreasing=T)] # sort by number of events for each gene
        ##check for at least 1 CNA event for each gene or each sample
        sum_gene<-colSums(topCNA)
        sum_sps<-rowSums(topCNA)
        if(min(sum_gene) == 0){cat(c('Found',sum(sum_gene == 0),'Gene without CNA events\n'),sep=' ')}
        if(min(sum_sps) == 0){cat(c('Found',sum(sum_sps == 0),'Samples without CNA events\n'),sep=' ')}
```

```
cat('Perform permutations\n')


        #define the number of permutations and perform the permutations
        N<-n
        result_mat<-matrix(0, nrow=N,ncol=ncol(topCNA)*(ncol(topCNA)+1)/2) #matrix to put in the permutation results
        Time_start<-Sys.time()
        for(m in 1:N)
          {
          per_mat<-matrix(0, nrow=nrow(topCNA),ncol=ncol(topCNA)) #matrix for each permutation
          for(r in 1:min(nrow(topCNA),ncol(topCNA)))
            {
      if(nrow(topCNA) <= ncol(topCNA)) #permutate by row if nrow is less than ncol
       {
       gene_vec<-c(1:ncol(topCNA))
       #check for full rank cols, full rank column will be removed for later permutation
       col_rank<-sum((sum_sps- rowSums(per_mat))[c(r:nrow(topCNA))]>0)
                full_gene<-which((sum_gene-colSums(per_mat))[gene_vec] >= col_rank)
                if(length(full_gene)>0 & (sum_sps[r]-sum(per_mat[r,]))> 0)
        {
        if(sum_sps[r]-sum(per_mat[r,])>=length(full_gene))
          {
          per_mat[r,gene_vec[full_gene]]<-1;gene_vec<-gene_vec[-full_gene]
          }
        else
          {
          ind_rm<-sample(length(full_gene),sum_sps[r]-sum(per_mat[r,]))
          per_mat[r,gene_vec[full_gene[ind_rm]]]<-1;gene_vec<-gene_vec[-full_gene[ind_rm]]
          }
        }
       #permutate after removing full rank columns
       if(length(gene_vec)>0)
        {
```

```
num_gene<-sum_sps[r]-sum(per_mat[r,])
if(col_rank>1)
        {
 ind_gene_pri<-which((sum_gene-colSums(per_mat))[gene_vec] %in% c(2:col_rank))
 ind_gene_sec<-which((sum_gene-colSums(per_mat))[gene_vec] ==1)
 if(num_gene <= length(ind_gene_pri))
  {
  temp<-sample(length(ind_gene_pri),num_gene)
          if(length(temp)>0)
   {
    per_mat[r,gene_vec[ind_gene_pri[temp]]]<-1
   }
  }
 else
  {
  temp1<-sample(length(ind_gene_sec),num_gene - length(ind_gene_pri))
  temp<-c(ind_gene_sec[temp1],ind_gene_pri)
          if(length(temp)>0)
   {
    per_mat[r,gene_vec[ind_gene_pri[temp]]]<-1
   }
  }
 }
else
 {
 ind_gene<-which((sum_gene-colSums(per_mat))[gene_vec] %in% c(1:col_rank))
 temp<-sample(length(ind_gene),num_gene)
          if(length(temp)>0)
  {
  per_mat[r,gene_vec[ind_gene[temp]]]<-1
  }
 }
}
```

```
 }
if(ncol(topCNA) < nrow(topCNA)) #permutate by column if ncol is less than nrow
 {
                    sps_vec<-c(1:nrow(topCNA))
 #check for full rank rows, full rank rows will be removed for later permutation
 row_rank<-sum((sum_gene- colSums(per_mat))[c(r:ncol(topCNA))]>0)
           full_sps<-which((sum_sps-rowSums(per_mat))[sps_vec] >= row_rank)
           if(length(full_sps)>0 & (sum_gene[r]-sum(per_mat[,r]))> 0)
  {
  if(sum_gene[r]-sum(per_mat[,r])>=length(full_sps))
    {
    per_mat[sps_vec[full_sps],r]<-1;sps_vec<-sps_vec[-full_sps]
    }
  else
    {
    ind_rm<-sample(length(full_sps),sum_gene[r]-sum(per_mat[,r]))
    per_mat[sps_vec[full_sps[ind_rm]],r]<-1;sps_vec<-sps_vec[-full_sps[ind_rm]]
    }
  }
 #permutate after removing full rank rows
 if(length(sps_vec)>0)
  {
  num_sps<-sum_gene[r]-sum(per_mat[,r])
  if(row_rank>1)
             {
    ind_sps_pri<-which((sum_sps-rowSums(per_mat))[sps_vec] %in% c(2:row_rank))
    ind_sps_sec<-which((sum_sps-rowSums(per_mat))[sps_vec] ==1)
    if(num_sps <= length(ind_sps_pri))
     {
     temp<-sample(length(ind_sps_pri),num_sps)
              if(length(temp)>0)
      {
       per_mat[sps_vec[ind_sps_pri[temp]],r]<-1
```

```
    }
   }
  else
   {
   temp1<-sample(length(ind_sps_sec),num_sps - length(ind_sps_pri))
   temp<-c(ind_sps_sec[temp1],ind_sps_pri)
            if(length(temp)>0)
    {
    per_mat[sps_vec[ind_sps_pri[temp]],r]<-1
    }
   }
  }
 else
  {
  ind_sps<-which((sum_sps-rowSums(per_mat))[sps_vec] %in% c(1:row_rank))
           temp<-sample(length(ind_sps),num_sps)
           if(length(temp)>0)
   {
   per_mat[sps_vec[ind_sps[temp]],r]<-1
   }
  }
 }
}
       }
     cor_mat<-t(per_mat) %*% per_mat
     cor_vec<-numeric()
     for(s in 1:ncol(cor_mat))
       {
       cor_vec<-c(cor_vec,cor_mat[s,c(s:ncol(cor_mat))])
       }
     result_mat[m,]<-cor_vec
     }
   Time_end<-Sys.time()
```

```
            #compare with the observation and generate the empirical p vallues
            original_data<- t(topCNA) %*% topCNA
            out_mat<-data.frame()
            ind<-0
            for(i in 1:ncol(topCNA))
              {
              for(j in i:ncol(topCNA))
                {
                gene1<-colnames(topCNA)[i]
                gene2<-colnames(topCNA)[j]
                no1<-original_data[i,i]
                no2<-original_data[j,j]
                no3<-original_data[i,j]
                no4<-no1+no2-2*no3
                no5<-mean(result_mat[,ind+j-(i-1)])
                no6<-no1+no2-2*no5
                pval<-sum(result_mat[,ind+j-(i-1)]<=no3)/N
            input<-
data.frame(Gene1=gene1,Gene2=gene2,No_Gene1=no1,No_Gene2=no2,No_Both=no3,No_OnlyOne=no4,No_BothExp=no5,No
_OnlyOneExp=no6,PvalueME=pval,PvalueCO=1-pval)
            out_mat<-rbind(out_mat,input)
                }
              ind<-ind+ncol(topCNA)+1-i
              }
    out_mat<-out_mat[order(out_mat$PvalueME),] #sort by PvalueME
    cat('Output results...\n')
    cat(c('Permutation time:','\n'),sep=' ')
    print(Time_end-Time_start)
    if(write == F){return(out_mat)}
    else{write.csv(out_mat,file='permutation_results.csv',row.names=F)}
    }
```