A STUDY OF GENOMIC ABERRATIONS IN GASTRIC ADENOCARCINOMA

ALVIN ENG KIM HOCK

MBBS(NUS), M.Med.(Surg), MRCS(Eng), FRCS(Edinb)

A THESIS SUBMITTED FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF BIOCHEMISTRY

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

This thesis would not have been possible without the help of Prof Kon Oi Lian who has patiently guided me at every step. I would also like to thank Louise Lee Sze Sing who was instrumental in assisting me with the data analysis and Leong Siew Hong for teaching me the basics of genomic research.

Our pathologists Dr Tan Soo Yong and Dr Lai Siang Hui who kindly agreed to read and verify all the tissue for this study. Magdalene Koh Hui-Kheng for assisting with the histopathology cores.

This study was conducted with funds from the National Medical Research Council of Singapore and the assistance of Mr. Dennis Lim Teck Hock.

Table of Contents

Summary

Despite declining incidence and mortality, gastric cancer remains the fourth most common cancer and the second leading cause of death in the world. Gastric carcinogenesis is believed to occur through one of 3 pathways, the commonest of which involves sequential changes in mucosal histology, from normal through intestinal metaplasia and dysplasia to overt carcinoma. We aimed to investigate the genomic changes that parallel these mucosal transformations as they progress along the pathway described by Correa in 1988.

57 specimens representing the histological types of overt carcinoma, dysplasia, intestinal metaplasia and adjacent histologically normal mucosa were obtained from the archived formalin-fixed paraffin-embedded pathology blocks of 17 patients. Genomic DNA was extracted from each specimen. Comparative genomic hybridization was performed using a validated 2464-BAC clone array having an average inter-clone interval of 1.4 Mb.

Our results revealed that all 4 histological types harbored extensive genomic changes that were highly similar. Further array CGH experiments conducted with tissue harvested from non-cancer gastrectomy specimens showed no evidence of significant copy number aberrations. Additional experiments found that the distant margin blocks of the same cancer patients had a distinctly different genomic signature compared to the earlier 57 specimens.

Several prospective sets of specimens that were harvested and processed in our laboratory confirmed that the genomic profile of gastric mucosa at the margin of a cancer resection is almost normal while the copy number aberrations in adjacent histologically normal gastric mucosa mirror those found in the tumor itself.

Several regions of interest that were found in our study included the +20q13, +8z23, -19p13 and +17q21 cytobands. These copy number aberrations were present in the adjacent mucosa as well as in the tumors.

The genome-wide study of adjacent normal mucosa in gastric cancer with array CGH has not been reported before and our findings are consistent with and provide genomic evidence for field cancerization in gastric adenocarcinoma. Our findings in gastric carcinoma are supported by recent discoveries of genomic, proteomic and nanoscale structural abnormalities in histologically normal adjacent colonic, prostatic, pancreatic and pulmonary tissue from cancer patients.

The concept of field cancerization was first proposed in 1953. This theory suggests that chronic exposure to a DNA-damaging agent such as a chemical compound or an infection like *H.pylori* leads to the clonal expansion of inappropriate cell types that exhibit genetic instability. This premalignant state would eventually lead to transformation into overt carcinoma. The field cancerization theory mirrors the Correa hypothesis and it provides some explanation for the frequency of recurrence in gastric cancer patients.

The understanding of gastric carcinogenesis as a field cancerization event would provide the impetus to focus resources on the study of premalignant histologically normal gastric mucosa that harbors the initiators of gastric carcinogenesis.

List of Tables

List of Figures

List of Abbreviations

ACAVIS	Array CGH Analysis and Visualization software
BAC	Bacterial artificial chromosome
CGH	Comparative genomic hybridization
CpG	Cytosine p guanine
CT	Computerized tomography
FAP	Familial Adenomatous Polyposis
FFPE	Formalin-fixed paraffin-embedded
GCEP	Gastric Cancer Epidemiology and Molecular Genetics Program
HDGC	Hereditary diffuse gastric cancer
HNPCC	Hereditary Nonpolyposis Colon Cancer
IM	Intestinal metaplasia
LCM	Laser Capture Microdissection
LOH	Loss of heterozygosity
LOWESS	Locally weighted scatterplot smoothing
MSI	Microsatellite instability
NSAID	Non-steroidal anti-inflammatory drug
PCR	Polymerase chain reaction
SNP	Single nucleotide polymorphism
SPEM	Spasmolytic polypeptide expressing metaplasia
SSC	Saline-sodium citrate buffer
UCSC	University of California at Santa Cruz
UCSF	University of California San Francisco
WGA	Whole genome amplification

Chapter 1

Introduction and Literature Review

1.1 Gastric cancer epidemiology

Despite a major decline in incidence and mortality rates over the last fifty years, gastric cancer remains the fourth most common cancer and the second leading cause of cancer death in the world (1). More recently, developing countries have tended to predominate in incidence. Changes in diet and improvements in hygiene are generally considered as being responsible for the decrease in incidence rates in the developed world (2). Male-to-female incidence ratios are usually about 1.5 to 2.5 with higher ratios for intestinal-type cancer and higher risk populations (3).

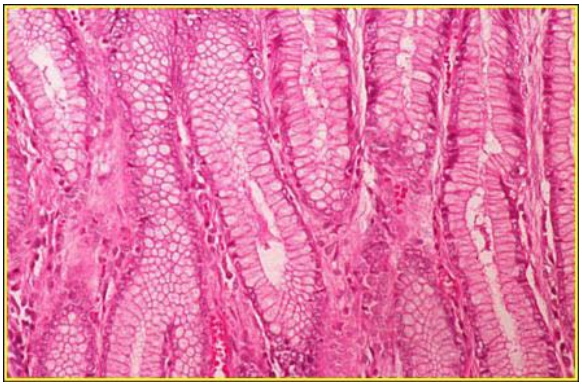
The incidence of gastric cancer in Singapore has likewise been decreasing. However, it remains firmly within the top five malignancies in the country. The latest census shows that it is the 4th most common malignancy and the 3rd greatest cause of cancer-related mortality in both males and females combined (4).

Most cases of gastric cancer present at an advanced stage and this is reflected in the fact that the mortality rate of gastric cancer in a population is usually higher than its incidence rate. The possible exceptions to this are countries with a high incidence which have developed mass screening programs. Identifying and treating gastric cancer at an early stage has the effect of prolonging overall survival and this has been observed in Japan in the last 15 years.

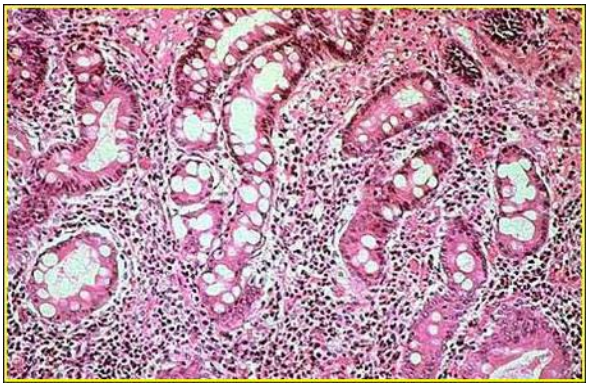
The Singapore Gastric Cancer Epidemiology and Molecular Genetics Program (GCEP) established in 2003 involves active mass screening of a cohort of 4000 patients in an attempt to determine possible targets for primary or secondary prevention in order to reduce the incidence of gastric carcinoma (5).

1.2 Gastric cancer pathology

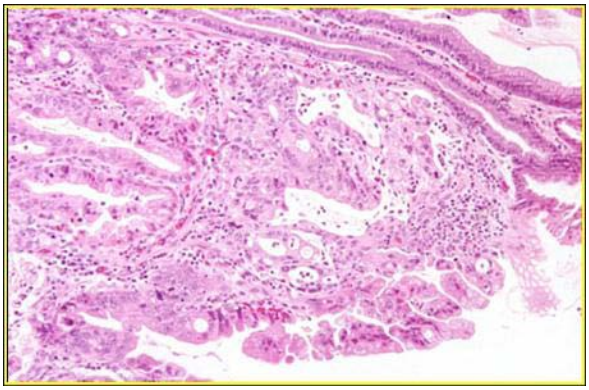
It is generally recognized that there are 2 main histological types of gastric carcinoma as first described in 1965 (6). The Lauren classification defines these as: (a) the intestinal type which is characterized by the metaplastic transformation of gastric-type mucosa to an intestinal type with abundant goblet cells; and, (b) the diffuse type which is defined by the presence of poorly differentiated signet ring cells. Both types may also co-exist thereby giving rise to a third entity of 'mixed' pathology.



Normal gastric epithelium



Gastric intestinal metaplasia



Gastric adenocarcinoma

Figure 1. Histology of gastric mucosa

The intestinal type is the more common variant seen and it is associated with an increased incidence of chronic atrophic gastritis and gastric atrophy. The diffuse cancers do not have this association. It is believed that intestinal metaplasia (IM) is the result of an inflammatory reaction which may be precipitated by ingestion of certain substances or by the presence of an infection such as *Helicobacter pylori*.

The occurrence of gastric dysplasia has been postulated to be a further step in the development of intestinal-type gastric cancer (7) although it is known that it may on occasion regress. The problems associated with histological interpretation of dysplasia are well-documented and these include inter-observational variation as well as the difficulty in differentiating high-grade dysplasia from intramucosal carcinoma (also known as early gastric cancer). The Vienna classification (8) (9) now provides for more accurate diagnosis of dysplastic lesions. Nevertheless, the difficulty of diagnosing dsyplasia accurately has hindered studies involving DNA or RNA as fresh frozen specimens cannot be read with the required degree of accuracy while formalin-fixed paraffin-embedded tissue is usually of suboptimal quality for genetic assays.

The other important category of precancerous stomach lesions are gastric mucosal polyps. These may be divided into 3 main categories: fundic gland polyps; hyperplastic polyps, and adenomas. The latter 2 have a slightly increased risk of progressing to carcinoma, with adenomas generally recognized as being of greater significance.

1.3 Etiology & Risk Factors

1.3.1 Risk Factors

With the exception of genetic syndromes, by far the strongest established risk factor for gastric cancer is *H. pylori* infection. Male gender, smoking, previous gastric resections and adenomatous polyps have also been associated with a higher incidence of gastric carcinoma. Epstein-Barr virus has also been reported to be responsible for approximately 5% of stomach malignancies and this subtype of gastric cancer has been shown to have distinct molecular and clinicopathologic characteristics (10).

Infection:	*Helicobacter pylori*
	Epstein-Barr virus
Atrophic gastritis	
Previous partial gastrectomy	
Adenomatous gastric polyps	
Blood group A	
Type III intestinal metaplasia	
Smoking	
High salt intake and/or preserved foods	
Genetic:	Familial adenomatous polyposis (FAP)
	Hereditary diffuse gastric cancer (HDGC)
	Peutz-Jeghers Syndrome
	Hereditary Nonpolyposis Colon Cancer (HNPCC)
	Li-Fraumeni Syndrome (inherited TP53 mutation)

Table 1. Risk factors for gastric cancer

1.3.2 Etiology

It has been postulated that there are at least 3 important pathways that lead to cancer in the stomach: (a) stepwise morphological transformation involving intestinal metaplasia; (b) diffuse type gastric carcinoma which involves signet ring cells thought to arise from the stem cell zone; and , (c) spasmolytic polypeptide expressing metaplasia (SPEM) where the gastric glands become filled with cells that express the polypeptide TFF2 (TreFoil Factor-2 also known as SP) (11).

The fundamental mechanisms underlying these pathways generally involve some degree of genomic instability. Several phenotypes of instability have been identified in gastric cancer (12).

The chromosomal instability phenotype is associated with mutation in genes that control the segregation of genetic elements. Chromosomal rearrangement or losses or gains of chromosomes can lead to either oncogene activation or tumor-suppressor gene inactivation.

The microsatellite instability (MSI) phenotype is characterized by defective repair of DNA replication. Inefficiencies of one or more of the mismatch repair genes can cause MSI which then results in frameshift mutations, thus altering the translation of DNA into protein products.

The third phenotype involves the cytosine p guanine (CpG) island methylator. Abnormal methylation of guanine and cytosine-rich regions results in silencing of tumor-suppressor genes leading to uncontrolled cellular growth and malignancy.

The recent discovery of cancer stem cells has led to the intriguing possibility that these immortal cells may be a key initiator of gastric carcinogenesis (13) (14). The stem cell may either be an organ-specific indigenous gastric stem cell or a bone

marrow-derived cell (BMDC) recruited to the gastric epithelium as a result of chronic inflammatory stress.

1.3.3 Hereditary diffuse gastric cancer (HGDC)

Diffuse-type gastric carcinoma is distinguished by the absence of defined premalignant lesions and poorly differentiated histology (6). It is also associated with *H. pylori* infection and is sometimes described as 'linitis plastica' alluding to a macroscopic appearance of widespread thickening involving the entire organ.

The discovery of the genetic events leading to diffuse gastric carcinoma is one of the success stories of modern genomics. A kindred of New Zealand Maoris that had diffuse-type carcinoma were found to have hereditary mutations of CDH1, a tumor-suppressor gene which codes for the protein E-cadherin (15). This protein mediates homophilic cell-cell interactions and establishes cell polarity. Loss of both alleles of the gene results in reduced expression of cadherin and this is found in up to 50% of all gastric cancers and up to 83% of diffuse carcinomas (16).

1.3.4 Correa's hypothesis

Also known as the intestinal pathway of gastric carcinogenesis, this hypothesis is central to our study as intestinal-type carcinoma is the predominant form in our population. Pelayo Correa first postulated in 1975 that nitroso compounds arising from ingested nitrites, in the presence of an impaired mucous barrier, may be the initiating step in a cascade of events leading to overt carcinoma (17).

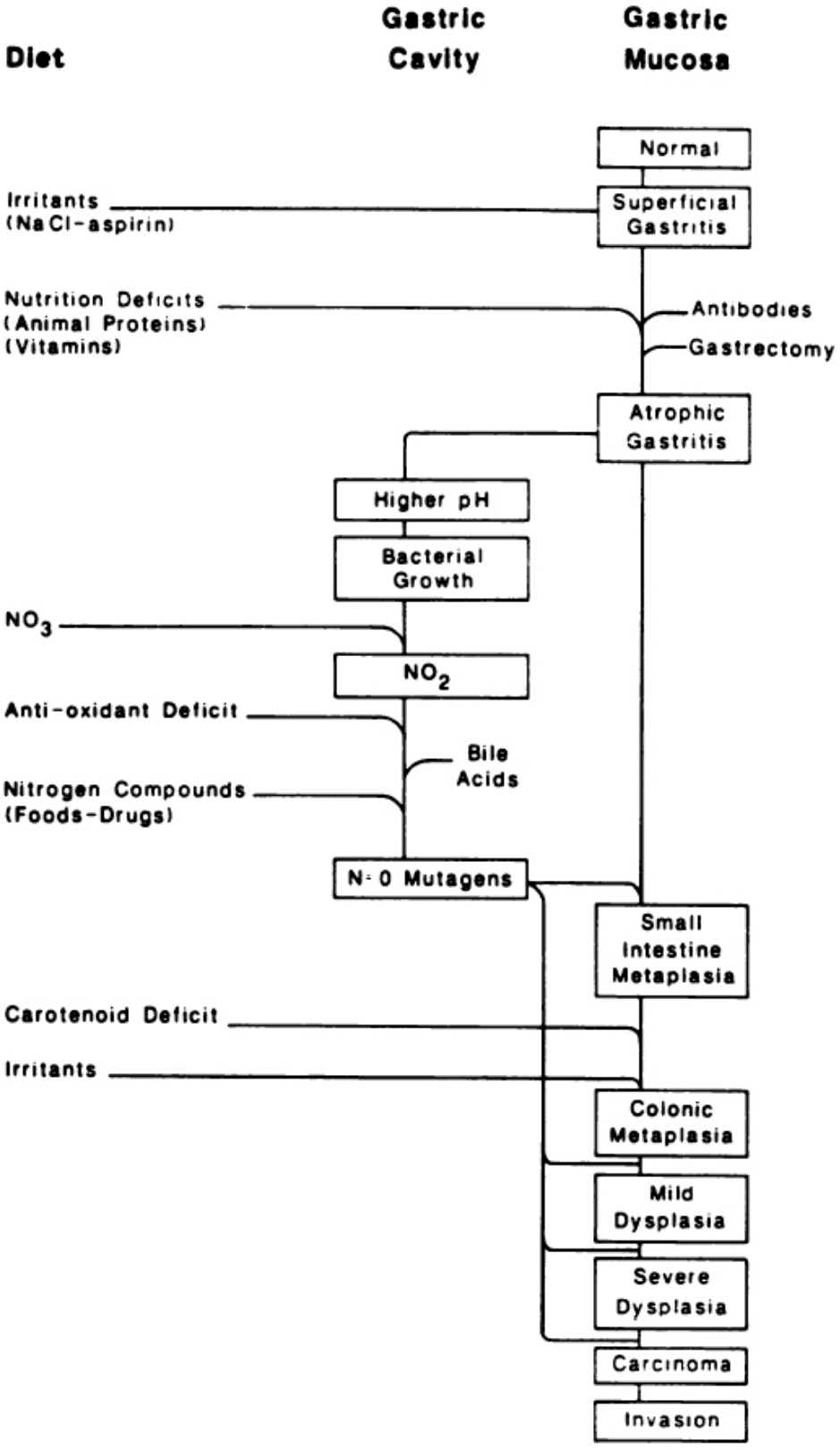


Figure 2.Correa's hypothesis of gastric cancer etiology (7)

The Correa model of gastric carcinogenesis implicates four distinct histological entities: normal mucosa, intestinal metaplasia, dysplasia and carcinoma. Assuming that accurate samples are obtained, it would then be possible to elucidate the molecular and genomic signatures of each histological type. The accumulation of genetic alterations in a linear or parallel route to overt carcinoma may then be described much as it already has in colorectal malignancies (18).

1.4 Screening for Gastric adenocarcinoma

A mass screening program for gastric cancer has existed in Japan since 1960 (19). Despite intensive research for the last 49 years, the only recommended tools for screening today remain diagnostic contrast radiography and endoscopy.

The last 20 years has seen rapid advances in technology for biomedical research. The search for biomarkers is particularly interesting as it may one day provide a simple tool for mass screening of any number of diseases, gastric cancer among them. The advantages of a biomarker cannot be overstated as the cost of any blood test or genetic test would almost certainly be at least an order of magnitude less than that of endoscopy. The convenience of a serum biomarker would also encourage a population to come forward for screening.

Biomarker discovery and genetic research are inextricably linked. A biomarker may be a protein or even a genetic test itself. Thus one possible avenue for biomarker discovery would lie along the route of research into abnormalities in the genomic DNA of cancer patients.

1.5 Management of gastric cancer

The diagnosis of gastric cancer is in almost all instances made on diagnostic endoscopy and biopsy. This is an invasive procedure and relatively expensive. As early gastric cancer may be asymptomatic or present with non-specific symptoms such as dyspepsia, the majority of patients are usually diagnosed at stage II or worse unless there is a nationwide screening program in place.

Surgical removal of the primary tumor and regional lymph nodes is the only curative option for gastric cancer. Adjuvant chemotherapy and radiotherapy provide adjuncts to curative surgery and also serve to slow tumor progression in advanced cases. Neoadjuvant therapy may reduce tumor volume with the goal of eventual curative resection.

Staging of the disease prior to surgery and at follow-up after surgery is usually with CT scans and endoscopy. The problem with this is that microscopic disease is not detectable with these methods and when macroscopic recurrence occurs it usually signifies metastatic or incurable disease. Thus the issue of recurrence, particularly in the locoregional lymph nodes, at the resection site and on peritoneal surfaces, constitutes a difficult diagnostic and treatment problem.

In general, 5-year survival rates for gastric cancer are approximately 20% worldwide except in Japan where the mass screening program and aggressive early treatment has contributed to 5-year survival rates of up to 60% (20). Local recurrence rates can be as high as 54% (21) (22).

Genomic and molecular markers that can predict disease patterns such as lymph node metastasis (23) or survival (24) can prove to be a valuable tool in

diagnosing or prognosticating gastric cancer patients. Biomarkers are also useful in optimizing the choice of adjuvant therapy (25) (26).

Stage Grouping

Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T1	N1	M0
	T2a/b	N0	M0
Stage II	T1	N2	M0
	T2a/b	N1	M0
	T3	N0	M0
Stage IIIA	T2a/b	N2	M0
	T3	N1	M0
	T4	N0	M0
Stage IIIB	T3	N2	M0
Stage IV	T4	N1, N2, N3	M0
	T1, T2, T3	N3	M0
	Any T	Any N	M1

Summary

Stomach	
T1	Lamina propria, submucosa
T2	Muscularis propria, subserosa
T2a	Muscularis propria
T2b	Subserosa
T3	Penetrates serosa
T4	Adjacent structures
N1	1 to 6 nodes
N2	7 to 15 nodes
N3	>15 nodes

Table 2. TNM staging adapted from UICC 6th edition (2002)

1.6 Current research directions in gastric cancer

The development of high-throughput technologies such as microarrays has ushered in an era of research characterized by the extensive use of statistics and bioinformatics. Microarrays can be classified in various ways. Arrays can be constructed on glass slides, silicon substrate or even beads. The genetic probes on the arrays may be complementary-DNA, oligonucleotides or small PCR fragments. These probes are typically deposited on the substrate by spotting with fine-pointed pins, inkjets or photolithography. Arrays can be designed for single channel or double-channel usage depending on the need for absolute quantitation versus relative estimation of one sample in comparison to another. Microarrays may be used to detect DNA or RNA. Gene expression studies typically employ cDNA arrays while SNP (single nucleotide polymorphism) studies usually involve oligo-arrays.

Gastric cancer, like any other malignancy, is characterized by multiple genetic and epigenetic alterations. Intense research into the molecular biology of gastric cancer over the past 20 years has revealed 3 pathways for gastric carcinogenesis as mentioned in section 1.3.2. The 2 classical pathways are shown overleaf. The more recently described SPEM pathway has yet to be fully characterized.

By far the most well known is the intestinal pathway and this is to be expected since it is the most common form of gastric carcinoma encountered in clinical practice. However, the breakthrough discovery of E-cadherin has catapulted the diffuse pathway to prominence in recent years. All these pathways are characterized

by alterations of the genome in 3 fundamental ways: chromosomal instability, microsatellite instability and epigenetic changes such as DNA methylation.

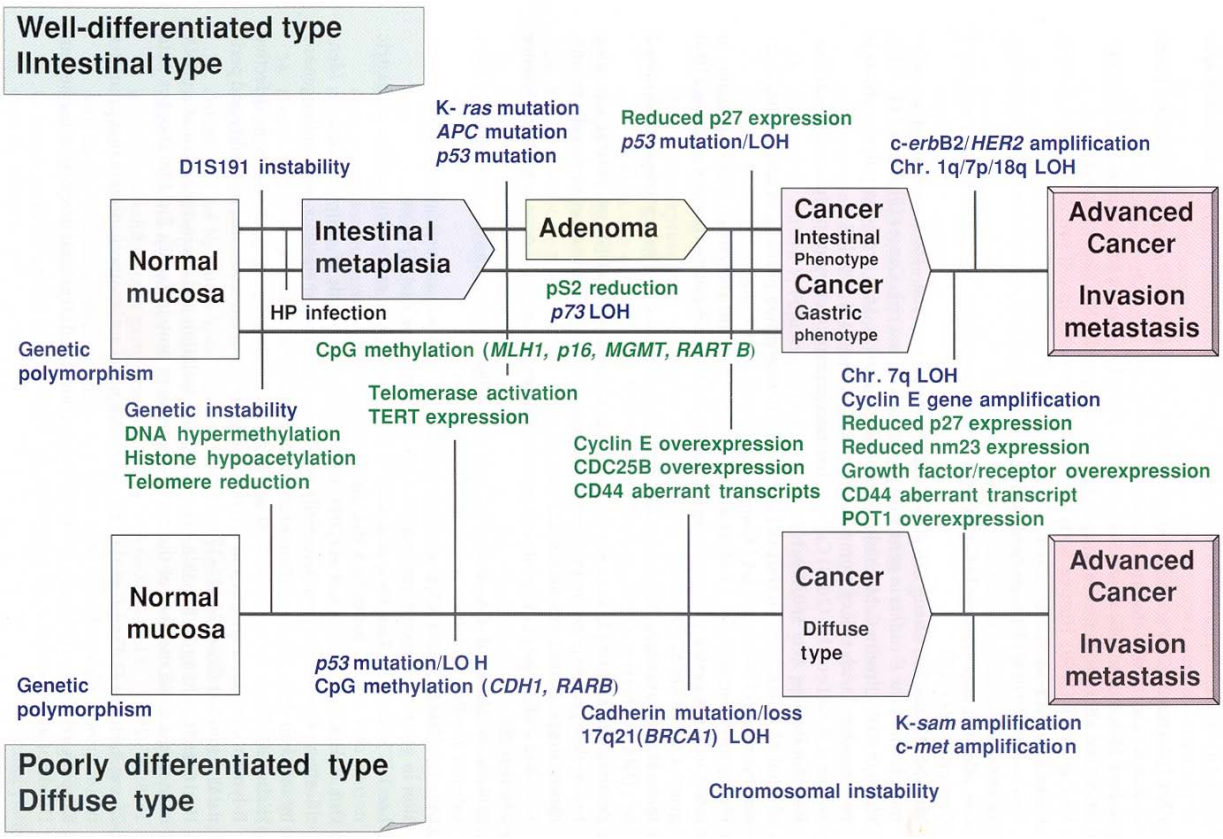


Figure 3. Genetic (blue) and epigenetic (green) alterations in gastric carcinogenesis. [Adapted from pg 70 of reference (27)]

One of the limitations of conventional molecular research is that it fails to address non-coding regions of the genome i.e. the gene deserts. Several techniques such as comparative genomic hybridization (CGH) have been developed to address this shortcoming and our laboratory has had some experience with these.

A previous study in our laboratory using metaphase-spread conventional comparative genomic hybridization (CGH) had demonstrated significant copy number gains and losses in gastric cancer tissue (24).

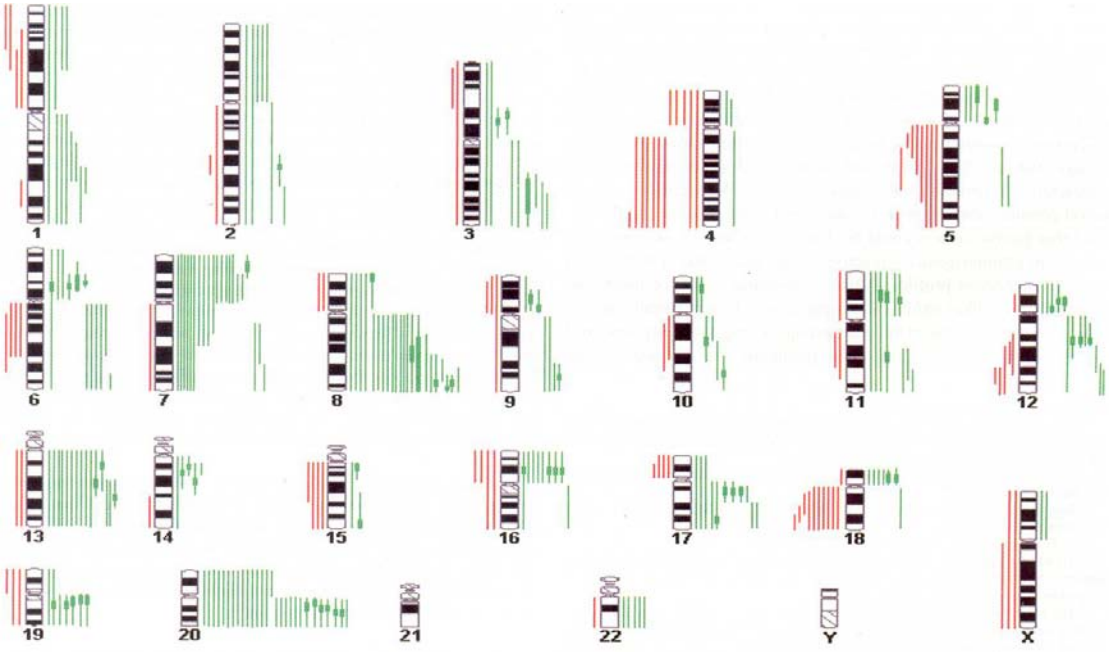


Figure 4. Chromosomal gains and losses in gastric cancer patients. Gains are shown as green lines and losses as red lines. Thick solid lines are highly amplified regions. (24)

1.7 Array CGH

The chromosomal changes such as gene amplification and deletions can often be detected by an increase or decrease in the amount of genomic DNA within the cell. This was the basis of a technique first described by Kallioniemi in 1992 which utilized competitive simultaneous in situ hybridization of fluorescent-labeled tumor and normal DNA in equimolar quantities to a normal human metaphase spread. Regions of relative amplification and deletion could then be identified by measuring the color

ratio of the two fluorescent dyes (28). This technique is now known as comparative genomic hybridization (CGH).

However, usage of metaphase chromosomes limits the detection of abnormalities involving short regions (< 20 Mb) of the genome. Microarray technology when applied to CGH, using a spotted array of mapped sequences instead of metaphase chromosomes overcomes the limitations of conventional CGH (29). The initial attempts were made with cDNA arrays but eventually the use of BAC-arrays has come to be recognized as a better way to determine regions of chromosomal gains and losses. The resolution of the array would then be a function of the length of the spotted sequences and the distance between the sequences on the human genome.

BAC is an acronym for bacterial artificial chromosome. It was developed in 1992 as a means of cloning long sequences (>300kb) of the human genome and it remains a useful tool for accurately replicating long sequences of human DNA (30). A BAC-array is a DNA-microarray that uses BAC clones as the spotted probes instead of the usual cDNA or oligonucleotides.

The advantages of BAC array CGH over conventional metaphase-spread CGH include higher resolution (1 Mb vs. 20Mb), simultaneous coverage of the entire genome and the requirement of smaller amounts of test DNA (300-500 ng vs. 1 µg)

1.8 Objectives of this study

The objective of this study is to utilize BAC array CGH to document the genomic aberrations in matched samples of gastric carcinoma, dysplasia, intestinal metaplasia and adjacent normal mucosa. The intention is to discover whether or not there is a steady progression of genomic copy number changes that parallels the transformation of susceptible mucosa into overt carcinoma. This could be the first step in an effort to discover possible regions of translocation, duplication or deletion. Although outside the scope of this study, the eventual potential discovery of break-points or duplicated/deleted genes could provide possible diagnostic, therapeutic or prognostic markers that can improve the clinical management of patients with gastric cancer.

Chapter 2
Materials & Methods

2.1 Obtaining samples

Records for all patients who had undergone gastrectomy for cancer at the Singapore General Hospital for the last 5 years were traced. Their pathology records were screened to identify gastrectomy specimens that contained all 4 histological types that we required for our study: adjacent normal mucosa, intestinal metaplasia, dysplasia and overt carcinoma.

A total of 15 suitable gastrectomy specimens were obtained in this manner. The original formalin-fixed paraffin-embedded (FFPE) tissue blocks were then traced from the archives of the Department of Pathology. Fresh slices from these blocks were fixed on slides and read by our collaborating pathologists to confirm that the blocks were suitable for our purposes.

Two additional sets of blocks containing all 4 tissue types were obtained from collaborators in Malaysia. These were processed in the same manner and had diagnosis and suitability re-confirmed by our pathologists.

We had the following inclusion criteria:

1. Only gastric adenocarcinomas were included in this study
2. All tissue was to be obtained from formalin-fixed paraffin-embedded blocks
3. All 4 histological types had to be present from blocks harvested from the same patient at the same operation. “Adjacent normal” specimens are histologically normal samples of gastric mucosa taken from the same paraffin block as abnormal tissue. “Distant / Far normal” specimens are only taken from blocks that are specifically labeled as the proximal or distal resection margins.

2.2 Core & Slice

The initial plan was to sample slices from the archived blocks using Laser Capture Microdissection (LCM) (31). However, this was not possible for our study as there was no expertise available within the Department of Pathology at that time for the procedure.

In order to overcome this obstacle to the study, we designed another method of sampling the blocks. We had available a machine used for constructing tissue microarrays. Using this hollow 'punch' device usually employed for obtaining cores for tissue microarrays, we were able to obtain cores of tissue from the blocks.

The procedure was as follows:

1. Slices taken from each block were read by the pathologist to identify areas for core punch biopsy
2. 1 mm diameter 'punch cores' were obtained from the blocks
3. A 40-micron height section was taken from the mucosal end of the punch core
4. A standard slice was taken from the top and bottom of this 40-micron height section and prepared on a glass slide
5. The top and bottom slices were read by a pathologist to confirm that only the correct tissue type was present.

Fig 5. Punch cores

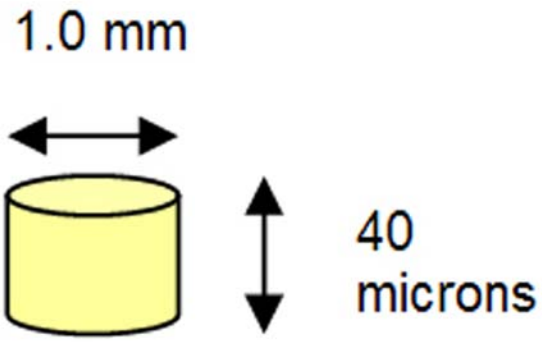


Fig 6. Section of the 'punch core'

In order to verify that the sampling method was accurate for our purposes, genomic DNA was extracted from a xenoimplanted tumor established from gastric cancer cell line (SNU-5) and tested on CGH and aCGH using recommended protocols. The results were compared against the known genomic profile of the carcinoma in our records. At a slice thickness of 40 microns, we were able to obtain enough DNA of sufficient quality that the aCGH profile of this extracted DNA matched the known genomic signature of the SNU-5 cancer.

A literature search revealed that a similar form of microdissection had just been described by another group (32) (33). The method described by Paris *et al.* used a hollow bore instead of a tissue micro-arrayer punch. We also differed in that we did not use the entire core but instead opted to use only a thin section of the core, thereby allowing for an additional verification step of the top and bottom slices of this section. We believe that the accuracy of our method would be enhanced since the possibility of non-target tissue within the 40-micron-height section would be minimized.

Since LCM is employed on very thin single slices mounted on glass slides, the potential disadvantage of our sampling method compared to LCM would be the possibility of harvesting non-target tissue within the 40-micron space. However, given the minute amounts of DNA available from a typical LCM specimen, whole genome amplification (WGA) is inevitably necessary. WGA would potentially introduce

artefactual copy number aberrations if the genome is not uniformly amplified. WGA methods like multiple displacement amplification (34), degenerate oligonucleotide-primed PCR (35), ligation-mediated PCR (36) and primer extension preamplification (37) are known to introduce copy number bias of dispersed genomic regions (38). The advantage of our sampling method is that it allows isolation of sufficient DNA from the sample itself, precluding the necessity for an additional WGA step.

2.3 DNA extraction

We used a commercial kit (PureGene from Gentra Systems Inc) to extract the genomic DNA from the formalin-fixed paraffin-embedded tissue (FFPE) sections. The protocol is detailed in Appendix 1. Briefly, the process involves de-paraffinization of the sample with xylene which is subsequently removed with 100% ethanol.

A cell lysis solution and proteinase K are then added in the second step which typically lasts 3 hours to overnight. This is followed by RNAse A treatment before proceeding with protein precipitation.

Finally the DNA is precipitated with isopropanol and glycogen. The cell lysate is centrifuged at 16000 g for 5 minutes and the supernatant drained to obtain a pellet of purified DNA which is then hydrated to 20µL of solution.

The DNA concentration is then quantified with Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc.). The typical yield from a 40-micron section was 30-40 ng/µL giving an overall yield of 600-800 ng. The DNA is then stored at 4°C until required.

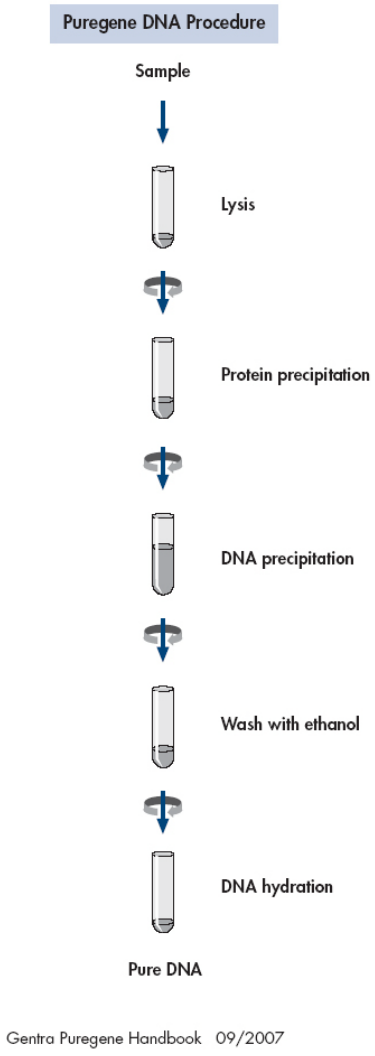


Fig 7. Flowchart for purification of DNA from FFPE tissue.

2.3.1 Reference DNA

The procedure of CGH necessitates a reference DNA sample for use in the competitive hybridization process. For our controls, we elected to use a pooled reference DNA comprising equal amounts of DNA harvested from formalin-fixed paraffin-embedded (FFPE) splenic tissue from 15 normal human males.

The reasons for this decision are:

1. In order to study the adjacent normal tissue profile, we could not use the histologically normal adjacent gastric tissue itself as the reference DNA sample.
2. The use of patient blood as a reference DNA posed 2 problems:
    a. The blood was often not available for most patients in our study
    b. The use of lymphocyte DNA of a much higher quality than the FFPE test specimens could introduce biases in the detected copy number results.
3. Since none of the patients had their own matched non-gastric FFPE tissue for use as a reference DNA source, the reference DNA was sourced from patients not part of the study group.
4. Pooled genomic DNA from 15 patients was used as a reference to minimize the possibility that 1 sample alone may have some idiosyncratic copy number aberration itself.
5. FFPE splenic tissue was used as few stomachs (or indeed any other organ) are usually removed in surgery unless there is a gross abnormality. Spleens are the exception as traumatic life-threatening splenic rupture is often routinely treated with splenectomy. These spleens are normal in size, structure and histology.

The pooled spleen reference DNA was compared to a DNA sample from a lymphocyte source which we had previously identified as normal. The resulting array image can be seen in section 3.1.2 and the corresponding karyogram in section 3.1.5. This was taken as confirmation that our pooled DNA was a valid reference point for our study.

2.4 Digestion of genomic DNA

This is the first step in the process of labeling DNA for hybridization (see Appendix 2). We used DpnII as the restriction enzyme in this step and the mixture was incubated at 37°C for at least 5 hours to allow the reaction to run to completion.

2.5 Purification of DNA

The digested products had to be purified in order to filter out unnecessary fragments that could have added to the 'noise' in the hybridization images. We used another commercial kit for this stage (QIAquick PCR Purification Kit from Qiagen Inc.) (see Appendix 2).

2.6 Labeling and hybridization

We obtained our BAC arrays from the University of California San Francisco (UCSF) Comprehensive Cancer Center Microarray Core facility. The specific array used was the HumArray 2.0 with an average spacing between clones of 1.4Mb (39). This BAC array comprised 2464 BAC clones spotted in triplicate (7392 spots) on a coated glass slide.

The protocol for BAC array hybridization was modified from that used by the UCSF core facility (http://cancer.ucsf.edu/array/protocols/index.php). The detailed protocol can be found in Appendix 2 and Appendix 3.

Briefly, we started with equal amounts (at least 500ng) of test and reference genomic DNA. The DNA was first denatured at 99°C with a random primer solution (Bioprime DNA labeling system from Invitrogen Inc.).

The mixture was then cooled on ice before adding Klenow fragment DNA polymerase (Bioprime DNA labeling system from Invitrogen Inc.) together with a mixture of 0.2 mM unlabeled dATP, dCTP, and dGTP; 0.1 mM unlabeled dTTP. Finally, either Cyanine-3-conjugated dUTP (test DNA) or Cyanine-5-conjugated-dUTP (reference DNA) was added to the mixture. (The cyanine-conjugated-dUTP dyes were sourced from Amersham/GE Healthcare). The entire mixture was then incubated at 37°C for at least 4 hours.

We used Microcon YM-30 Centrifugal Filter Units (from Millipore Inc.) to remove unincorporated nucleotides from the labeling reaction. At this stage it was possible to assess the labeling efficiency by the intensity of the color of the flow-through. The concentration of the labeled product was then measured with the Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc.).

As preparation for the hybridization process, we combined equal amounts Cy3-dUTP-labeled test DNA and Cy5-dUTP-labeled reference DNA with human Cot-1 DNA (from Invitrogen Inc.) and precipitated the mixture using 3M pH5.2 sodium acetate and ice-cold 100% ethanol. The samples were allowed to fully precipitate for 60 minutes at -20°C and then centrifuged at 16,100 rpm at 4°C for another 60 minutes to produce a violet-colored pellet of labeled genomic DNA. The pellet was then left to dissolve in the dark for an hour in a 60µL of a pre-hybridization solution comprising 10% dextran sulfate, 2× SSC, 50% formamide, 4% SDS, and water.

The labeled gDNA mixture was then denatured at 73°C and then incubated at 37°C for an hour to allow pre-annealing of the Human Cot-1 DNA to the labeled probes.

The array boundaries on the glass slide are virtually invisible to the naked eye and we marked these using a diamond-pen under phase-contrast microscopy. We then applied Hybaid EasiSeal 65µL Frames (Cat.No.HBOSSSEZ2E from Fisher Scientific Inc.) around each array. The arrays were then placed on a slide warmer at 37°C for 10 minutes.

The pre-hybridization solution was again employed, this time as a wetting solution on the slide arrays. Once the wetting solution was re-aspirated, the hybridization mixture itself was applied to the array. The glass slides were placed in a horizontal position arrays facing up in a slide box containing some washing solution (50% formamide and 2× SSC at pH7) in the base to maintain humidity. The box was sealed with parafilm and placed on a slow rocker at 37°C for 48-68 hours in the dark.

Post-hybridization, the slides were washed in a solution of 50% formamide and 2× SSC at pH7 at a temperature of 50°C for 20 minutes and then in PN buffer (0.1M Na_2HPO_4, 0.1% nonidet P40) at room temperature for 15 min. A final rinse in 2X SSC solution preceded the serial dehydration with ethanol solutions. The slides were then spun-dried at 800 rpm in a centrifuge for 2 minutes prior to imaging.

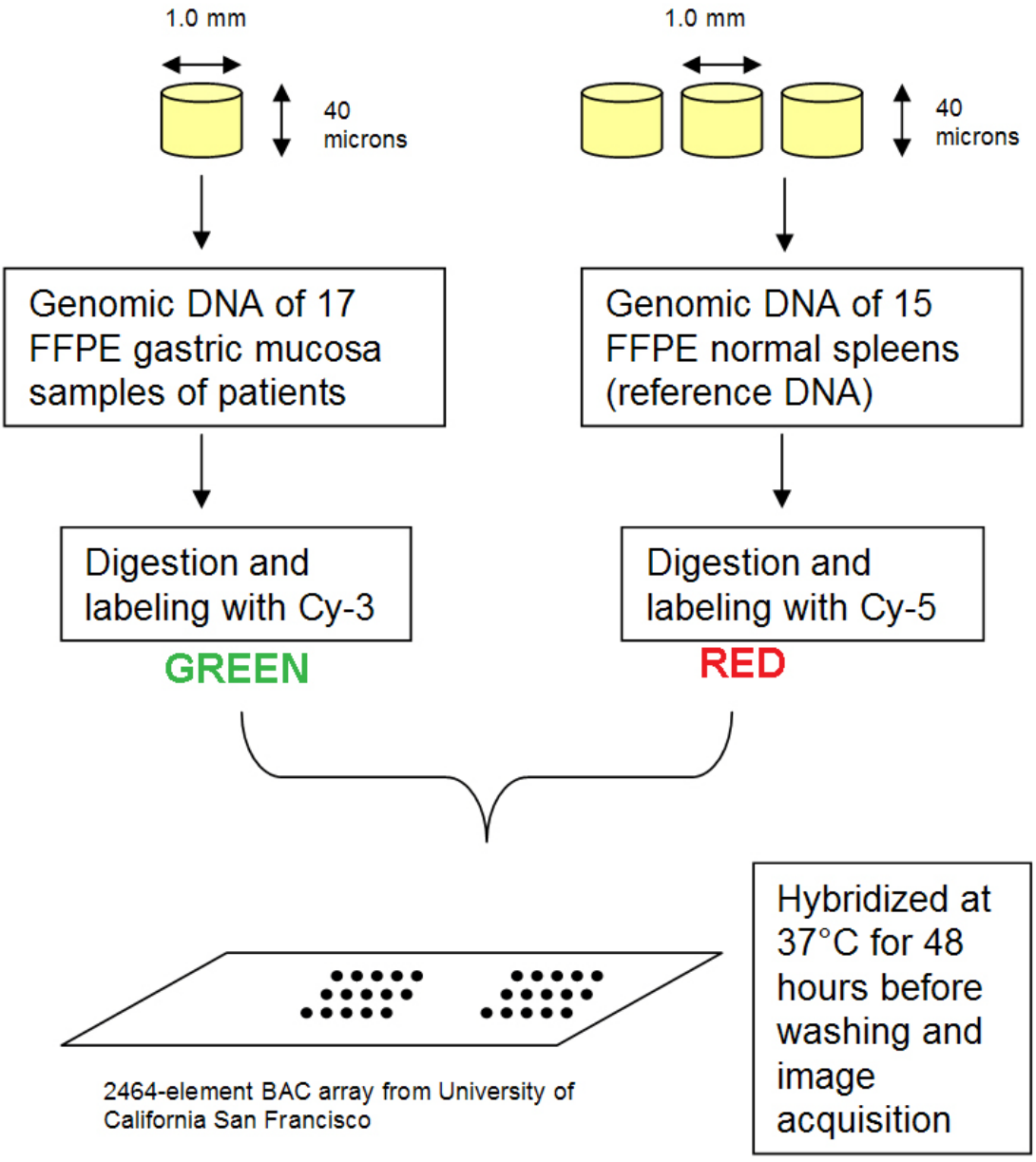


Fig 8. Diagram summarizing the hybridization process

2.7 Imaging and post-processing

We obtained our array images using an Axon GenePix 4000B laser scanner (Molecular Devices Inc.). This is a dual-laser scanning system at wavelengths of 532 nm (green) and 635 nm (red)

Fig 9. Genepix laser scanner

The combined color image was then obtained with green signifying a relative abundance of test gDNA and red a relative deficiency of test gDNA. Yellow would signify relatively equal amounts of both test and reference gDNA (see images in section 3.1.2).

The combined color image was then broken down to its component monochrome images at 532 nm and 635 nm (obtained directly from the scanner). The monochrome images were then rotated through 90 degrees in preparation for post-processing beginning with SPOT and SPROC software.

SPOT is the software developed at UCSF to analyze the array images. SPOT functions to provide statistics about each spot on the array (such as $\log_2$ ratios of the total integrated Cy3 and Cy5 intensities) in addition to performing local background correction for each spot (40). SPROC is the companion program to SPOT that maps each spot on the array to a specific clone and chromosome position, and averages over replicate spots in order to output a final ratio value for each clone on the array (40).

SPROC contains information on a number of clones which have been found by UCSF to be 'bad' clones. These are essentially clones that did not transfer adequately during the manufacture of the array (i.e. when the array was printed on the glass slide at UCSF). Using SPOT and SPROC, a modified SPOT file is first created. This is put through a normalization process using the Statistical Microarray Analysis (SMA) package in the R environment (www.r-project.org). The normalized $\log_2$ (test/reference) ratios are then used as the new input into the modified SPOT file. This new SPOT file is then used to run SPROC again to obtain a final SPOT and SPROC output file for further analysis.

2.8 Problems with the hybridization process

2.8.1 Quality of DNA from FFPE tissue

Numerous reports abound on the difficulty of obtaining good quality DNA from formalin-fixed tissue (41) (42). Although formalin is excellent at preserving the morphological structure of tissues, it is also a crosslinking agent that induces chemical modifications and fragmentation of nucleic acid structures (42). Although the gold standard for molecular analyses remains unfixed fresh or snap-frozen tissues these

preservation methods cannot be used for our study because they do not provide accurate morphological details sufficient to distinguish the histological features of metaplastic and dysplastic mucosa within the stomach.

In order to gauge the quality of our extracted DNA, we ran several gels to determine the degree of fragmentation of the genetic material. From the image in Figure 10 below it is clear that the DNA from FFPE tissue comprised smaller fragments compared to DNA from a blood lymphocyte sample. This was a clear indicator that we could expect poorer results than we had from fresh tumor tissue.

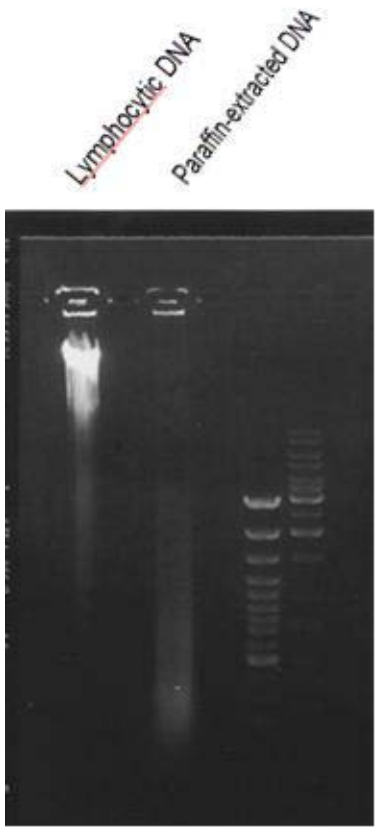


Fig 10. DNA from FFPE tissue comprises significantly smaller fragments. First marker is GeneRuler 100bp DNA Ladder Plus (Fermentas) and the second is GeneRuler 1kb DNA Ladder (Fermentas).

2.8.2 Quality of hybridization results

The procedures for hybridization when we began our study in 2004 were relatively primitive compared to the alternatives for automated hybridizations today. As such there was a steep learning curve in our initial efforts. Our first few attempts at hybridization were unsuccessful in large part due to small oversights in the complicated hybridization or washing process. Examples include loss of the labeled probes at some stage; uneven coverage of the array by the hybridization mixture and increased background noise from particulate contamination.

Fortunately, these obstacles are largely operator-dependent and once we mastered the protocol, there were few further errors.

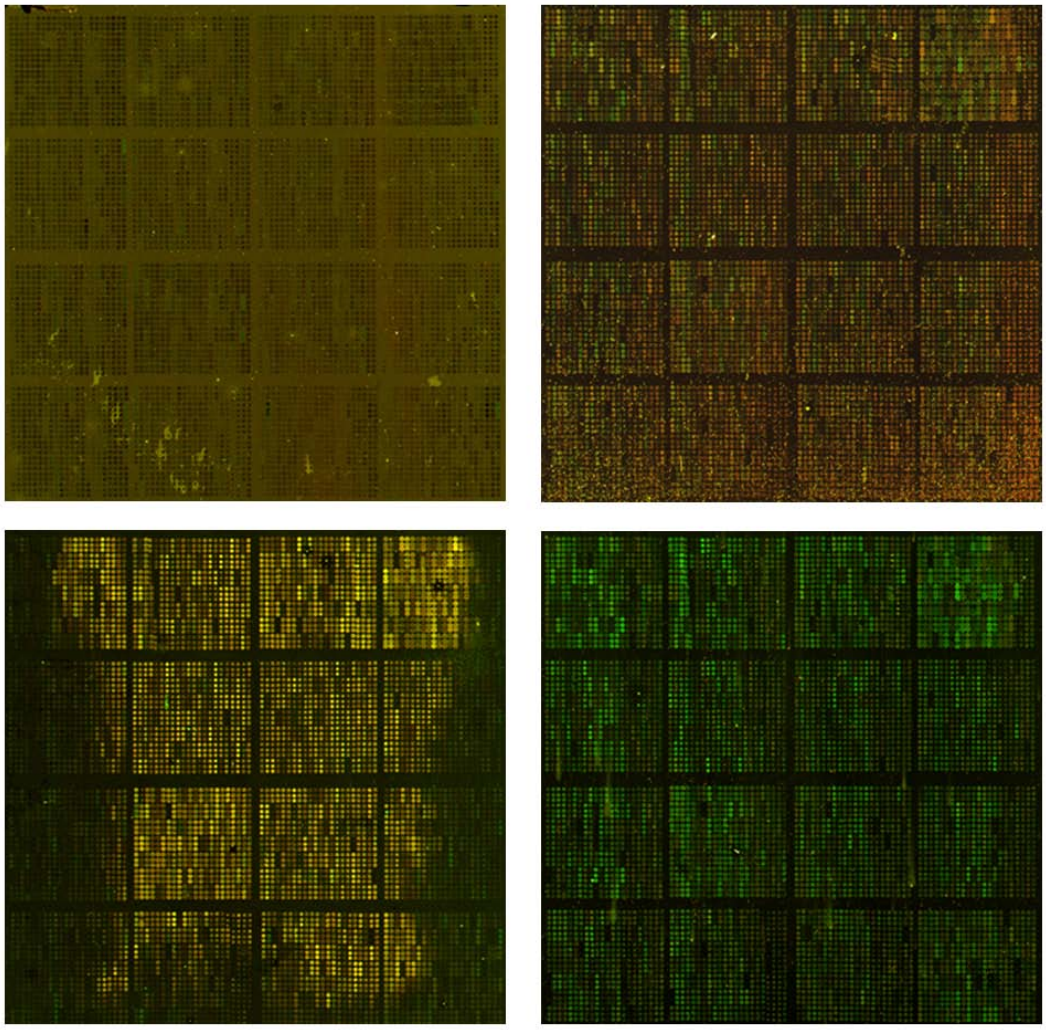


Fig 11. Examples of poor hybridizations

2.9 Determination of threshold

Unlike conventional CGH on metaphase spreads where $\log_2$ (test/reference) values of more than +0.3 signify amplifications and less than -0.3 signify deletions, the determination of significant copy number changes in array CGH is less straightforward. Measurement variation varies from hybridization to hybridization and hence the threshold of one may differ from another.

We adopted the method described by Douglas *et al*. (43). The first step was to establish regions of modal copy number in independent normal versus normal hybridizations. We used our pooled spleen reference DNA for this purpose and performed 3 sets of hybridizations. Based on the autosomal chromosomes, a threshold $\log_2$ ratio value of +/- 0.232 representing the 99% confidence interval of normal copy number was determined. Thereafter, modal regions in subsequent hybridizations involving test versus reference samples were defined by the above threshold, and used to calculate the coefficient of variation and 99% confidence intervals. Log_2 ratios falling above and below these 99% confidence intervals were then deemed as amplifications and deletions.

In order to further refine our data analysis specific to the identification of potential regions of changes, we opted to exclude copy number changes reported by only one or two neighboring clones. We thus required changes in at least 3 contiguous clones before we considered a region of genomic DNA to be amplified or deleted.

2.10 Data analysis and the development of ACAVIS

We discovered that it was difficult to visualize the overall gross changes simply by analyzing the datasets of the 2464 clones in software like Microsoft Excel alone. We were therefore obliged to develop our own software for this purpose.

Array CGH Analysis and Visualization (ACAVIS) is the result of our collaboration with faculty members from Nanyang Polytechnic. The program is written in Java and primarily functions to provide graphical representation of the numerical data from SPOT and SPROC.

The images generated include genome-wide karyograms as well as representations of individual chromosomes. Options exist to view the data as lines or as outliers/points only. In addition, the ability to represent up to 20 different samples in one image at the same time vastly simplifies the search for obvious regions of differences.

In addition to its graphical functions, ACAVIS integrates several statistical functions such as filtering and LOWESS (Locally Weighted Scatter plot Smoothing) which allow us to analyze the data from various perspectives. It can also show the frequencies of gains or deletions as a sidebar on the chromosome.

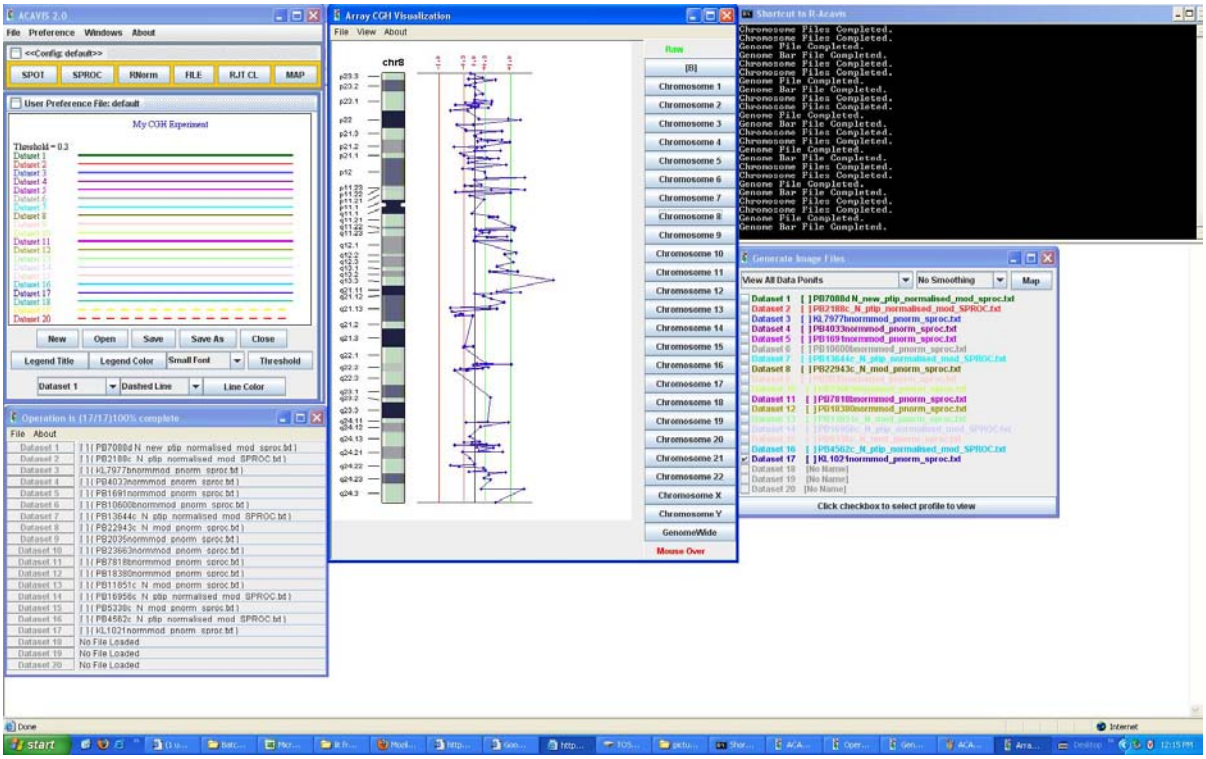


Fig 12. Screenshot of ACAVIS showing the chromosome 8 profile of an individual sample

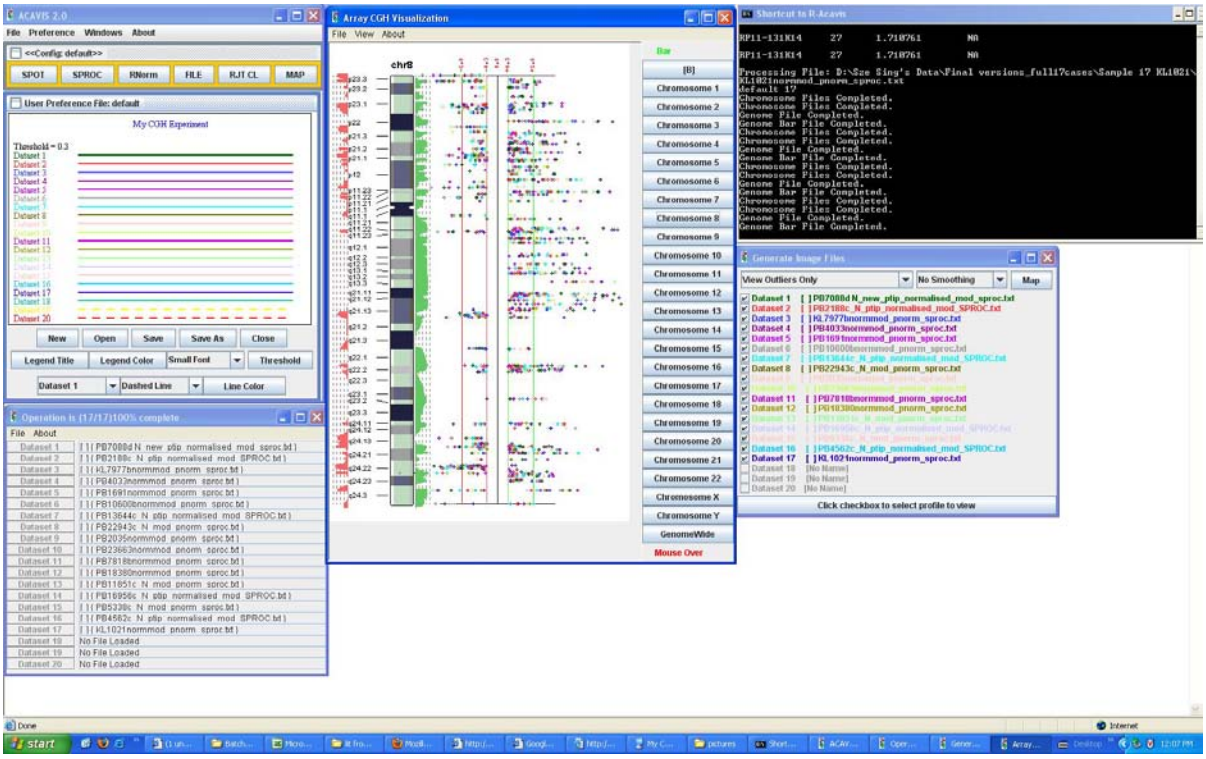


Fig 13. Screenshot of ACAVIS showing the chromosome 8 profile of 17 samples

Chapter 3
Results & Initial Analysis

3.1 Sample results

3.1.1 Sample acquisition results

A total of 57 specimens were obtained and histologically confirmed by our collaborating pathologist. As illustrated in the Table 3, the majority of patients had intestinal-type carcinomas with only one having diffuse-type cancer. Unfortunately, we were unable to obtain further clinical information on the 2 specimens from Malaysia.

Patient Number	Sex	Age	Cancer stage (AJCC/UICC)	Histological type (Lauren's)	Survival (months) *	Recurrence (months) #
1	M	79	IIIA	Mixed	A (29)	N
2	F	77	IIIA	Intestinal	A (31)	N
3	-	-	-	-	-	-
4	M	45	IV	Intestinal	A (78)	N
5	F	80	IIIA	Mixed	A (31)	N
6	M	78	IIIB	Intestinal	D (39)	R (25)
7	M	62	IA	Intestinal	A (60)	N
8	M	63	IIIB	Intestinal	A (51)	N
9	M	81	IIIA	Intestinal	A (31)	N
10	M	63	II	Intestinal	A (48)	N
11	M	78	IIIA	Intestinal	-	-
12	M	62	II	Intestinal	A (85)	N
13	M	81	IIIB	Mixed	-	-
14	M	60	IV	Intestinal	D (11)	-
15	M	69	II	Diffused	D (79)	N
16	F	74	II	Intestinal	A (31)	R (23)
17	-	-	-	-	-	-

* **A** = alive; **D** = deceased

N = no recurrence; **R** = recurrent cancer

Numbers within brackets are months from the date of gastrectomy to survival and recurrence status

Table 3. Details of the 17 patients

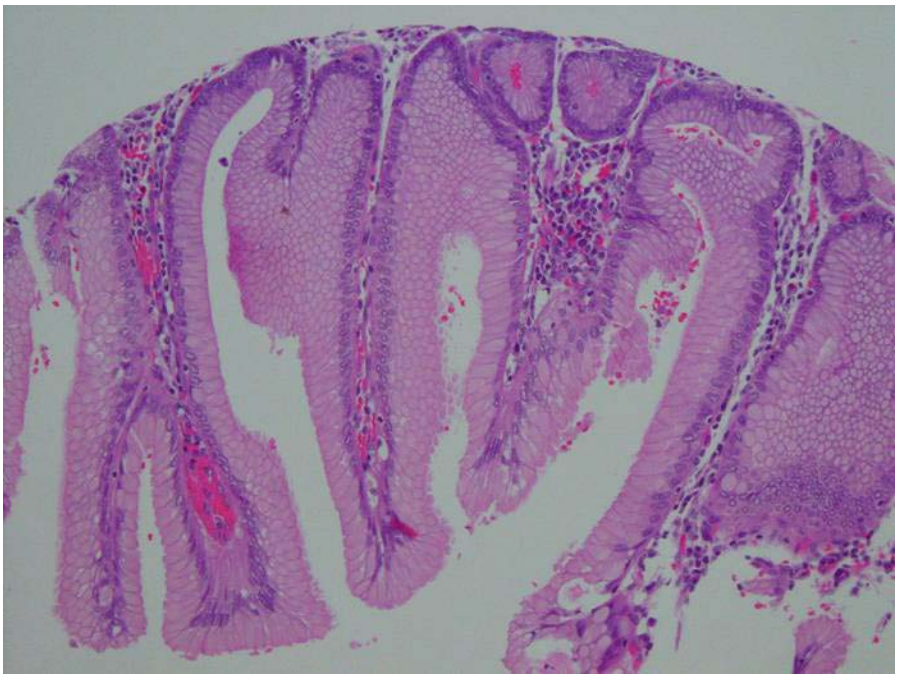
Of the 17 patients, we were unable to recover some tissue types from the archived FFPE blocks. This was, in almost all cases, due to tissue quality issues which were flagged by our pathologist. Table 4 below illustrates the sample types available for hybridization for each patient.

Patient No	Adjacent Normal	Metaplasia	Dysplasia	Tumor	Total
1	√	√	√	√	4
2	√	√	√	√	4
3	√	×	√	√	3
4	√	√	×	√	3
5	√	√	√	√	4
6	√	√	√	√	4
7	√	√	√	√	4
8	√	√	√	√	4
9	√	√	√	√	4
10	√	√	×	√	3
11	√	√	√	√	4
12	√	×	√	√	3
13	√	×	√	√	3
14	√	×	√	√	3
15	√	×	√	√	3
16	√	×	×	√	2
17	√	×	×	√	2
					57

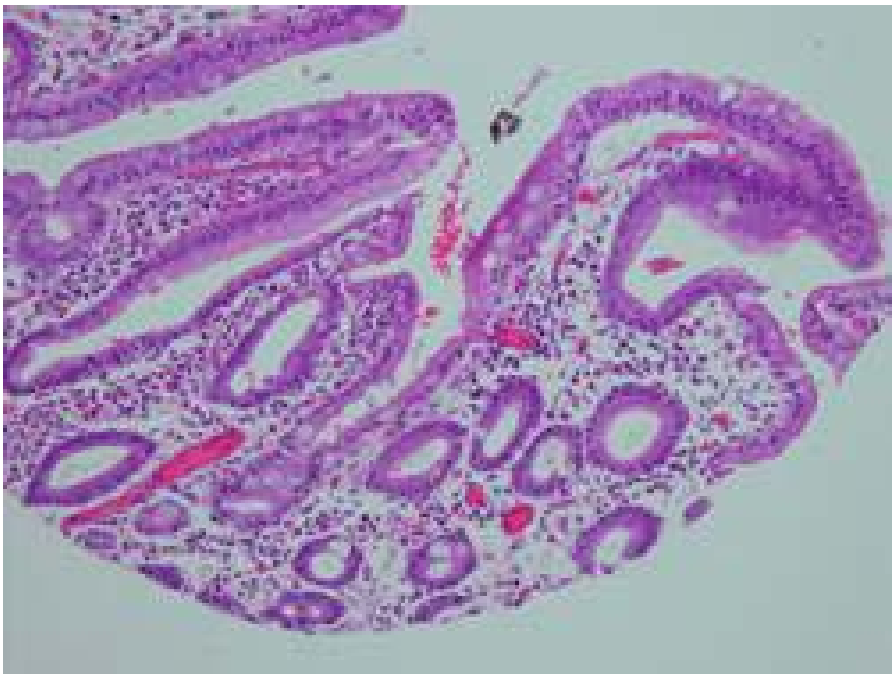
Table 4. Specimens by tissue type

Given that one patient had diffuse-type carcinoma, 2 had indeterminate pathology and 3 had mixed-type pathology by Lauren classification, we were hesitant to include them in our analysis since our initial goal was to investigate the genomic changes along the Correa pathway of intestinal-type carcinogenesis.

However, as will be evident in the later analysis, stratification by Lauren type will prove to be of lesser import than our eventual findings.

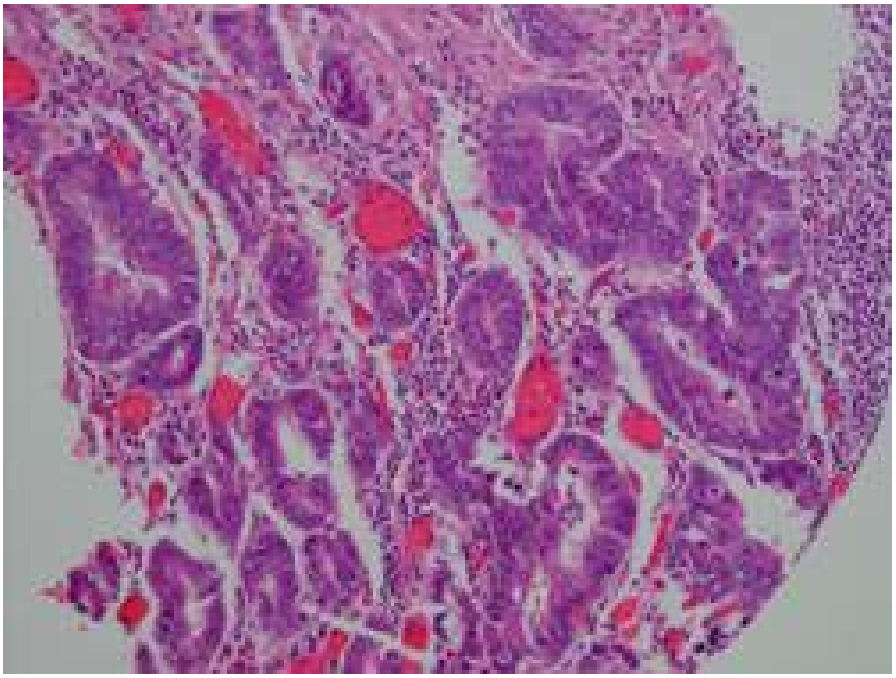


(a) Adjacent normal (top slice of a 40 micron section)

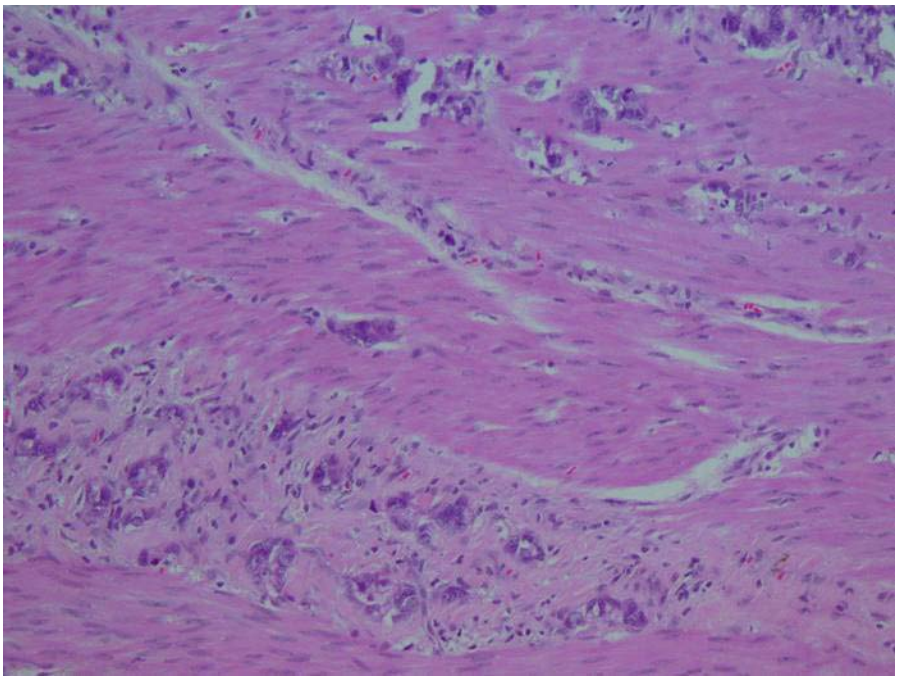


(b) Intestinal Metaplasia (bottom slice of a 40 micron section)

Fig 14(a) & (b).Histology from 40 micron sections



(c) Dysplasia (close up view of a top slice)



(d) Carcinoma (close up view)

Fig 14 (c) & (d) (cont.) Histology from 40 micron sections

3.1.2 Hybridization images

Figures 15, 16 and 17 are typical hybridization images obtained for different tissue types immediately after acquisition with the Axon GenePix scanner:

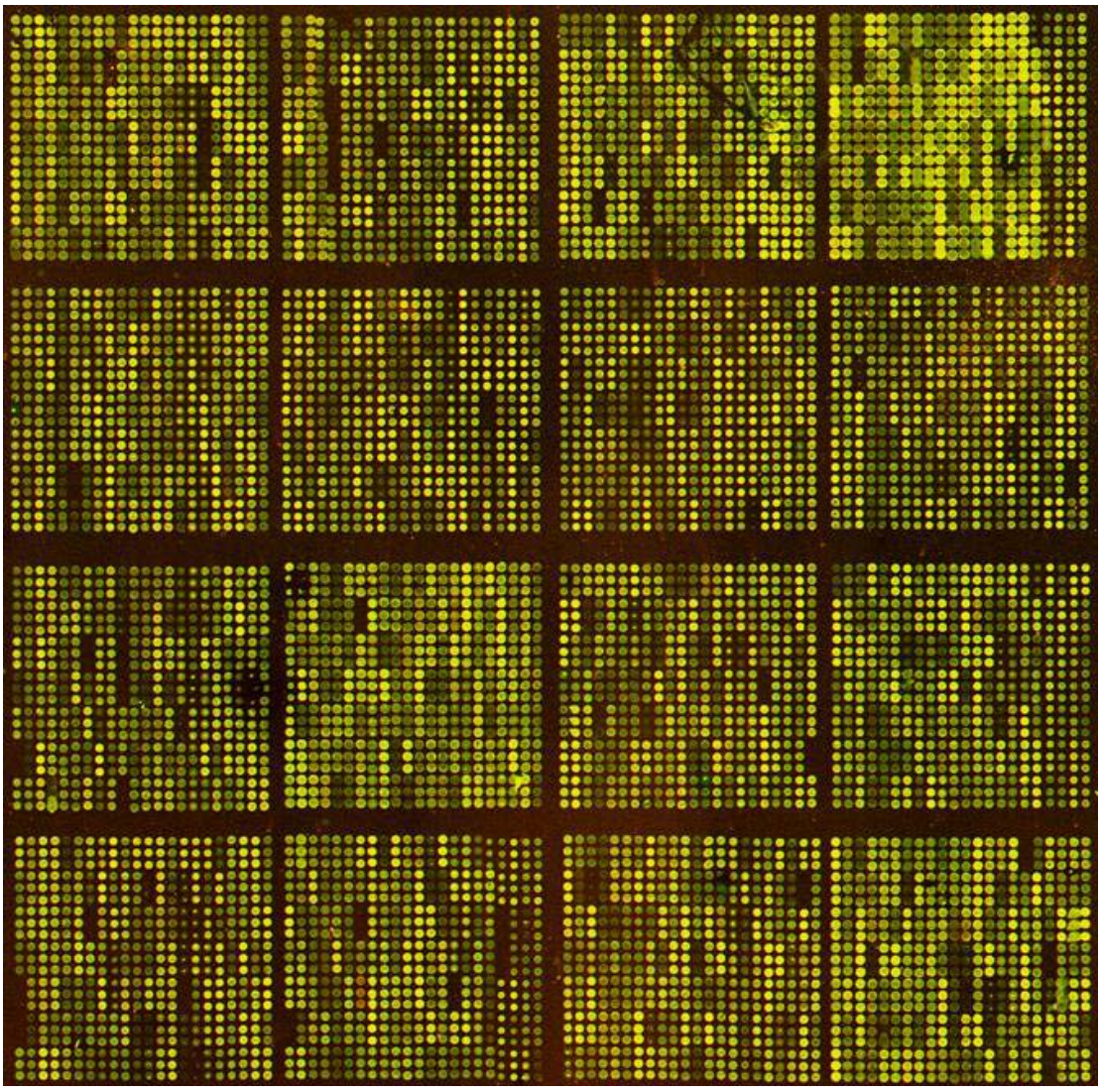
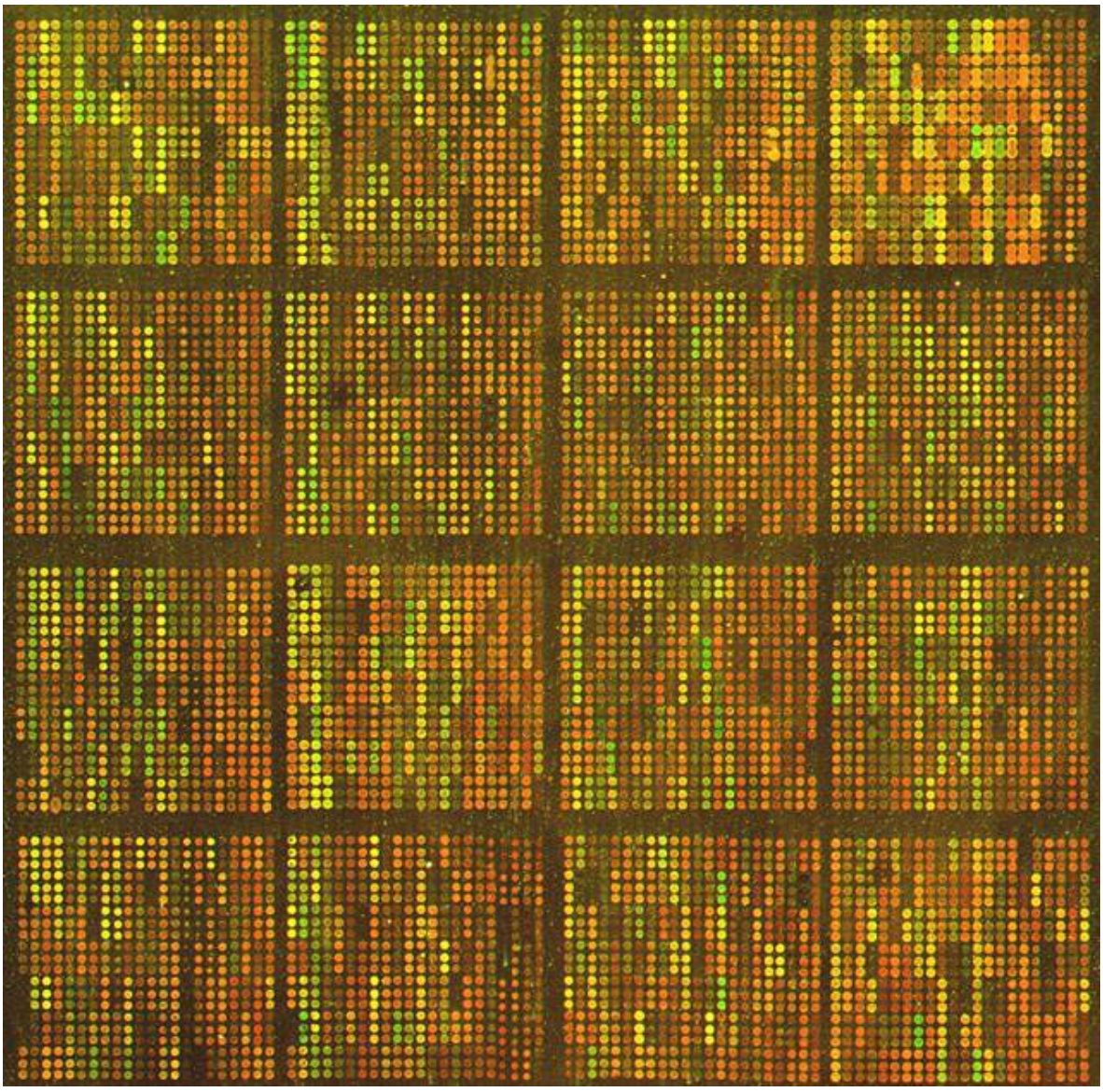


Fig 15. Hybridization image of lymphocyte normal versus pooled spleen reference showing no obvious red or green areas to suggest copy number abnormalities



Fog 16. Hybridization image of adjacent histologically normal gastric mucosa of a gastric cancer patient versus pooled spleen reference

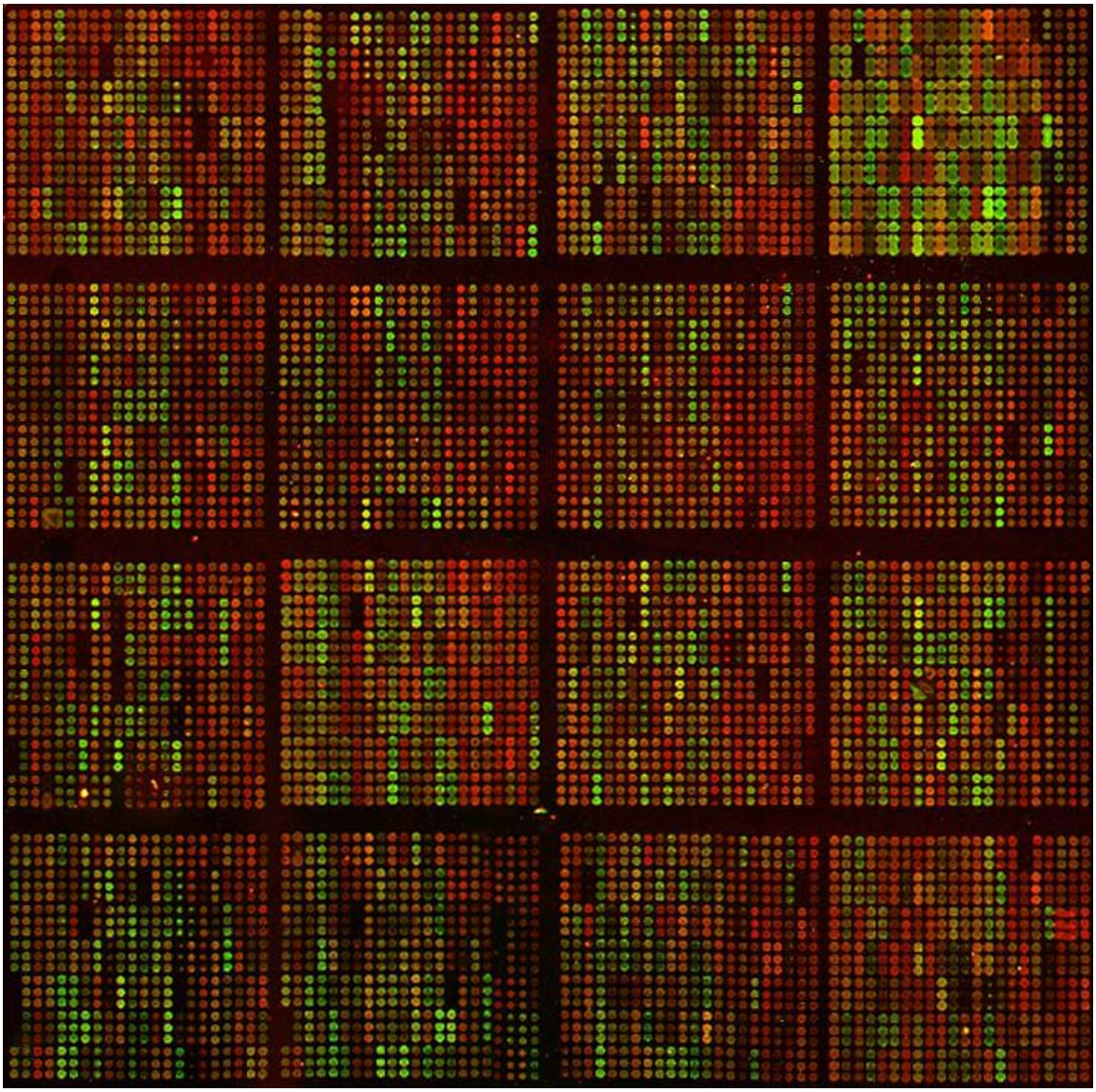


Fig 17. Hybridization image of overt gastric carcinoma versus pooled spleen reference

3.1.3 Image processing

The SPOT and SPROC software required the images to be separated into individual 532nm and 635 nm intensities before the values could be entered. SPOT also read the data from each clone in a horizontal fashion thus necessitating rotation of the image. Figures 18, 19 and 20 are typical hybridization images at various stages of post-processing:

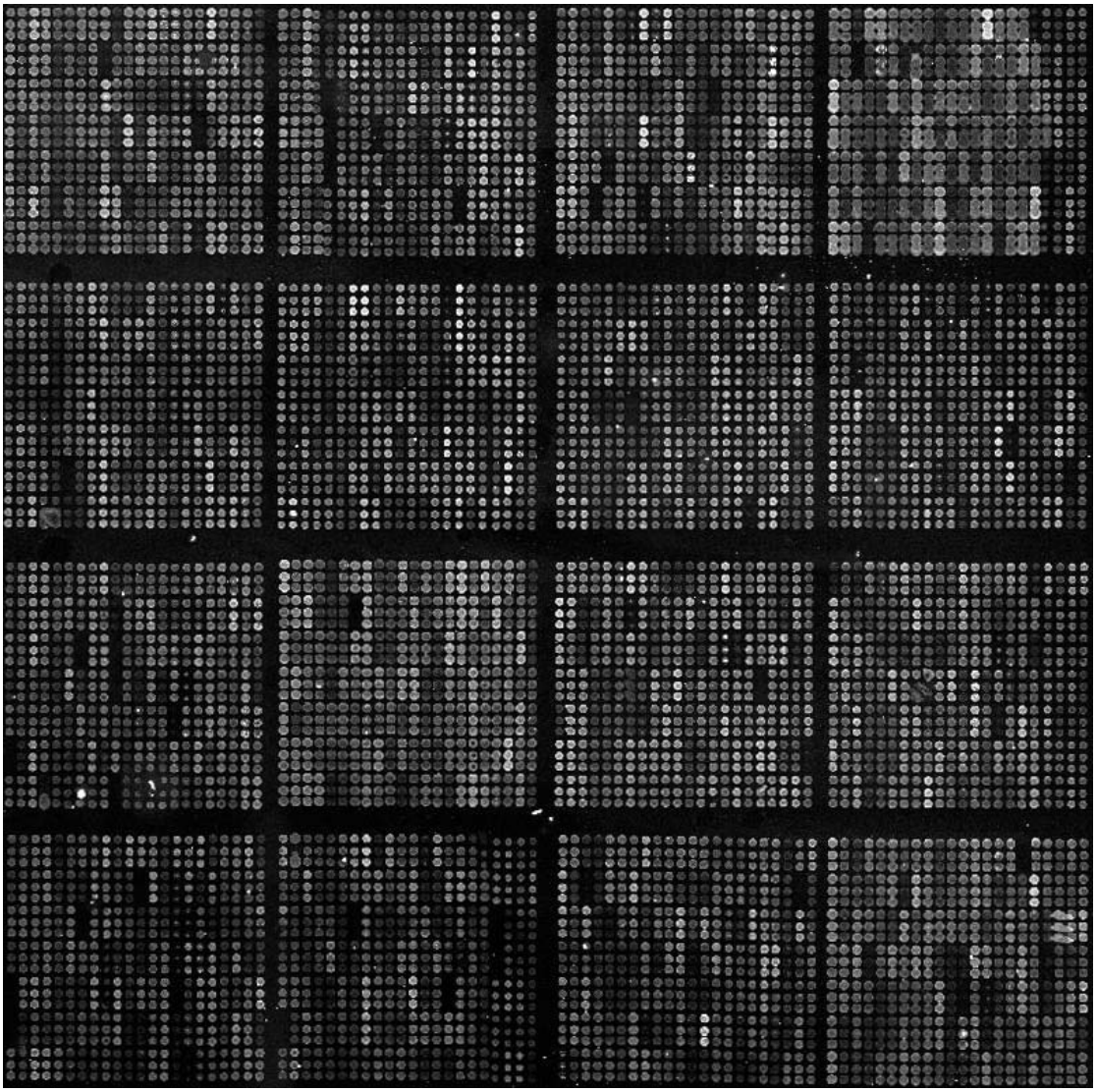


Fig 18. Single channel (Cy3) monochrome image

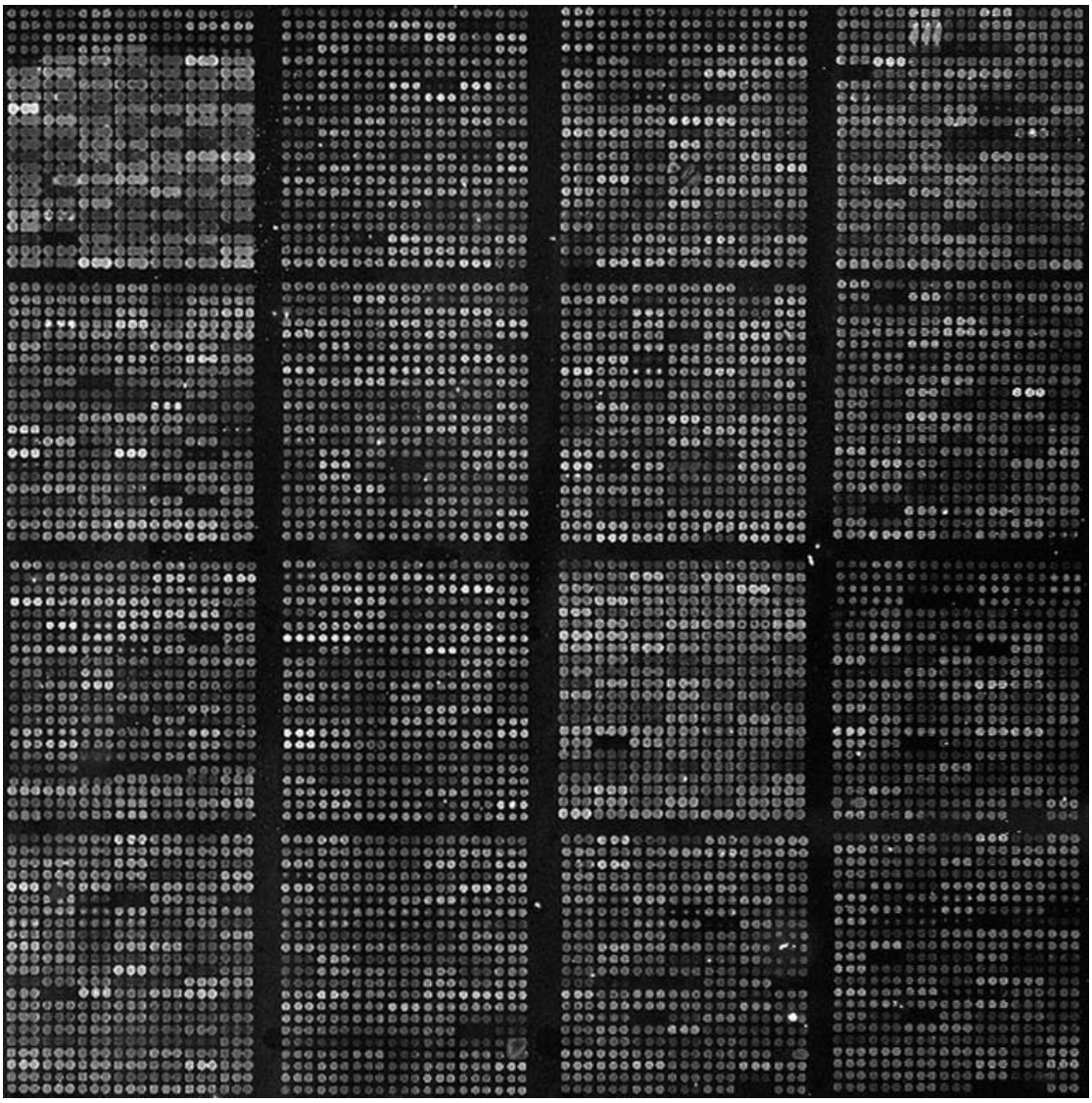


Fig 19. Single channel (Cy3) monochrome image after rotation with Adobe Photoshop

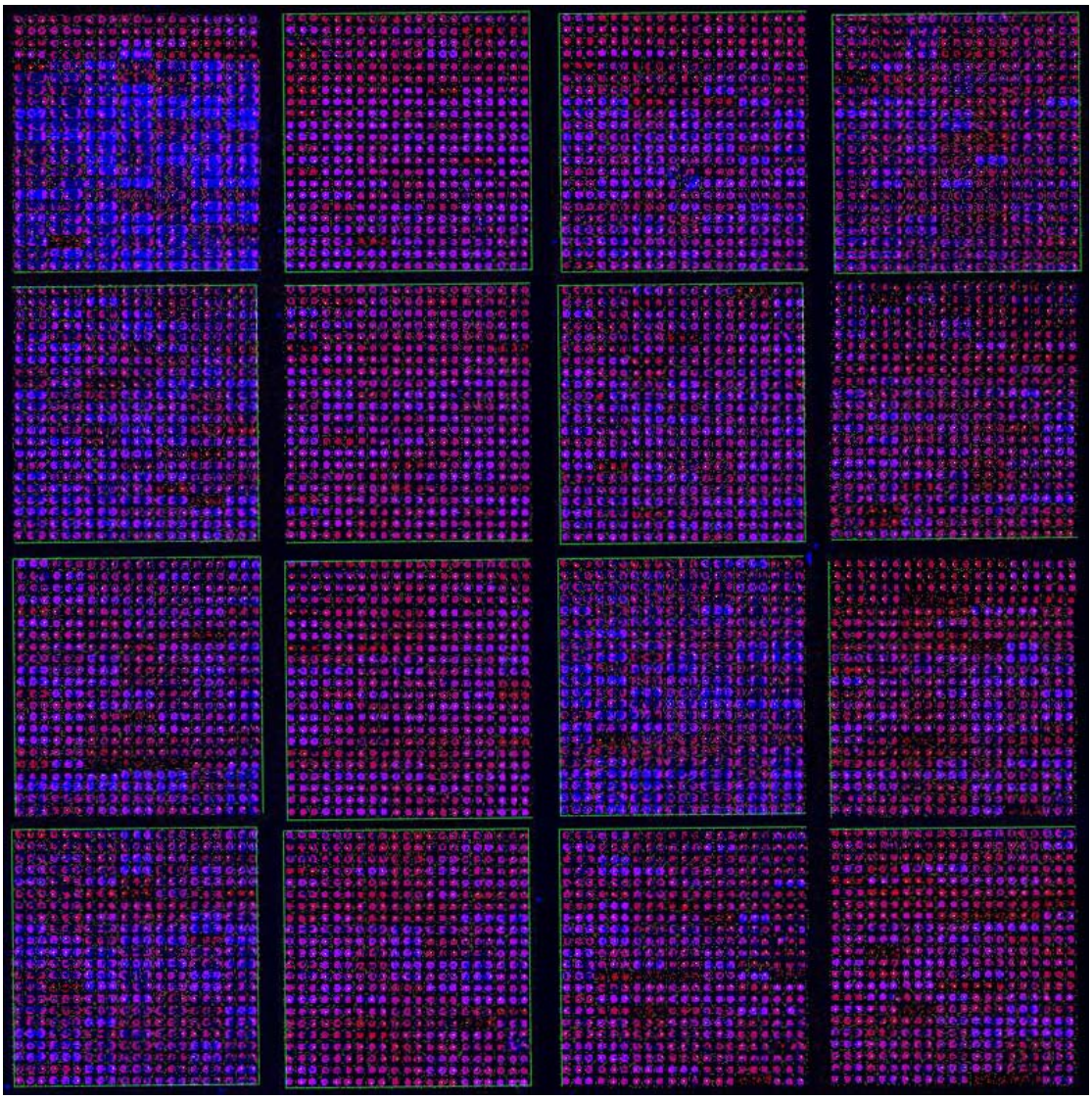


Fig 20. Image after processing with SPOT

3.1.4 Conversion of image data to copy number values expressed as Log_2 ratios

We used SPOT and SPROC (from the UCSF Microarray Core website) to convert the intensity data from the GenePix scanner to numerical data and Log_2 ratios. The screenshots in Figures 21 and 22 show the typical output from these programs.

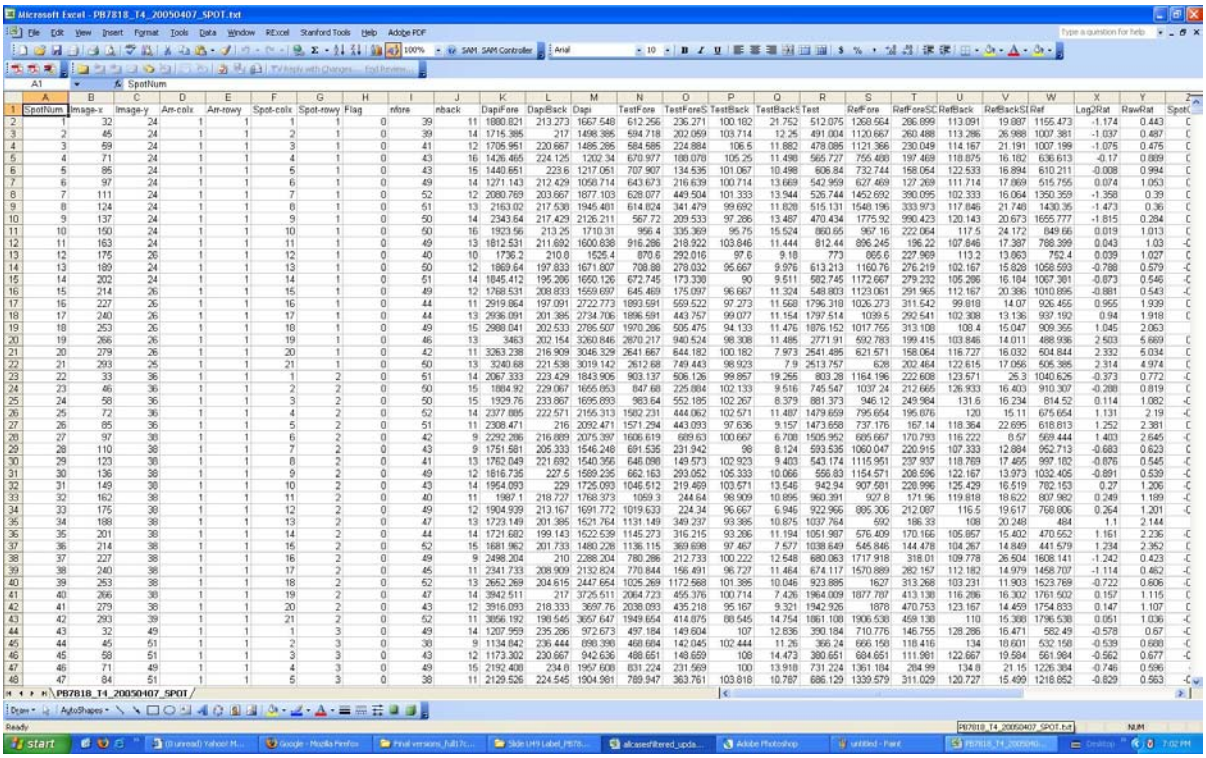


Fig 21. Screenshot showing a typical SPOT output in Microsoft Excel format

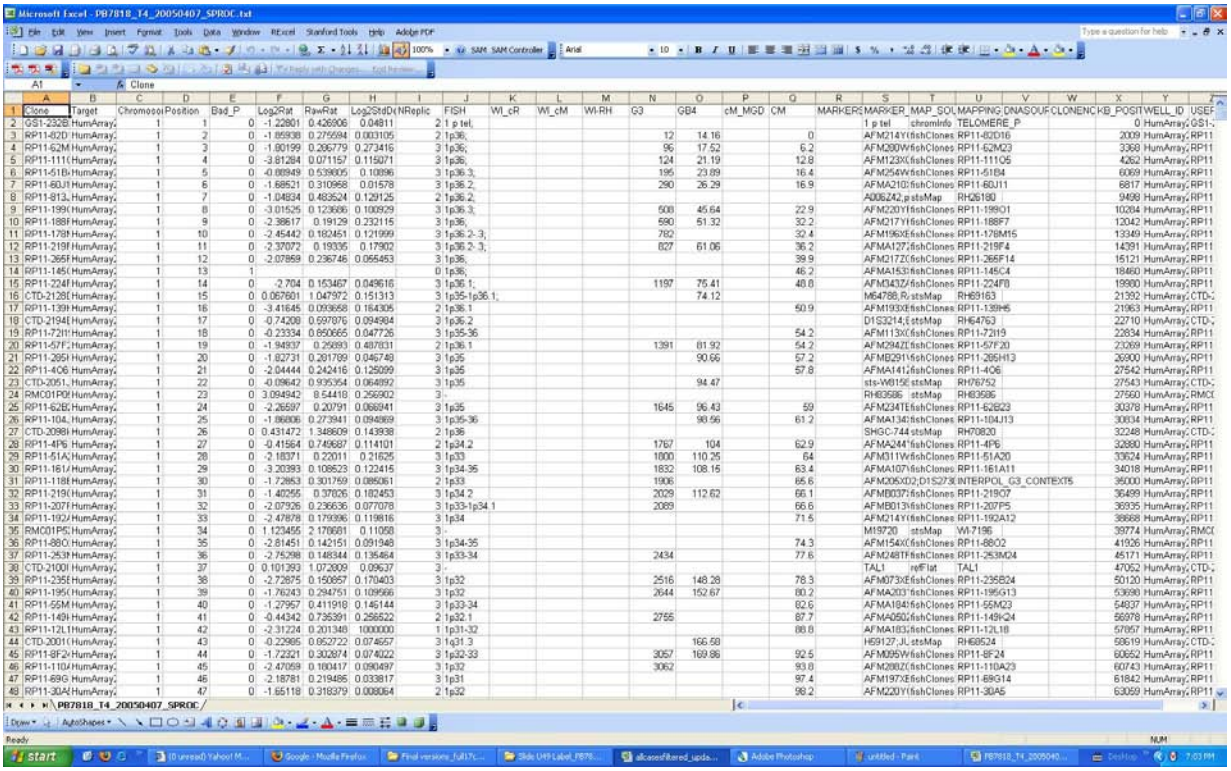


Fig 22. Screenshot showing a typical SPROC output in Microsoft Excel format

As mentioned in section 2.7 earlier 'bad' clones were removed at this stage before we applied the Statistical Microarray Analysis (SMA) package in the R environment before regenerating the final SPOT and SPROC output for further analysis.

3.1.5 Conversion of data to graphical representation of Copy Number

Figures 23-27 illustrate the typical graphical output from ACAVIS in the line format and outlier format for both genome-wide karyograms and individual chromosomes.

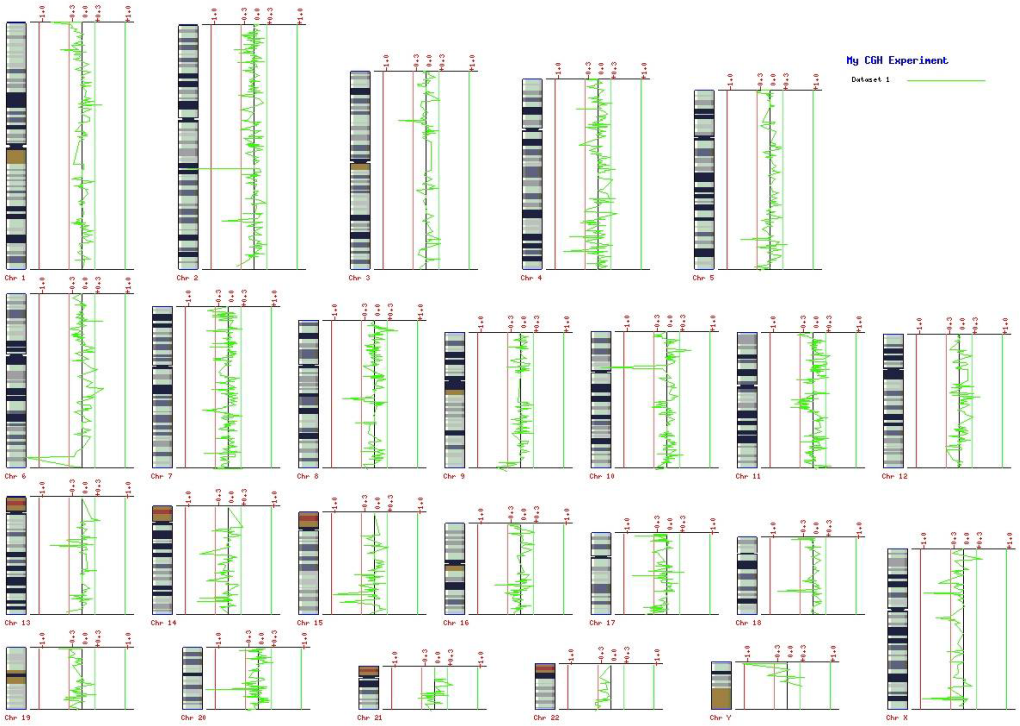


Fig 23. Genome-wide karyogram of lymphocyte normal versus pooled spleen reference showing minimal copy number changes well within what is acceptable as normal

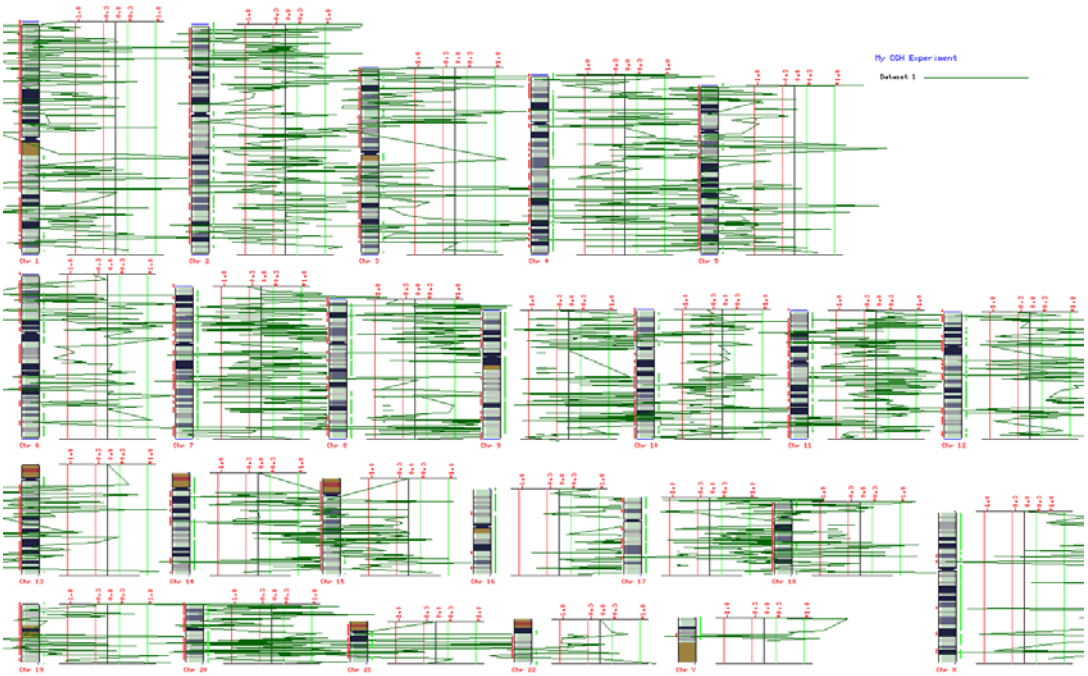


Fig 24. Genome-wide karyogram of carcinoma vs. pooled spleen reference showing gross abnormalities in copy number

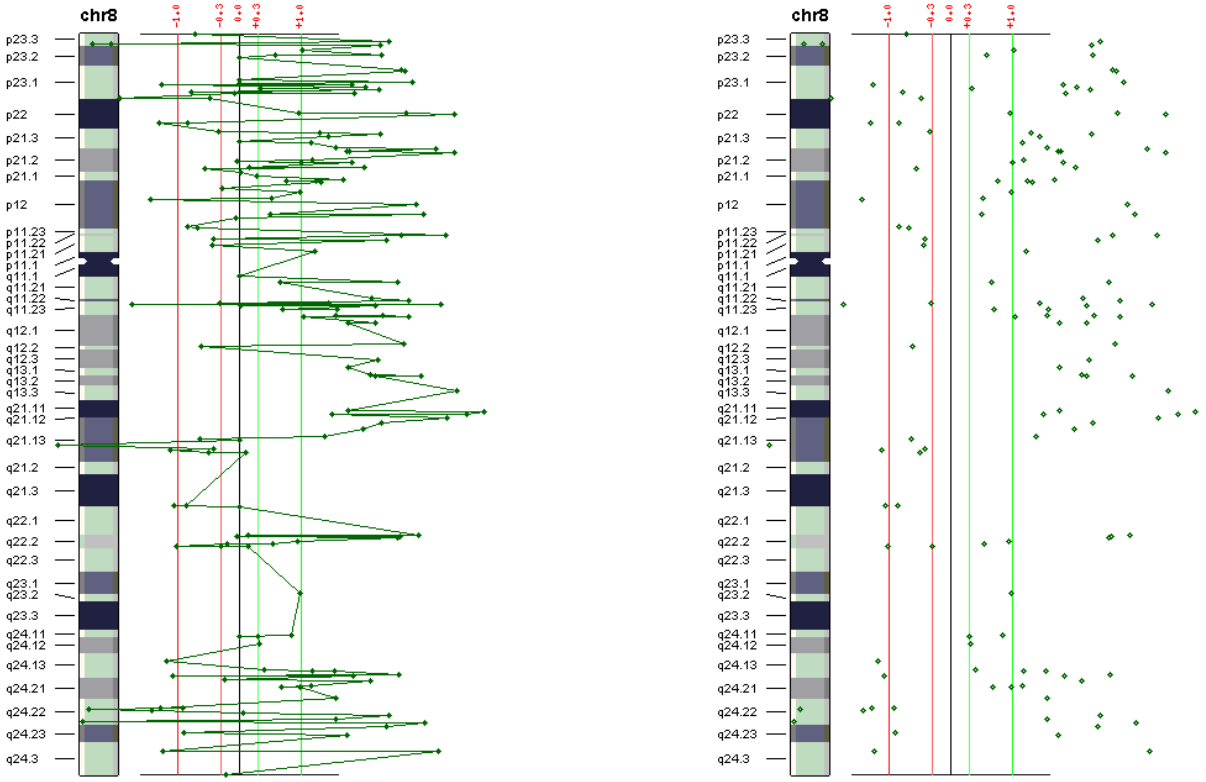


Fig 25. Magnified view of chromosome 8 in a carcinoma vs. pooled spleen reference in line format (left) and in outlier format (right)

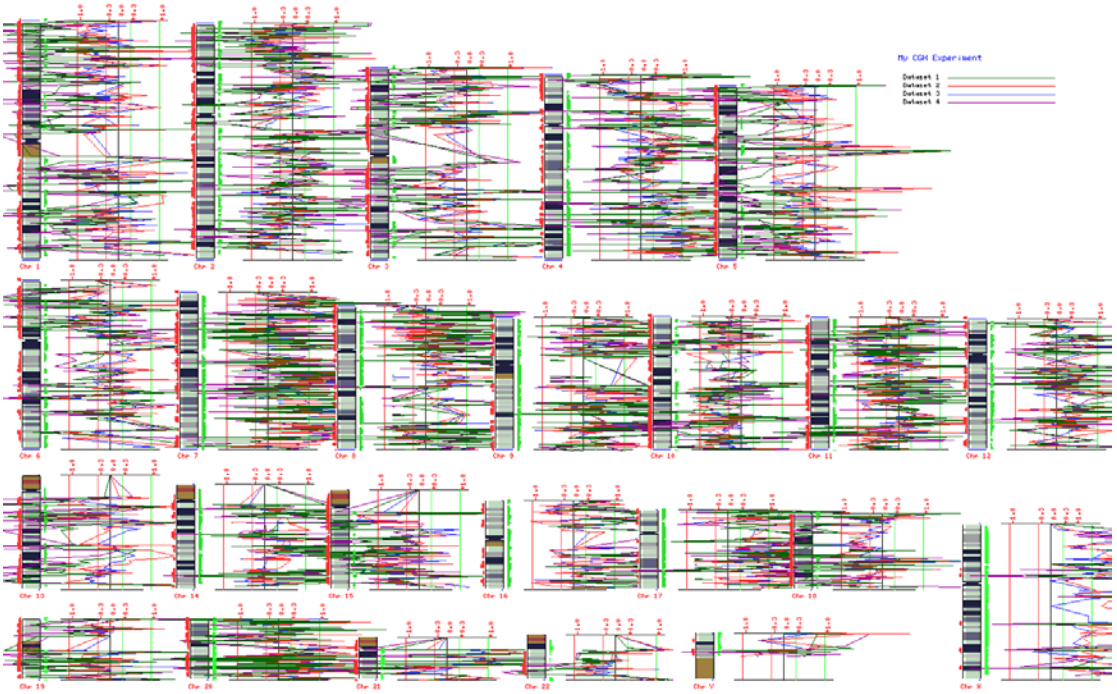


Fig 26. Combined genome-wide karyogram of 4 hybridizations from the same patient. Green represents adjacent normal; Red represents metaplasia; Blue represents dysplasia; Purple represents carcinoma

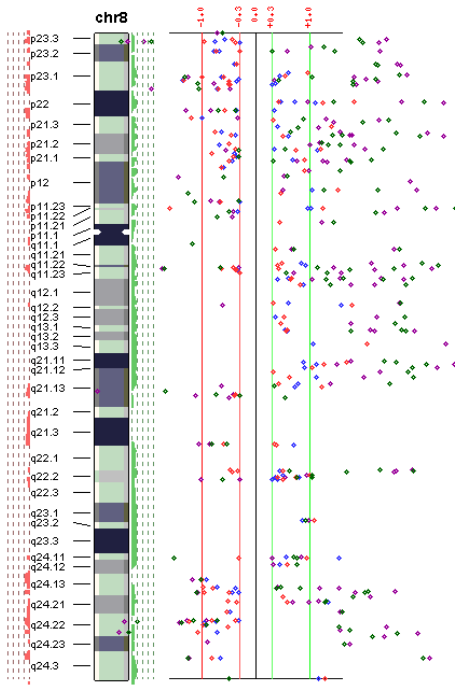


Fig 27. Magnified chromosome 8 (in outlier format) from the preceding karyogram

3.2 Combined results for all 17 patients

57 samples were successfully hybridized versus the pooled spleen reference as shown in the Table 5.

Patient	Normal	Metaplasia	Dysplasia	Tumor
1	hyb	hyb	hyb	hyb
2	hyb	hyb	hyb	hyb
3	hyb	n/a	hyb	hyb
4	hyb	hyb	n/a	hyb
5	hyb	hyb	hyb	hyb
6	hyb	hyb	hyb	hyb
7	hyb	hyb	hyb	hyb
8	hyb	hyb	hyb	hyb
9	hyb	hyb	hyb	hyb
10	hyb	hyb	n/a	hyb
11	hyb	hyb	hyb	hyb
12	hyb	n/a	hyb	hyb
13	hyb	n/a	hyb	hyb
14	hyb	n/a	hyb	hyb
15	hyb	n/a	hyb	hyb
16	hyb	n/a	n/a	hyb
17	hyb	n/a	n/a	hyb

Table 5. 57 hybridizations from the 17 patients

3.2.1 Similarity of copy number profiles between the tissue types

The combined results of all 17 patients yielded an interesting pattern. All the 4 tissue types in each patient tended to appear highly similar in terms of the general trend of amplifications and deletions. This result was consistent both in graphical format and in the form of raw data as seen in the Figures 28 and 29.

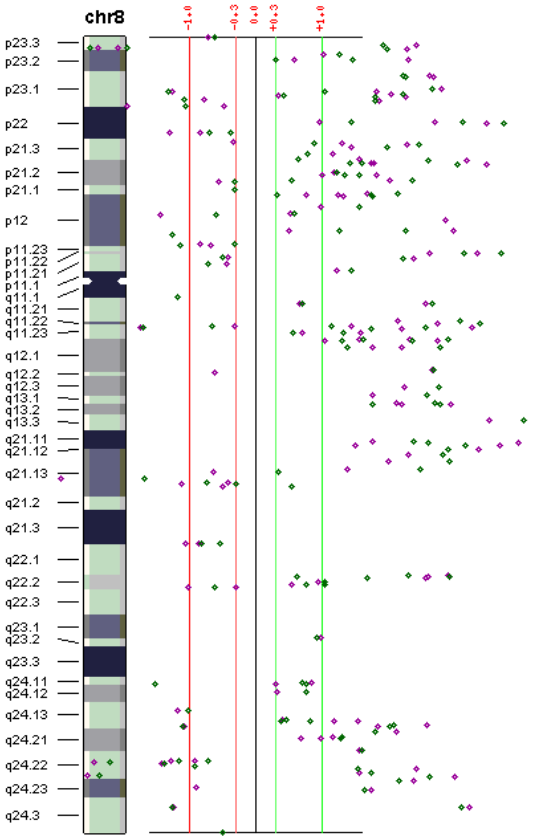


Fig 28. Chromosome 8 profiles of adjacent normal (green) and cancer (purple) in one patient showing a similar pattern of copy number abnormalities

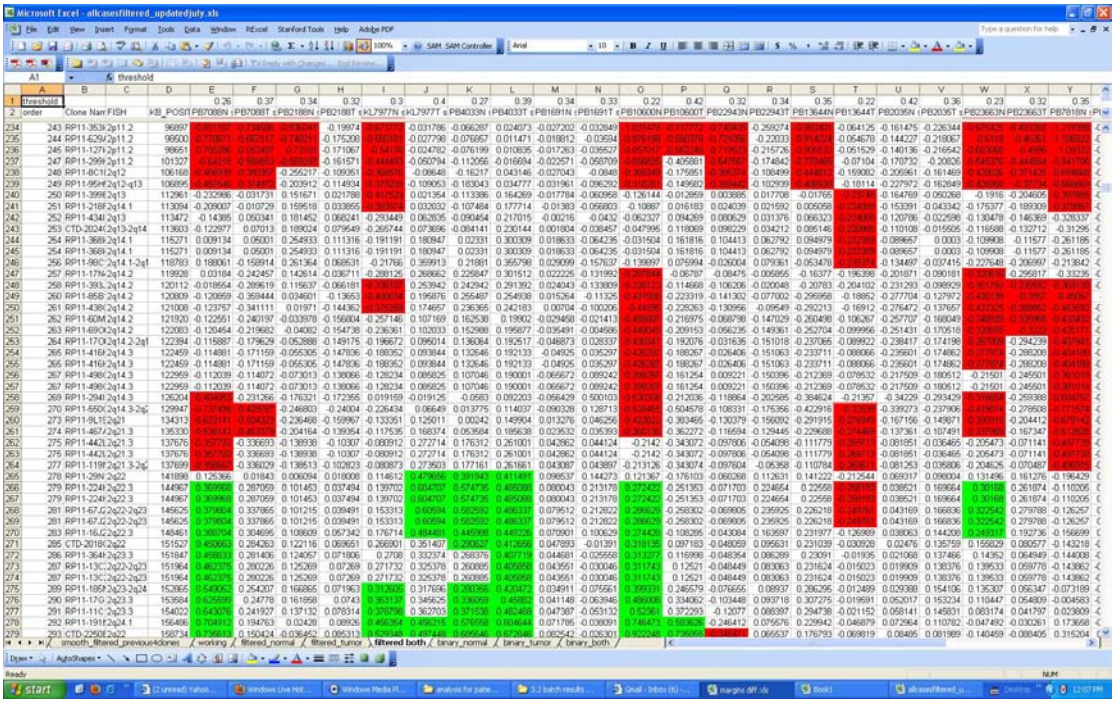


Fig 29. Screenshot of Excel spreadsheet showing similar areas of copy number abnormalities in 17 adjacent normal and 17 cancer samples.

3.3 Analysis of the combined results for 17 patients

3.3.1 Combined karyogram of all 17 patients

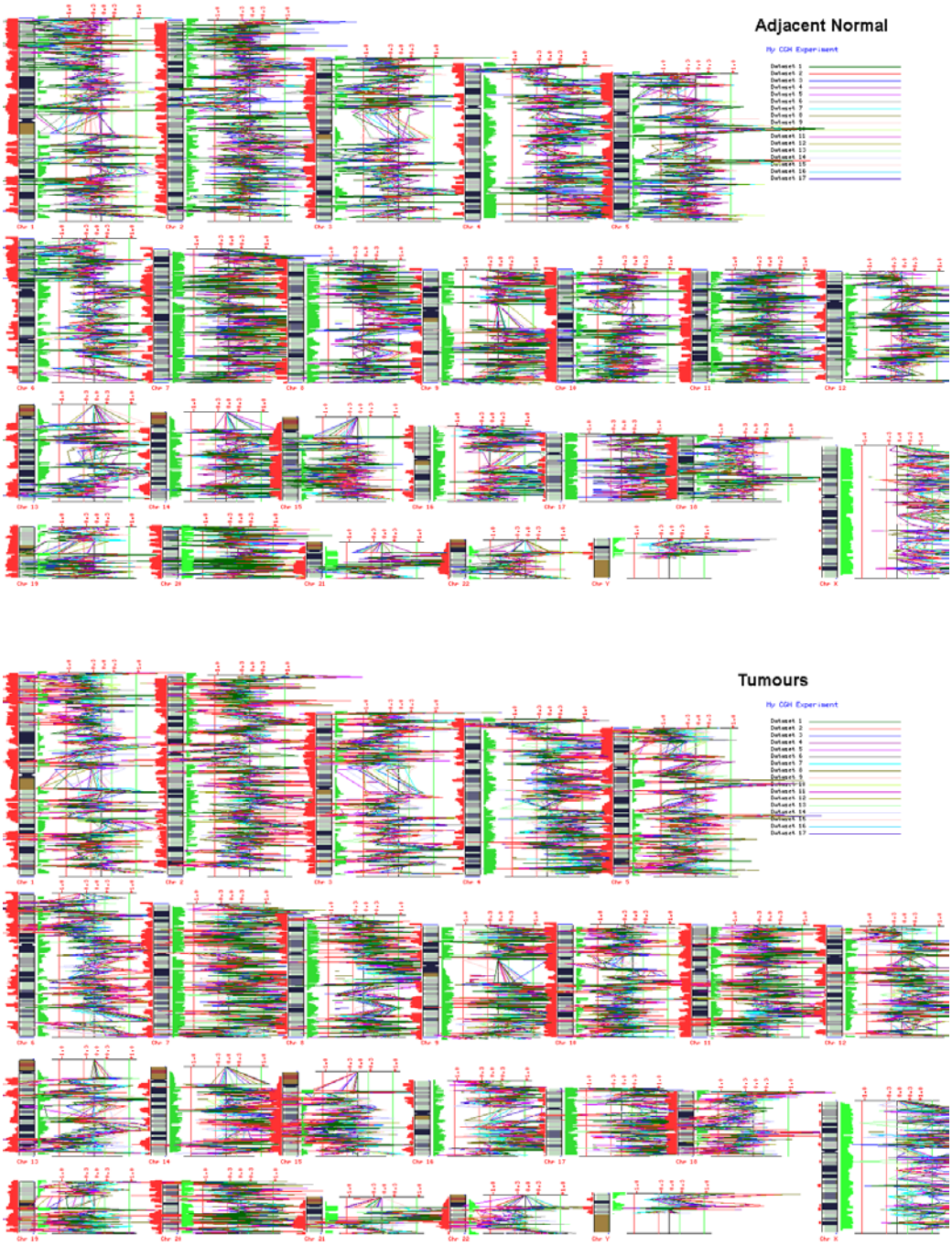


Fig 30. Genome-wide karyograms of adjacent normal (above) and carcinoma (below) for all 17 patients shows similarities although the sheer amount of overlapping data precludes close comparison on this view.

In order to better examine this finding of widespread similarity, we decided to focus on 2 tissue types instead of 4. Our reasons were:

1. All 17 patients had adjacent normal and tumor tissues allowing a greater sample size as a basis for comparison
2. The theoretical difference between tissue types should be greatest between adjacent normals and tumors.

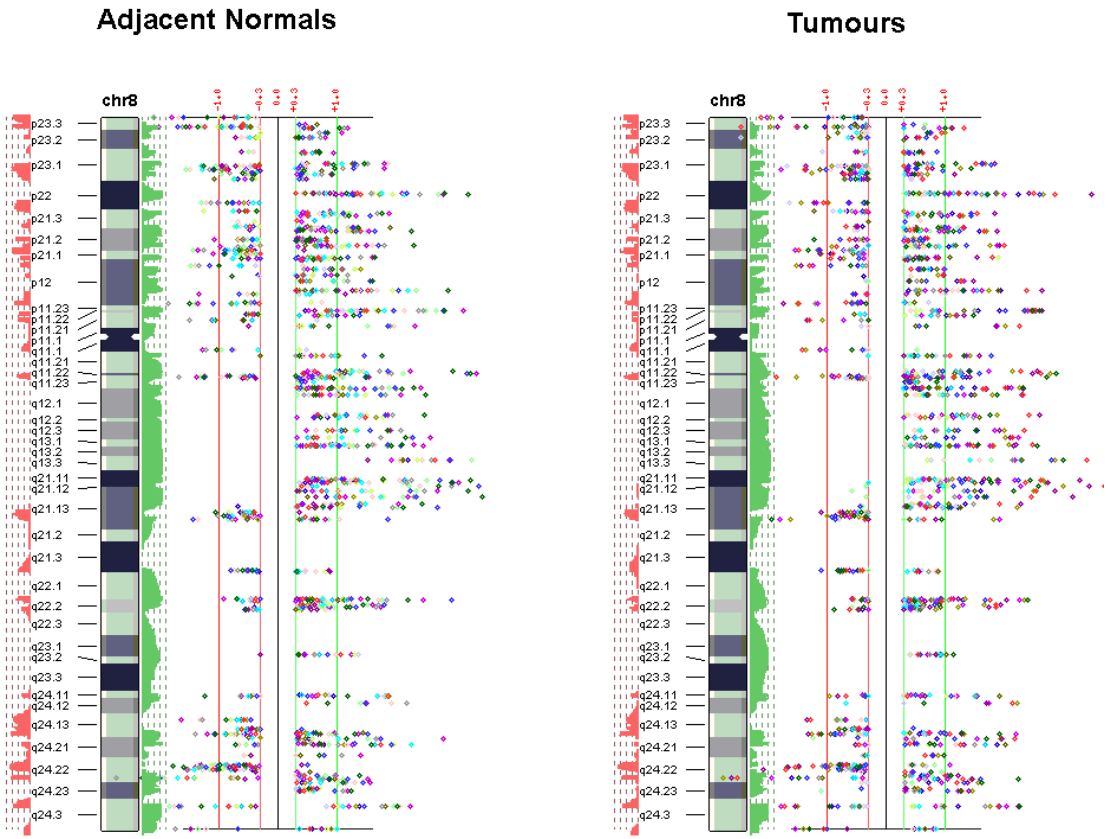


Fig 31. Magnified view of chromosome 8 for all 17 patients shows similar copy number changes between adjacent normals and carcinoma. The green and red bars represent the frequency of the copy number abnormality occurring with values ranging from 0 to 17 patients. There are 17 colors of dots representing the 17 patients.

3.3.2 Additional Tables & Graphs for all 17 patients

The bar charts in Figure 32 demonstrate that the similarity of adjacent normal and tumor tissue types occurs across the entire genome in our group of 17 patients.

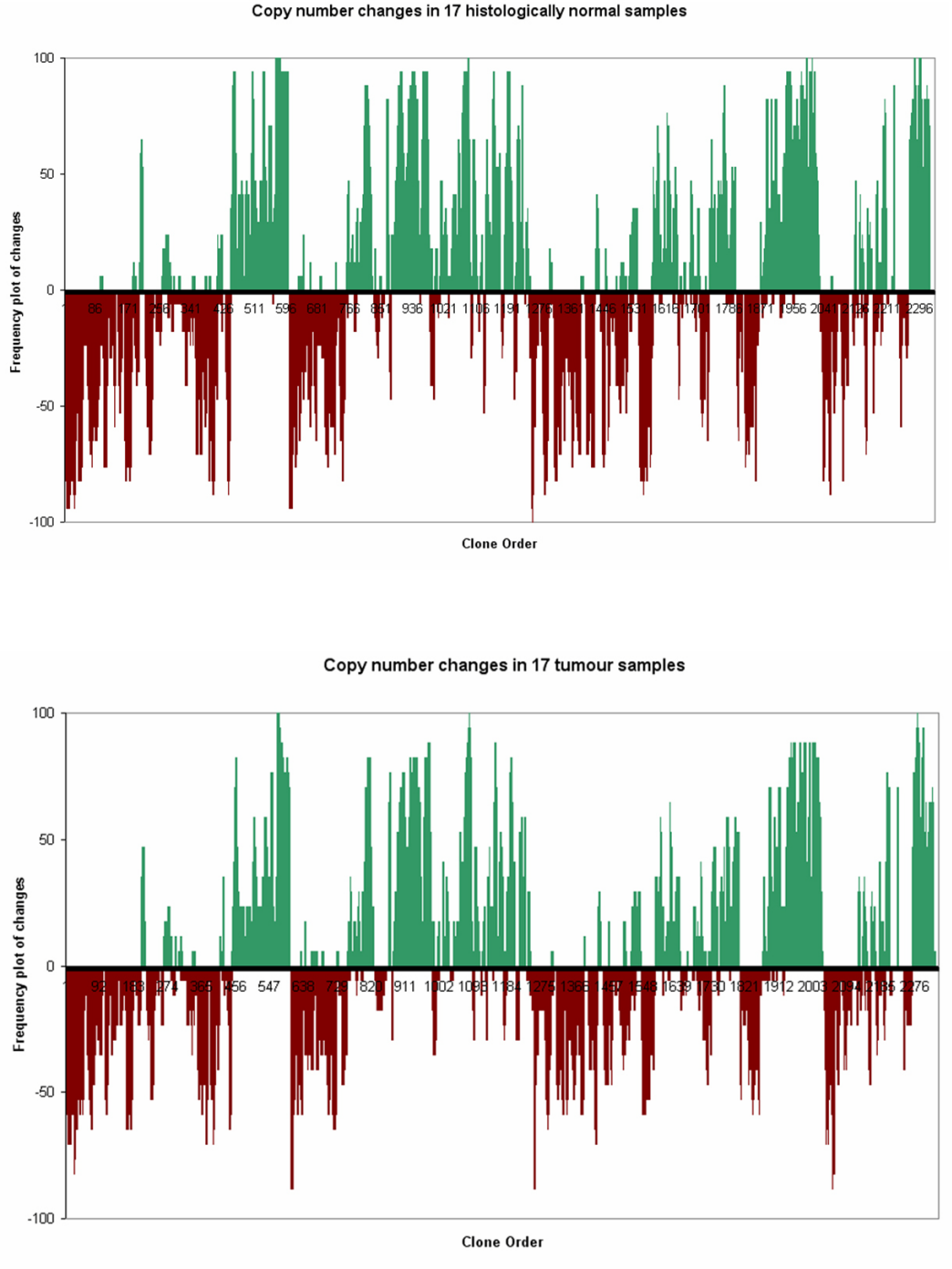


Fig 32. Bar charts of clone position on the x-axis versus % frequency (out of 17) on the y-axis

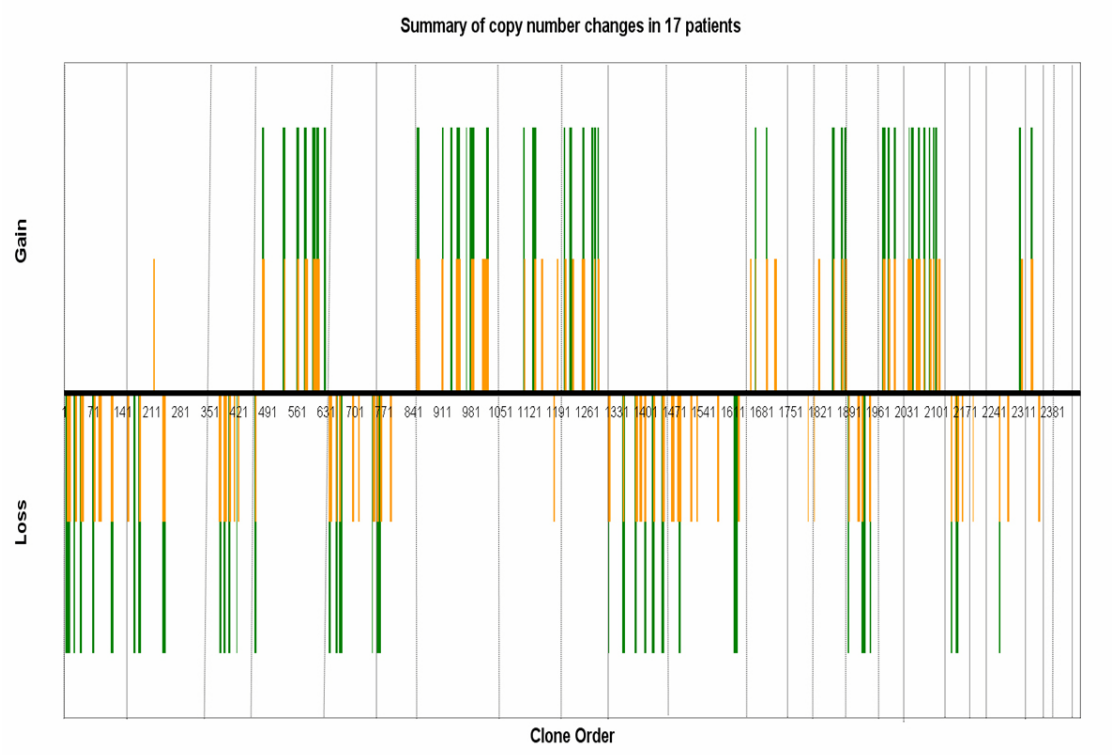


Fig 33. Bar chart summarizing the copy number changes present in≥ 50% of 17 patients. Changes in adjacent normal mucosa appear as short orange bars and in tumor mucosa as long green bars. (Bar lengths do not denote frequency of occurrence.) The changes are ordered according to clone order from chromosome 1 to Y. Grey vertical lines demarcate chromosomes

The chart in Figure 33 illustrates that with few exceptions (e.g. position 211) the changes that occur in a majority of the 17 adjacent normals tend to be mirrored in the corresponding 17 cancers as well and vice versa.

Table 6 on the next page summarizes the regions of similar copy number changes in both tumors and adjacent normal gastric mucosa. Examination of chromosome 8q, which we know to be highly amplified in gastric cancer, reveals that at least 3 cytobands are also amplified in adjacent normals.

Region	Cytoband	Location (KB)	Δ	Region	Cytoband	Location (KB)	Δ	Region	Cytoband	Location (KB)	Δ
1	1pterp34.2	0-41926	Del	32	8q11.21q11.23	52199-53073	Amp	63	15q14	31731-32651	Del
2	1p31.3p32.3	53698-60743	Del	33	8q12.1q21.13	55552-79384	Amp	64	15q22.2	58459-58647	Del
3	1p31.3	63059-63825	Del	34	8q22.1q22.2	98877-99005	Amp	65	15q22.2q22.31	61304-63585	Del
4	1p13.2p21.2	101060-111807	Del	35	9p22.3p23	10432-14333	Amp	66	15q23q24.1	65199-70593	Del
5	1q23.1q23.3	156282-159634	Del	36	9p21.1p21.2	27537-32970	Amp	67	15q25.1	76260-76930	Del
6	1q32.1	199369-202893	Del	37	9q21.32q21.33	80348-85079	Amp	68	15q25.1q25.3	82928-85888	Del
7	2p23.2p24.2	18611-29382	Del	38	9q22.33	96225-97131	Amp	69	16p11.2q12.1	29955-50758	Amp
8	2p11.2q11.2	84963-98651	Del	39	9q22.33q31.1	97948-98443	Amp	70	16q21	60574-62316	Amp
9	3p21.31p22.1	42270-45301	Del	40	9q31.3q32	108423-110120	Amp	71	16q22.1q222.2	68772-71706	Amp
10	3p14.3p21.1	53680-55786	Del	41	10pterp15.3	0-2197	Del	72	16q23.1q23.3	78534-82157	Amp
11	3p13p14.2	62456-71461	Del	42	10p11.23p12.1	24840-30577	Del	73	17p13.1p13.3	624-9012	Amp
12	3q13.13q22.3	112584-138742	Del	43	10q11.21	43423-44121	Del	74	17p12q11.2	14414-30082	Amp
13	3q26.2	170127-170803	Amp	44	10q22.1	71740-72570	Del	75	17q12q21.31	33343-43736	Amp
14	4p15.33p16.3	4485-11085	Del	45	10q22.3	78276-81837	Del	76	17q21.32q22	47275-51088	Amp
15	4p15.1	28104-35527	Amp	46	10q23.3q24.2	96648-99789	Del	77	17q23.2q25.3	57550-78502	Amp
16	4q21.21q21.23	82882-85428	Amp	47	10q25.2q25.3	112985-114531	Del	78	18pterp11.32	0-714	Amp
17	4q27q28.1	123197-125135	Amp	48	10q26.2qter	127765-136000	Del	79	18p11.31p11.32	2655-3431	Del
18	4q31.21q31.22	144968-148000	Amp	49	11pterp15.5	0-524	Amp	80	18p11.22q11.2	9821-19622	Del
19	4q32.2qter	163266-192000	Amp	50	11p15.1	19235-20188	Del	81	18q12.1q12.2	30825-33435	Del
20	5p15.2p15.33	3055-11435	Del	51	11q23.2q24.3	112818-127680	Del	82	18q21.1	42181-46857	Del
21	5p13.3p15.1	16623-33747	Del	52	12q13.11q13.12	47334-47663	Amp	83	19p13.12q12	14304-33315	Del
22	5q11.2	53345-57985	Del	53	12q15q21.2	66706-75752	Amp	84	20q13.2q13.31	53054-55630	Amp
23	5q35.1q35.2	168407-173496	Del	54	13q12.3q13.1	29031-30769	Amp	85	21q21.1	16652-18596	Amp
24	6p24.3p25.1	4285-7900	Del	55	13q14.2q14.3	46357-51033	Amp	86	21q21.2q21.3	24047-26442	Amp
25	6p22.3p24.1	13498-17056	Del	56	13q33.3q34	107748-111287	Del	87	21q21.3	26789-29646	Amp
26	7p21.1p22.3	2141-16990	Amp	57	14q11.2	22529-22596	Del	88	21q22.13q22.2	35114-40277	Del
27	7p12.1p12.2	48000-52046	Amp	58	14q23.1q23.2	56641-61602	Amp	89	Xpterq22.1	0-99516	Amp
28	7q11.22q11.23	71275-72883	Amp	59	14q23.3	65448-65844	Amp	90	Xq22.2q23	101504-113066	Amp
29	7q21.11q21.3	77250-92475	Amp	60	14q24.3q32.11	73302-87879	Amp	91	Xq25q26.3	125353-134636	Amp
30	7q22.1q31.33	99973-125500	Amp	61	14q32.2q32.31	98738-99306	Amp	92	Xq27.3q28	145658-154000	Amp
31	7q33-7q36.1	132809-147908	Amp	62	15q11.2q12	22771-23232	Amp				

Table 6. Similar regions found in both tumor and adjacent normal samples

3.3.3 Clustering

The next attempt at classifying the data was involved determining if the genomic profiles of the samples clustered according to any particular pattern. We used Cluster 3.0 and TreeView, both written by Michael Eisen from the Howard Hughes Medical Institute at the University of California at Berkeley (http://rana.lbl.gov/EisenSoftware.htm).

Using average linkage unsupervised hierarchical clustering and TreeView, we generated the cluster diagram seen in Figure 34. The long image on the left is the entire group of clones going down vertically with the 37 samples going across horizontally. The image on the right is a magnified section taken from the main image with the tree at the top also magnified.

On the horizontal axis, C represents spleen versus spleen controls. T represents carcinoma and N represents adjacent histologically normal mucosa. I, D and M represent the Lauren classifications of intestinal, diffuse and mixed pathologies. For example, TI represents an intestinal-type tumor.

The dendrogram (at the top of Figure 34) and the order of the columns after unsupervised hierarchical clustering demonstrate that the controls are fundamentally different from all the other samples. The second conclusion that can be gleaned from the cluster diagram is that there is no evidence of any segregation between cancers and adjacent normals regardless of Lauren type. Neither is there any evidence of grouping according to Lauren type although it should be recognized that the small numbers of diffuse and mixed pathologies precludes any meaningful conclusions regarding this point.

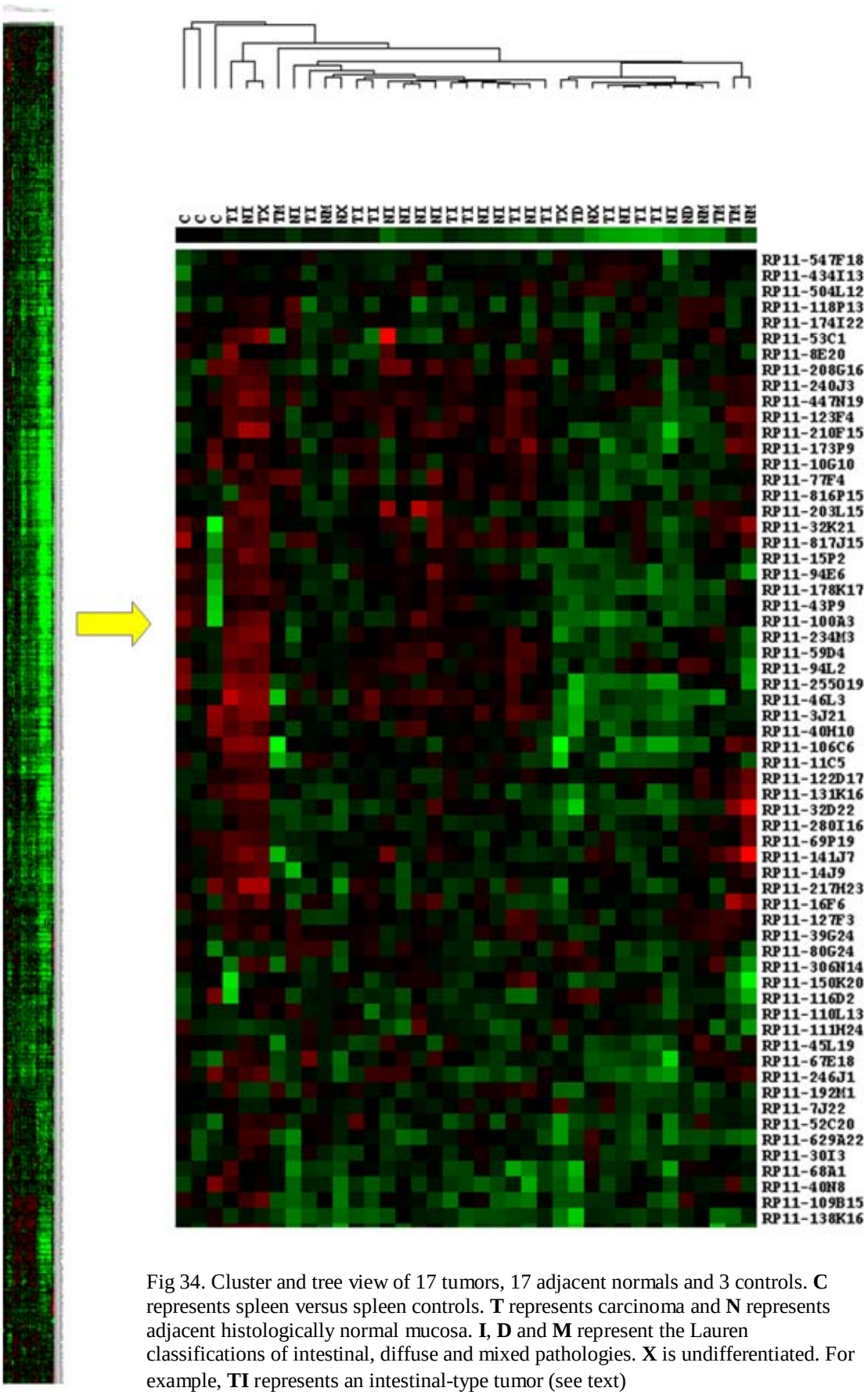


Fig 34. Cluster and tree view of 17 tumors, 17 adjacent normals and 3 controls. **C** represents spleen versus spleen controls. **T** represents carcinoma and **N** represents adjacent histologically normal mucosa. **I**, **D** and **M** represent the Lauren classifications of intestinal, diffuse and mixed pathologies. **X** is undifferentiated. For example, **TI** represents an intestinal-type tumor (see text)

3.3.4 Regions of interest different between normal and tumors

Despite the fact that the bar charts and the cluster diagram so eloquently illustrate the genomic similarity of tumor and adjacent normal tissues, it is nevertheless tempting to probe into possible differences between these two tissue types in an effort to discover possible regions of interest which may contain sequences that propel genetically altered adjacent mucosa down the road to overt carcinoma.

Using the data used to construct the bar charts in section 3.3.2, we searched for BAC clones for which the difference in the frequency of a copy number aberration was greater than 5%. The cytoband was then matched against the UCSC (University of California at Santa Cruz) Genome database to identify the RefSeq genes present in these regions.

No.	Clones	Cytoband	Position (KB)	Frequencies (%)		Change	Possible Genes
				Normal	Tumor		
1	RP11-138K16	1p21.2	99419-99595	41	53	Del	PALMD
2	RP11-94D19	3p14.2	60739-60909	47	53	Del	FHIT
3	RP11-19I19	5q11.2q12.1	58769-58928	41	59	Del	PDE4D
4	RP11-47N20	5q14.1q14.2	81388-81542	47	59	Del	APG10L, ATG10
	RP11-207B2	5q14.2	81757-81757	41	53	Del	
5	RP11-66E14	6p24.3	8799-8799	47	65	Del	
6	RP11-193J17	7q11.22	71146-71146	47	53	Amp	CALN1
7	RP11-9C22	7q32.1	127170-127170	47	65	Amp	SND1
	CTB-162H9	7q32.1	127287-127287	41	65	Amp	SND1
8	RP11-182C2	10q25.1	111415-111495	47	53	Del	ADD3
	RP11-182P7	10q25.2	111544-111544	47	53	Del	ADD3
9	RP11-265I6	18q12.3	35981-36148	47	53	Del	

Table 7. Frequency table of cytobands and genes in corresponding regions

From Table 7, it can be seen that the regions with the greatest differences in copy number frequencies between tumor and adjacent normal involve amplification of 7q32.1 and deletion of 5q11.2q12.1. The respective genes in those regions are SND1

and PDE4D. Little is known about SND1. PDE4D is known to encode functional proteins that degrade the cAMP, which itself is a key signal transduction molecule. Up-regulation of PDE4D may play an important role in epithelial-mesenchymal transition (44).

3.4 Initial conclusions

The obvious conclusion was that tumor and adjacent normal gastric mucosa were genomically alike. Given that the morphological and histological appearances of these two tissue types are vastly different from tumor, this was a difficult conclusion to accept initially.

The consideration of other possibilities to account for the experimental findings included possible tissue contamination, whether at the initial sampling stage or further down the line at the primer stage, or even at the hybridization stage if any reagents had been compromised. We repeated several hybridizations on samples with excess DNA using fresh reagents and clean equipment and our findings were similar.

Another potential source of bias was the FFPE tissue. Given the notorious cross-linking of DNA known to occur in this circumstance, we could not quantify the degree of its effect on our final results since we had no matched fresh specimens for comparison.

The unexpected experimental findings prompted the decision to embark on a series of further experiments to either confirm or refute our findings. To that end the next set of experiments were designed to contemplate the following questions:

1. Are there significant copy number changes in the margin blocks of the 17 patients?
2. Would freshly harvested gastric tissue with minimal formalin fixation processing produce similar results?
3. Are there significant copy number changes in DNA from the FFPE gastric mucosa of non-cancer patients?

Chapter 4

Further Experiments

In order to determine if our results were indeed true or perhaps due to experimental error, we proceeded to perform 3 further sets of experiments: (a) array CGH on the stomach tissue of non-cancer patients; (b) array CGH profile of the margin blocks from the initial set of patients; and, (c) array CGH on freshly harvested gastric cancer specimens. 'Margin blocks' contain paraffin-embedded formalin fixed tissue taken from the proximal and distal resection margins of gastrectomy specimens.

4.1 Stomach tissue from non-cancer patients

This experiment was designed to demonstrate or exclude the possibility that there was tissue contamination at some point along our sampling or hybridization procedures. Samples were obtained from gastrectomies for perforated or bleeding benign peptic ulcers

4.1.1 Methodology

Samples were obtained from patients who had undergone gastrectomy for non-cancer diagnoses. We were only able to obtain specimens from 2 patients who had undergone gastrectomy for large perforated ulcers. The relative scarcity of such gastrectomies today is testimony to the efficacy of proton-pump inhibitors and the sea-change in management of peptic ulcer disease.

Specimens from both patients were processed in a similar fashion, undergoing formalin-fixation and paraffin-embedding. The coring process and the verification of the 40 micron sections were also performed as previously described.

Subsequent DNA extraction, random primer labeling and hybridization were conducted in identical fashion as for the 57 previous specimens. The arrays were imaged and the results are shown below.

4.1.2 Results

Patient	Age	Sex	Race	Surgery
Non-cancer patients				
1	76	F	Chinese	2005
2	49	M	Chinese	2005
Cancer patients				
1	80	M	Chinese	2004
2	78	F	Chinese	2004
3	-	-	-	2004
4	50	M	Chinese	2000
5	81	F	Chinese	2004
6	81	M	Chinese	2002
7	66	M	Chinese	2001
8	65	M	Chinese	2002
9	83	M	Chinese	2004
10	65	M	Chinese	2002
11	80	M	Chinese	2003
12	69	M	Chinese	1999
13	85	M	Chinese	2000
14	63	M	Chinese	2002
15	76	M	Chinese	1999
16	75	F	Chinese	2004
17	-	-	-	2004

Table 8. Comparison of non-cancer (benign ulcer) patients with cancer patients

The epidemiological data of the non-cancer patients with benign ulcers who underwent gastrectomy is summarized in Table 8. No data is available on the NSAID usage and *Helicobacter pylori* status of the cancer patients. Surgery denotes the date the paraffin block was created.

The hybridization image in Figure 35 shows a relatively uniform yellow color across most of the BAC clones suggesting that there are few deletions or amplifications.

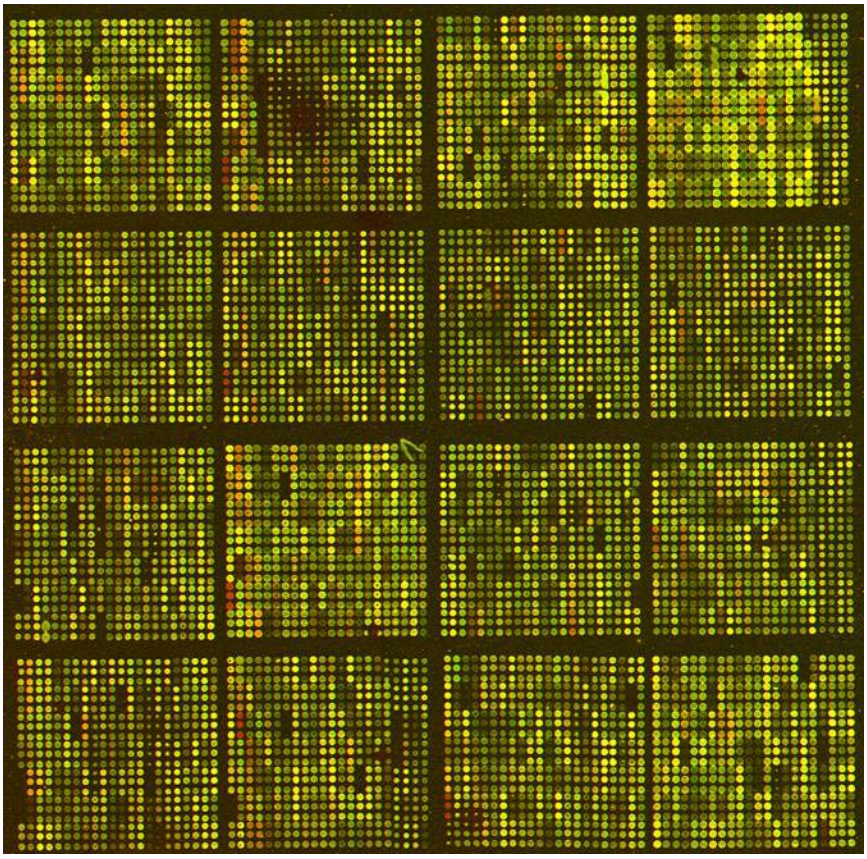


Fig 35. Hybridization image of gastric mucosa from non-cancer patient vs. pooled spleen reference

The genome-wide karyograms for both non-cancer patients in Figure 36 show that despite the yellow appearance of the hybridization image, there are a number of copy number changes present.

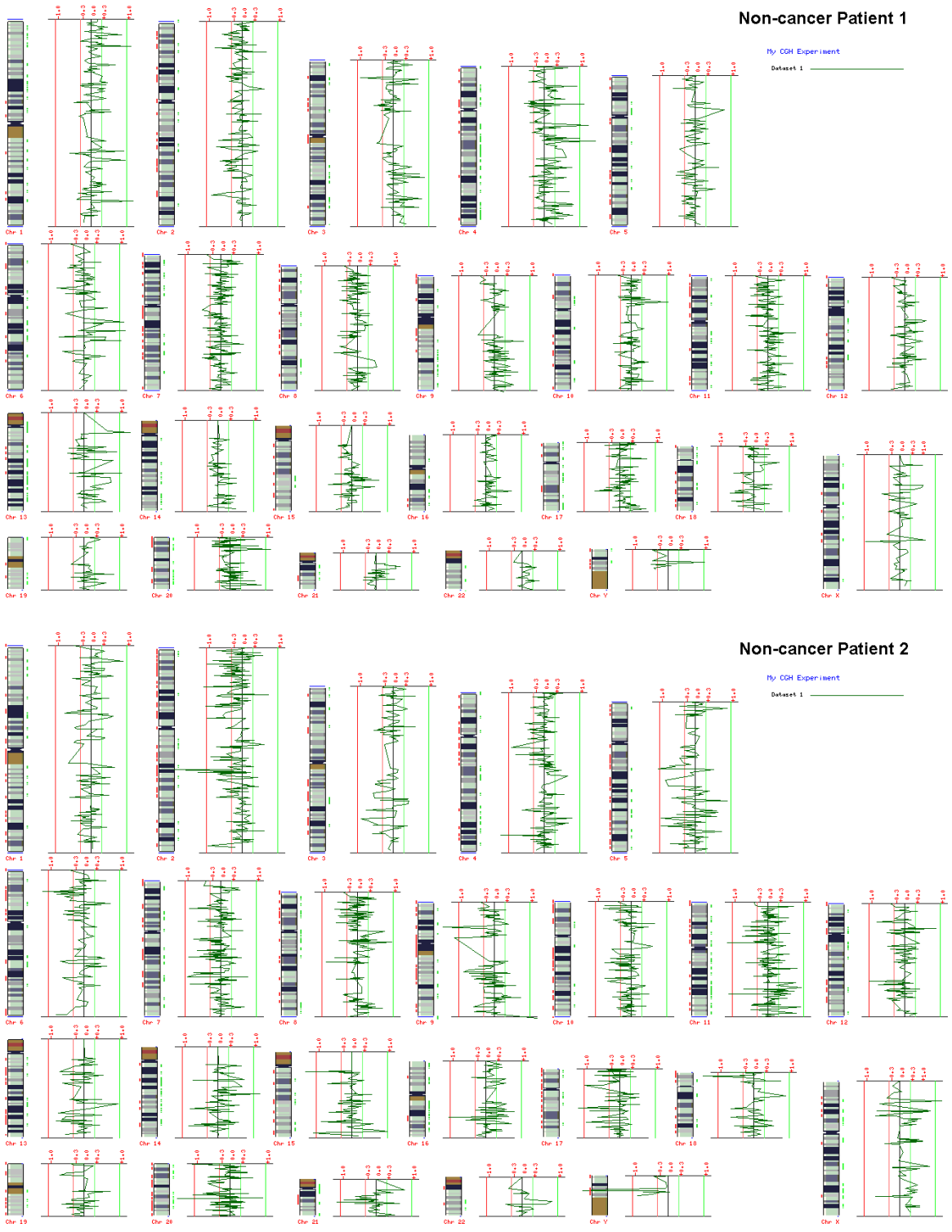


Fig 36. Genome-wide karyograms of both non-cancer patients

However, on closer inspection and comparison with the karyograms seen in Chapter 3, it is fairly evident that both the number and magnitude of the copy number aberrations are greatly reduced in the non-cancer patients. This is perhaps most obvious in magnified views of single chromosomes seen in Figure 37 below.

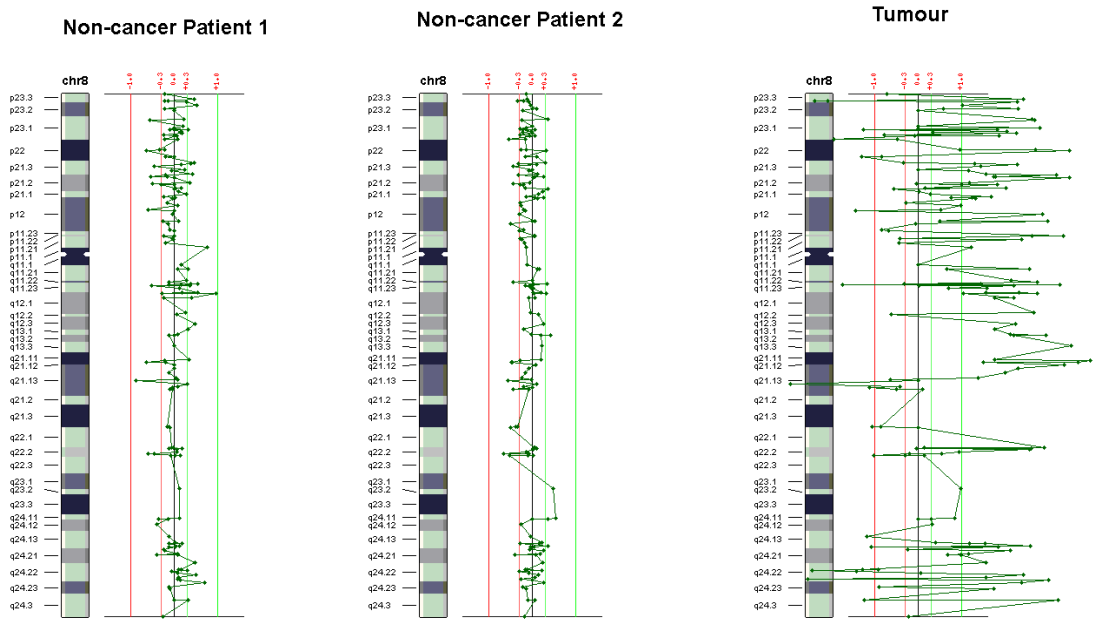


Fig 37. Chromosome 8 profile of both non-cancer patients compared to a tumor specimen

There is also a notable reduction in copy number aberrations in these non-cancer specimens in comparison to the adjacent normal specimens from both the archived tissues as well as from 3 newly-processed prospective samples as seen in Figure 42 (in Section 4.3.2)

The data from these 2 non-cancer patients was then added to the cluster analysis described in section 3.3.3. The resultant cluster and tree diagram is shown in Figure 38. The same abbreviations apply as in section 3.3.3 with the addition of U to represent the 2 non-cancer Ulcer patients.

It is immediately obvious that the 2 non-cancer patients cluster together with the controls away from all the cancer patients.

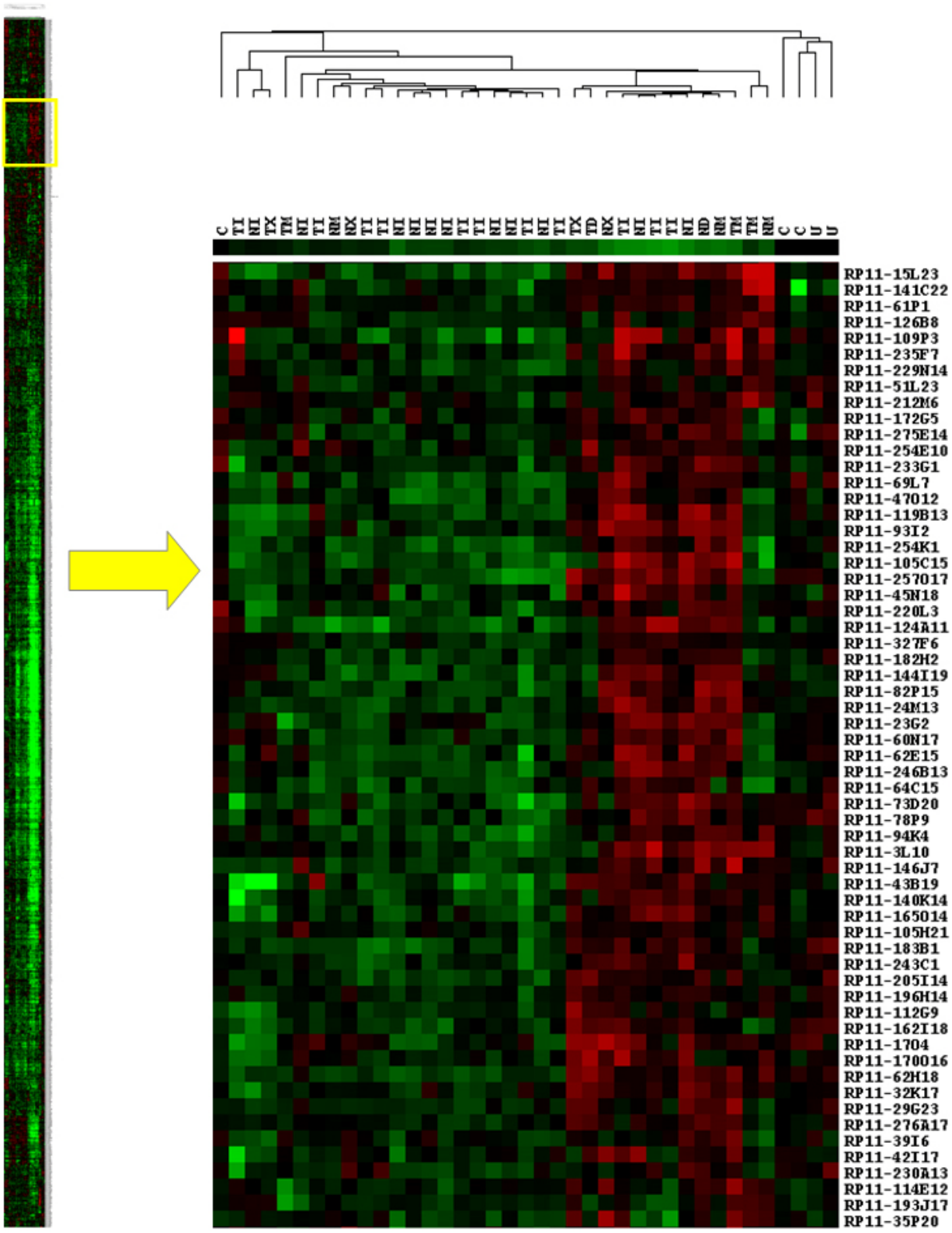


Fig 38. Cluster and tree view of 17 tumors , 17 adjacent normals, 2 non-cancer ulcers and 3 controls. **C** represents spleen versus spleen controls. **U** represents the ulcers. **T** represents carcinoma and **N** represents adjacent histologically normal mucosa. **I**, **D** and **M** represent the Lauren classifications of intestinal, diffuse and mixed pathologies. **X** is undifferentiated. For example, **TI** represents an intestinal-type tumor (see text)

4.1.3 Discussion

The obvious difference in the genomic signature of histologically normal gastric mucosa from cancer versus non-cancer patients is interesting. This difference could perhaps be attributed to one of two possibilities:

1. Histologically normal adjacent mucosa in gastric cancer is genomically abnormal with gross copy number aberrations.
2. The age of the archived tissue versus the recently processed (albeit FFPE) non-cancer tissue might be a deciding factor in the quality of the DNA, leading to differences in the genomic signature.

4.2 Distant resection margins of the original group of 17 patients

As mentioned earlier, thus far all the 'normal' specimens in the initial 57 samples in section 3 can be defined as 'adjacent normals'. This is because they were all harvested from the same block as the tumor specimens.

The purpose of this experiment is to determine if the distant resection margins are similar to the adjacent normals or the tumors.

4.2.1 Methodology

In order to perform this additional experiment, we attempted to trace the margin blocks of the gastrectomy specimens of our original 17 patients. However, we were only able to obtain proximal (gastric) and distal (duodenal) margin blocks for 8 of the 17 patients. Of the remaining 9 patients, several had no margin blocks available and a few had only proximal oesophageal margins after a total gastrectomy.

Table 9 summarizes the characteristics of the margin blocks we were able to obtain. Distances of the histologically uninvolved surgical margin to the histologically involved edge of the primary tumor are also listed to provide an idea of the magnitude in differences of location of these margin specimens from adjacent normal samples.

Patient Number	Sex	Age	Distance from uninvolved surgical margin (in cm)	
			Proximal margin	Distal margin
2	F	77	4.5	4
5	F	80	5	4
6	M	78	3	1.5
7	M	62	5	4
8	M	63	8	2.5
10	M	63	4	5
14	M	60	5.5	2.5
16	F	74	3	5

Table 9. Margin blocks of 8 patients

All the margin blocks were processed as described in Chapter 2. DNA extraction, random primer labeling and hybridization were also performed in an identical manner. Proximal gastric margins were considered 'Far Normals'. Distal duodenal margins do not comprise gastric tissue and are not considered gastric margins for our purposes here.

4.2.2 Results

Since the primary aim of this experiment is to determine whether or not the margins (Far Normals) have a similar genomic profile as the adjacent normals and the tumors, another cluster diagram was constructed. In Figure 39, T signifies tumor, N signifies adjacent normal and F signifies far normal. The numerals after each alphabet denote the patient number.

The appearance of the dendrogram at the top would give the impression that unlike the non-cancer patients, these margins do not segregate on a first-order branch. However, closer inspection will reveal that despite this, there is a real clustering of the Far Normals (proximal margin samples) away from the tumors and the adjacent normals. All the margin samples are on the far right of the cluster diagram indicating that it highly probable that the margins are at some level fundamentally different from the tumor and adjacent normals.

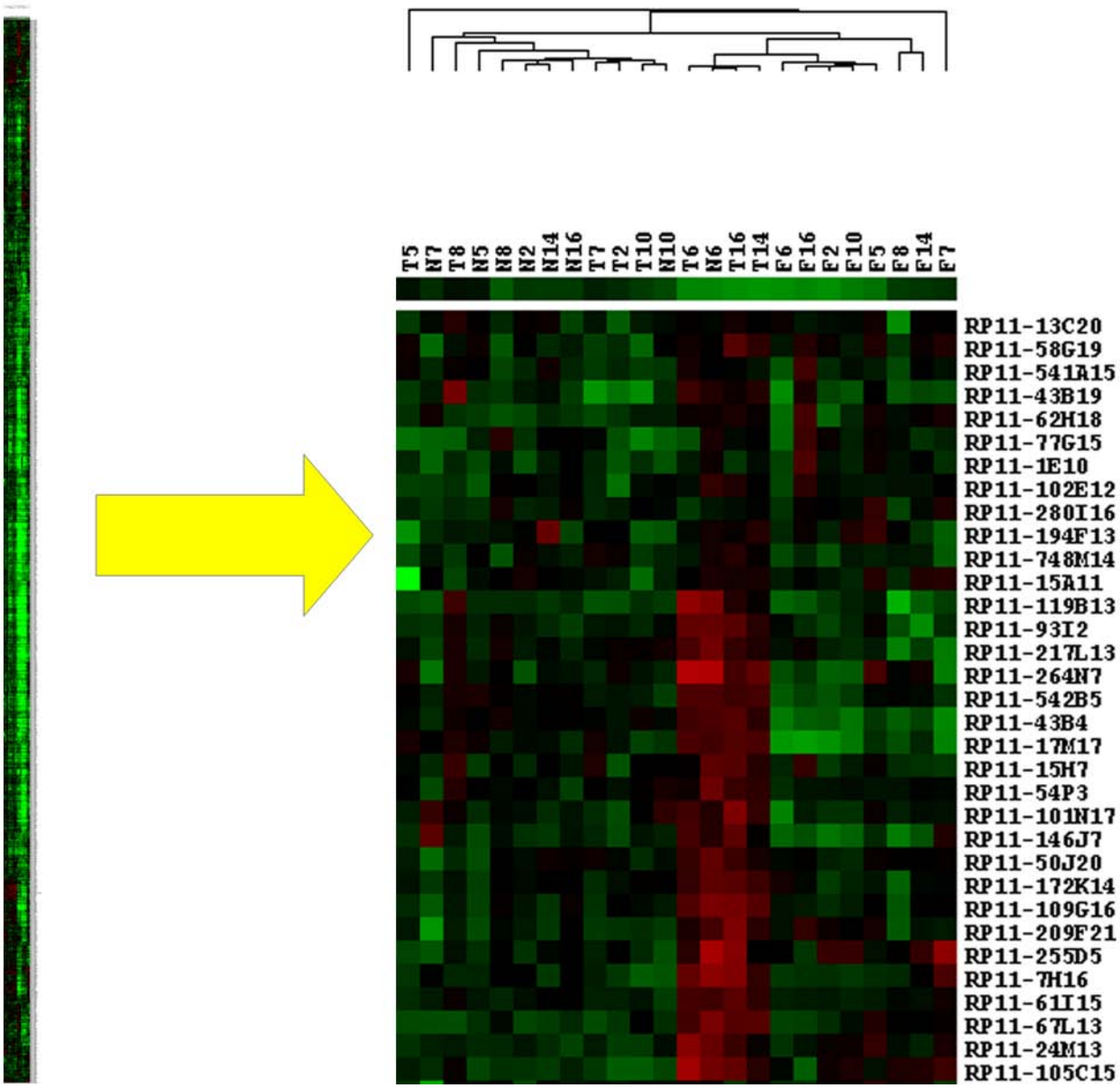


Fig 39. Cluster diagram of 8 tumors (T), 8 adjacent normals (N) and 8 far normals (F)

Taking the comparison one step further, we used the distal duodenal margins to subtract away the ‘noise’ in our array CGH signatures. This was done by excluding any genomic abnormalities that appeared in the uninvolved duodenal samples as well since the duodenal samples do not constitute gastric tissue

The cluster diagram was then reconstructed using this dataset (Figure 40) and the difference between the far normals and the adjacent normals became more pronounced with first order differences emerging in the dendrogram. The far normals again cluster tightly on the far right.

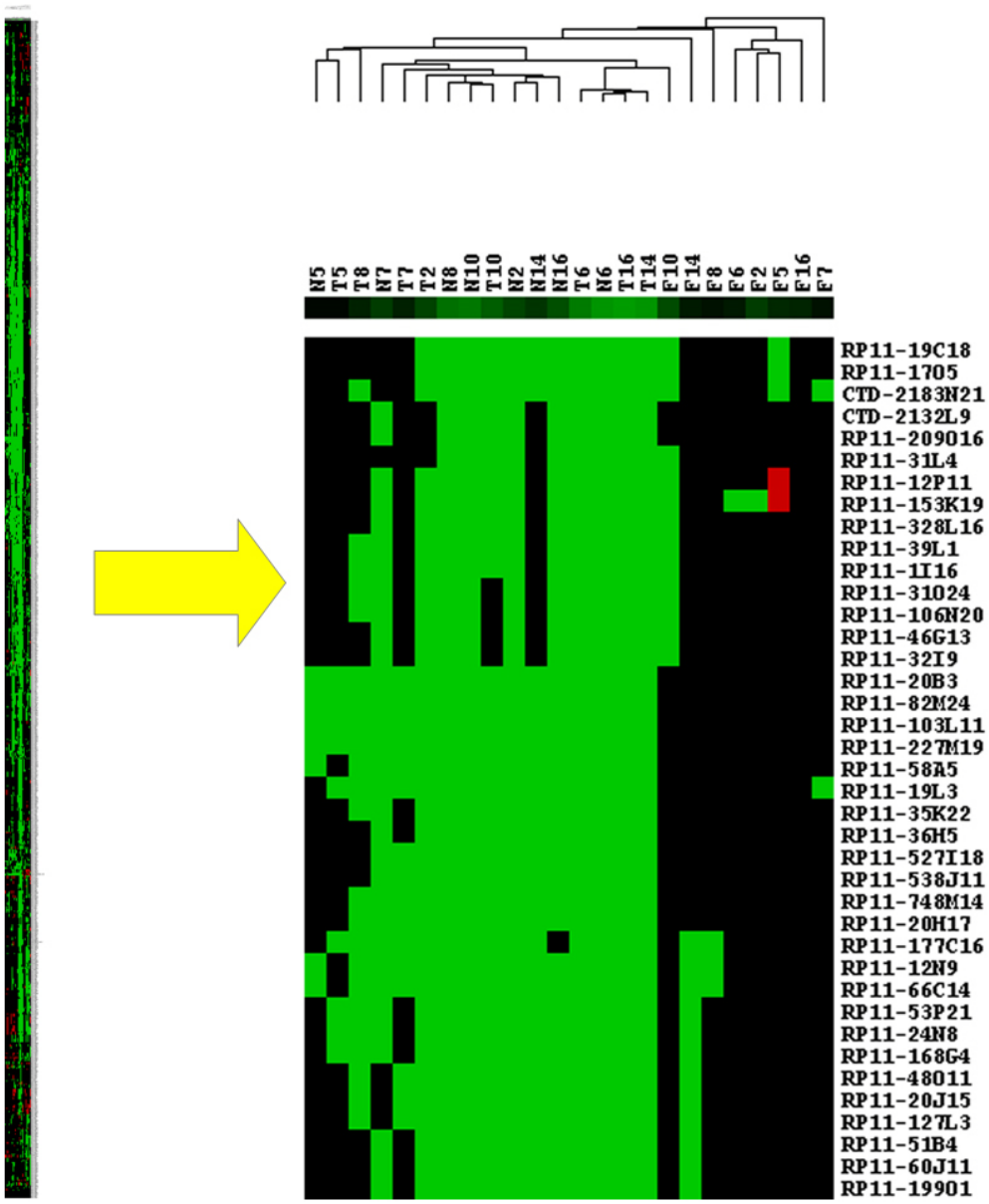


Fig 40. Cluster diagram after subtracting ‘noise’ in duodenal mucosa

Cytoband	Position (KB)	No. of clones	Amp /Del	RefSeq genes
1p13.2p13.3	110394-111807	3	Del	KCNA2, KCNA3, CD53, C1orf103, DRAM2, CEPT1, DENND2D, CHI3L2, LOC149620, CHIA, C1orf88, OVGP1, WDR77, ATP5F1,C1orf162, ADORA3, RAP1A,C1orf183, DDX20, KCND3
1q23.3	159634-159634	1	Del	UHMK1, UAP1
2p23.3	24594-27414	6	Del	**NCOA1**, C2orf79, CENPO, ADCY3, DNAJC27, POMC, EFR3B, DNMT3A, DTNB, ASXL2, KIF3C, RAB10, HADHA, GPR113, SELI, C2orf39, OTOF, C2orf70, KCNK3, CIB4, C2orf18, CENPA, DPYSL5, MAPRE3, TMEM214, AGBL5, LOC100128731, EMILIN1, KHK, CGREF1, ABHD1, PREB, C2orf53, TCF23, SLC5A6, C2orf28
3q26.2	170127-170244	2	Amp	**EVI1**, **MDS1**
4p15.1	30264-30741	3	Amp	PCDH7
4q31.21	144968-147357	6	Amp	GAB1, SMARCA5, LOC441046, GYPE, GYPA, GYPB, HHIP, ANAPC10, ABCE1, OTUD4, SMAD1, MMAA, LOC646603, ZNF827
4q34.3	183090-183250	2	Amp	
4q35.1q35.2	187703-188988	4	Amp	FAM149A, CYP4V2, KLKB1, F11, MTNR1A, FAT1
5p15.2p15.33	3055-11435	11	Del	IRX1, LOC340094, ADAMTS16, KIAA0947, FLJ33360, MED10, FLJ25076, LOC255167, NSUN2, SRD5A1, POLS, ADCY2, C5orf49, FASTKD3, MTRR, SEMA5A, SNORD123, TAS2R1, LOC285692, CCT5, FAM173B, CMBL, MARCH6, ROPN1L, CTNND2, DAP
5q35.1	168407-170151	3	Del	SLIT3, CCDC99, DOCK2, LOC100131897, FOXI1, LOC133874, LCP2, LOC257358, KCNIP1, KCNMB1
6p22.3p23	14557-16761	4	Del	JARID2, DTNBP1, MYLIP, GMPR, ATXN1
7q22.1	99973-100337	4	Amp	ZAN, EPHB4, TRIM56, SERPINE1, SLC12A9, TRIP6, SRRT, UFSP1, ACHE, MUC17
7q33	135996-136916	4	Amp	CHRM2, DGKI, PTN
7q34	141243-141243	1	Amp	LOC93432
8q21.13	76694-79384	2	Amp	LOC100192378, ZFHX4, PXMP3
9q21.32q21.33	80348-83093	5	Amp	RASEF, FRMD3, C9orf103, UBQLN1, GKAP1, KIF27, C9orf64, HNRNPK, RMI1, SLC28A3, NTRK2
11p15.1	19235-19960	6	Del	NAV2, LOC100126784
13q12.3q13.1	29031-30769	3	Amp	ALOX5AP, USPL1, C13orf22, HSPH1, B3GALTL, RXFP2, EEF1DP3, FRY, ZAR1L, **BRCA2**
14q32.2	98738-98738	1	Amp	
15q22.2	58459-58459	1	Del	RORA, NARG2
16q23.1	78534-78534	2	Amp	WWOX
16q23.3	82157-82157	1	Amp	
17p13.3	2494-2572	2	Amp	MNT, METT10D, LOC284009
17p13.1	8460-9012	3	Amp	LOC100128288, KRBA2, RPL26, RNF222, NDEL1, MYH10, CCDC42, SPDYE4, MFSD6L, PIK3R5, PIK3R6
17p11.2p12	14414-19537	6	Amp	HS3ST3B1, PMP22, TEKT3, CDRT4, FAM18B2, CDRT1, TRIM16, ZNF286A, TBC1D26, MEIS3P1, ADORA2B, ZSWIM7, TTC19, NCOR1, PIGL, CENPV, UBB, TRPV2, C17orf45, SNORD49A, SNORD49B, SNORD65, C17orf76, ZNF287, ZNF624, CCDC144A, LOC162632, LOC100129396, TNFRSF13B, MPRIP, PLD6, FLCN, COPS3, NT5M, MED9, RASD1, PEMT, RAI1, SMCR5, SREBF1, TOM1L2, LRRC48, ATPAF2, C17orf39, DRG2, MYO15A, ALKBH5, LLGL1, FLII, SMCR7, TOP3A, SMRC8, SHMT1, EVPLL, LOC339240, LGALS9C, LOC220594, FAM106A, CCDC144B, TBC1D28, ZNF286B, FOXO3B, TRIM16L, FBXW10, FAM18B, PRPSAP2, SLC5A10, FAM83G, GRAP, LOC400581, EPN2, B9D1, MAPK7, MFAP4, RNF112
17q12q21.2	38260-38919	6	Amp	C17orf37, GRB7, IKZF3, ZPBP2, GSDMB, ORMDL3, GSDMA, PSMD3, CSF3, MED24, SNORD124, THRA, NR1D1, MSL1, CASC3, RAPGEFL1, WIPF2, CDC6, **RARA**, GJD3, TOP2A
17q21.31	43736-43736	1	Amp	
17q21.33	48029-48315	2	Amp	NGFR, NXPH3, SPOP, SLC35B1, FAM117A
17q22	50701-51088	3	Amp	CA10
17q23.2	58226-58714	4	Amp	CLTC, PTRH2, TMEM49, TUBD1, RPS6KB1, RNFT1, DHX40P, HEATR6, LOC653653, CA4
18q21.1	43675-46857	4	Del	SMAD2, ZBTB7C, KIAA0427, SMAD7, DYM, C18orf32, RPL17, SNORD58C, LIPG, ACAA2, SCARNA17, MYO5B, CCDC11, MBD1, CXXC1, C18orf24, MAPK4, MRO, ELAC1, SMAD4
20q13.11	42402-42402	1	Del	PTPRT
21q21.1	16652-18596	3	Amp	C21orf34, CXADR, BTG3, C21orf91, NCRNA00157, CHODL, PRSS7

Table 10. Genomic abnormalities present in both adjacent normals and tumor but ABSENT in the Far Normals

The DNA from the margin blocks have a distinctly different genomic profile compared to the adjacent normals and the tumors. This result establishes several points:

1. The age of the archived tissue is unlikely to be a major factor in determining the outcome of our initial experiments. This is clearly shown by the fact that different genomic signatures can be obtained from blocks of an identical age, with distance from tumor being the only differentiating factor.
2. Distance from the primary tumor is a significant determinant of genomic instability in histologically normal gastric mucosa in cancer patients
3. The concept of a zone of 'cancerization' surrounding the primary tumor should be considered.

The detailed analysis of cytobands showing copy number aberrations present in both adjacent normals and tumors but absent in proximal margin tissues are summarized in the Table 10 on the preceding page. Several of the genes have been highlighted.

BRCA2 is a DNA repair gene that is most famously associated with breast cancer. There have been a number of reports that have found an association with gastric cancer as well (45) (46) (47) (48) and the risk of developing gastric cancer for carriers of BRCA2 mutations may be as high as 20 -60% (46).

MDS1 and EVI1-like gene were recently found to be aberrantly expressed in gastric cancer cells (49). It is believed that their action as one of the co-repressors of the TGF-β signaling pathway may be involved in gastric carcinogenesis.

4.3 Prospective gastric cancers formalin-fixed and paraffin-embedded (FFPE)

The initial 57 specimens from 17 patients were all acquired from pathology archives dating back up to 5 years. The aim of this experiment was to determine two things:

1. Is there a progression of changes from distant normal gastric tissue to adjacent normal gastric tissue in cancer patients?
2. Are there differences between archival FFPE tissue and freshly prepared FFPE issue?

4.3.1 Methodology

Three patients were identified prior to gastrectomy for cancer as being suitable candidates for tissue harvest. Their consent for tissue donation was obtained in the usual manner using our institution-standard procedure.

Once the stomach was resected, it was examined by a pathologist in the operating theatre complex. The pathologist then provided us with samples of the tumor itself, adjacent normal mucosa and distant proximal gastric margins.

The 3 tissue specimens from each patient were then processed with formalin-fixation overnight followed by paraffin-embedding the next day by our own laboratory staff.

Punch core biopsy, sectioning in to 40 micron wedges and verification of the top and bottom slices was performed as described earlier. DNA was then extracted and hybridized to our BAC arrays with the pooled spleen DNA as reference.

4.3.2 Results

Patient	Age	Sex	Race	Surgery
Prospetive patients				
1	69	M	Chinese	2005
2	67	M	Chinese	2005
3	76	M	Chinese	2005
Non-cancer patients				
1	76	F	Chinese	2005
2	49	M	Chinese	2005
Cancer patients				
1	80	M	Chinese	2004
2	78	F	Chinese	2004
3	-	-	-	2004
4	50	M	Chinese	2000
5	81	F	Chinese	2004
6	81	M	Chinese	2002
7	66	M	Chinese	2001
8	65	M	Chinese	2002
9	83	M	Chinese	2004
10	65	M	Chinese	2002
11	80	M	Chinese	2003
12	69	M	Chinese	1999
13	85	M	Chinese	2000
14	63	M	Chinese	2002
15	76	M	Chinese	1999
16	75	F	Chinese	2004
17	-	-	-	2004

Table 11. Epidemiological characteristics of the 3 prospective cancer patients

The epidemiological characteristics of the 3 prospective cancer patients in comparison to the other patients is summarized in Table 11 above. Surgery denotes the date the paraffin block was created.

The 3 patients were labeled A, B and C respectively. BAC array CGH results from this small group of patients were significantly cleaner than for our initial 17 patients. As an example, the karyogram in Table 41 represents the genomic profile of the proximal gastric margin from one of the patients. It is reasonably similar to the signature from non-cancer patients in section 4.1.2 and distinctly different from the genomic profiles of our initial 17 patients seen in chapter 3.

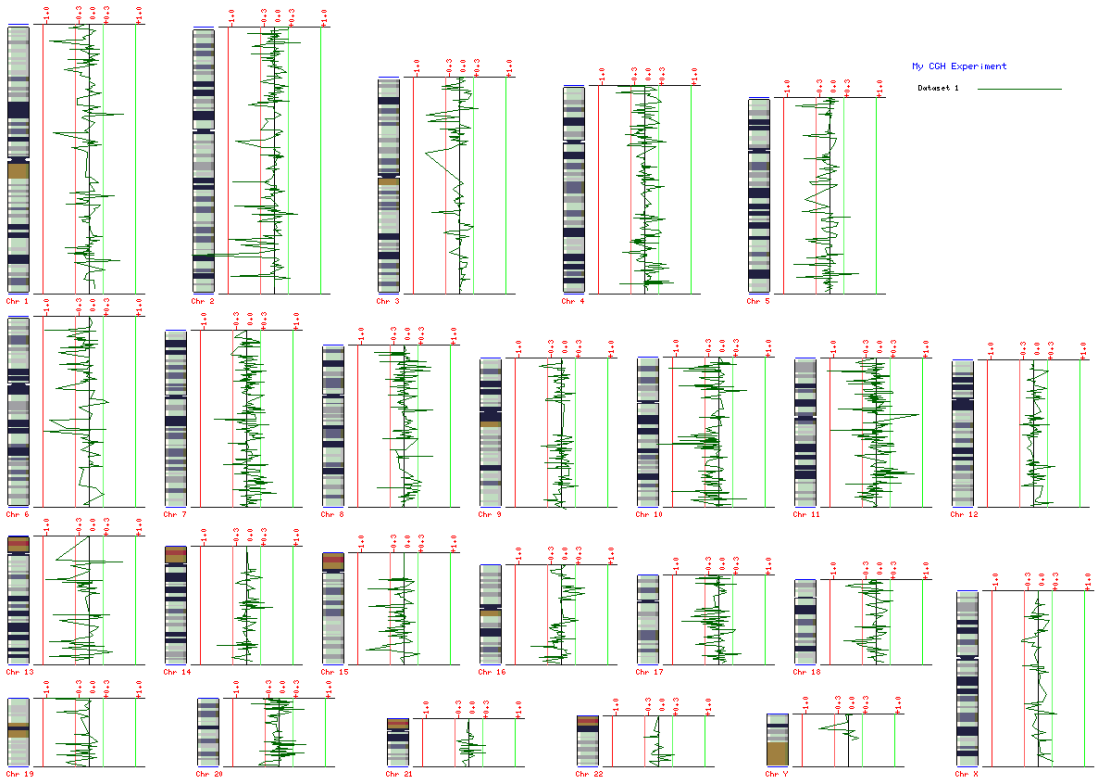


Figure 41. Genome-wide karyogram for the distant normal specimen of Patient A

The comparison is more obvious when we place the magnified single chromosome view of the 3 patients alongside the results from chapter 3. This is demonstrated in Figure 42.

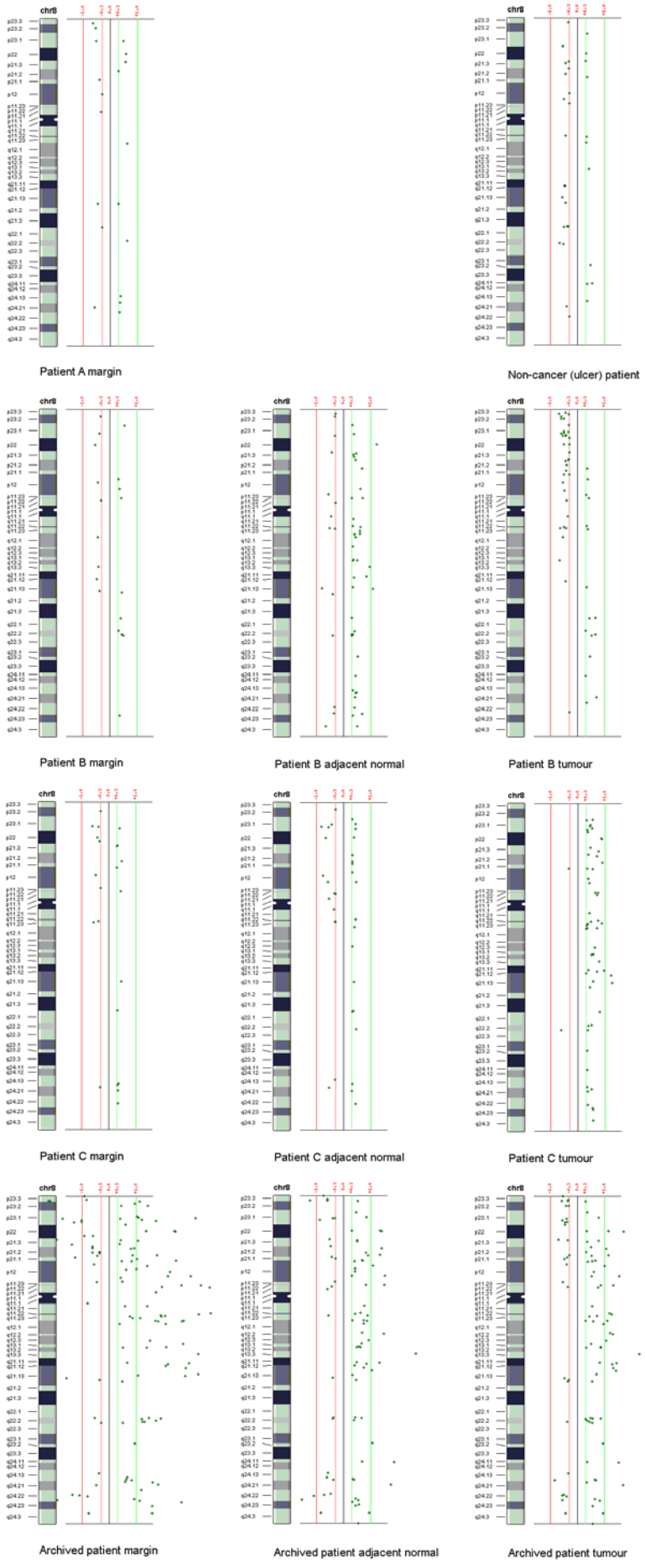


Figure 42. Chromosome 8 profile of different sample types from the 3 prospective patients (A, B, C) compared to similar tissue types of a patient from the initial set of archived specimens. Tumour and adjacent normal samples from Patient A were of insufficient quantity to perform aCGH. A non-cancer (benign ulcer) profile is at the top right corner for comparison purposes

Figure 43 presents in magnified view the progression of changes in the 3 specimen types obtained from this experiment.

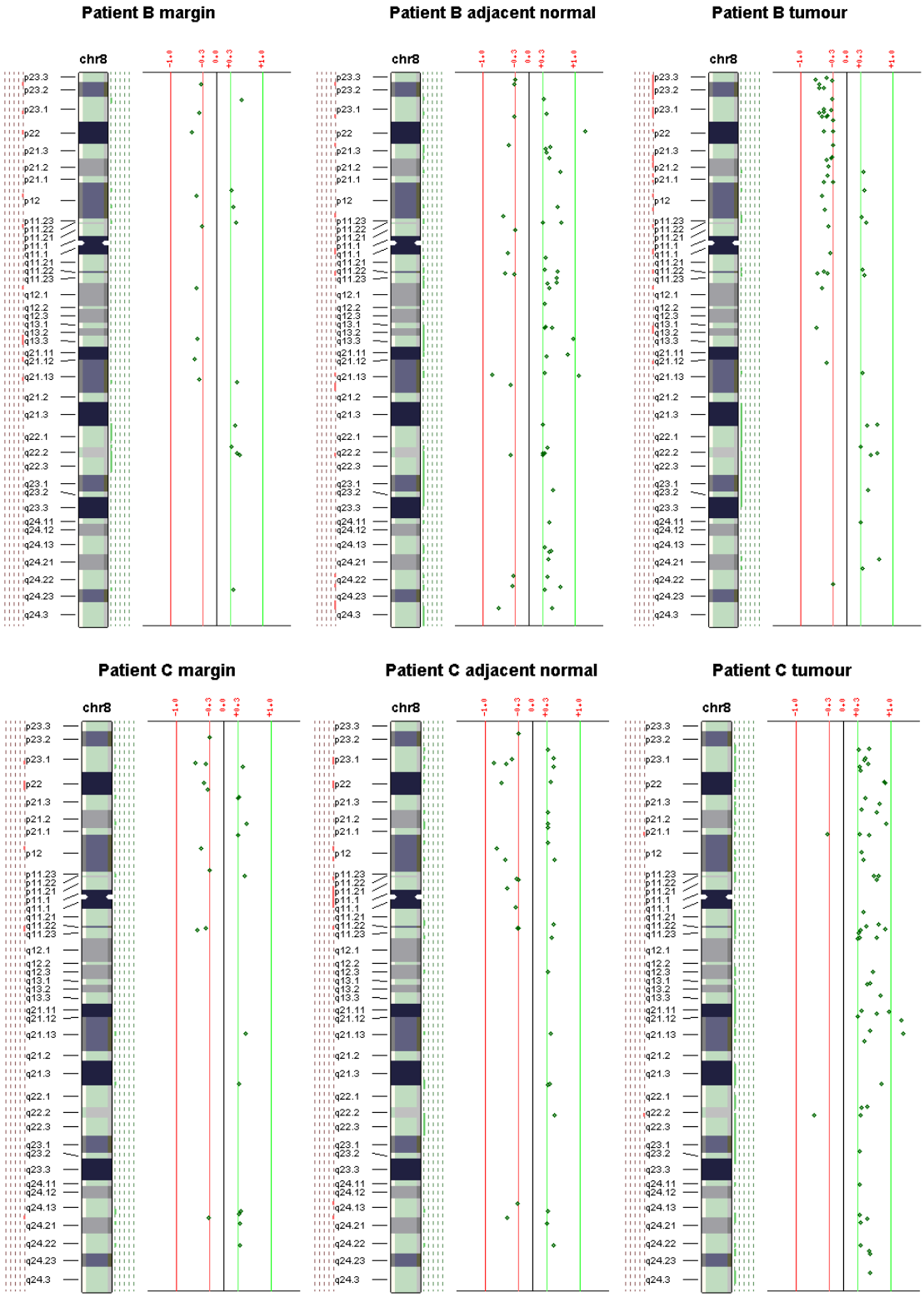


Fig 43. Chromosome 8 comparison across tissue types from Patients B & C

There is a visible progression of genomic abnormality in the 3 specimens obtained from each patient. The tumor specimens show the expected amplifications in chromosome 8 characteristic of most gastric cancer samples. The margin samples are relatively clean with only a few outliers, similar to the profile from non-cancer patients. The adjacent normal tissue is perhaps somewhere in between with a greater number of genomic changes than the margin samples. The adjacent samples are also characterized by greater amplitudes in the copy number changes.

Comparison of the tumor profiles of the 2 patients above with the tumor profiles of chromosome 8 in our earlier 17 patients also demonstrates a 'cleaner' signature despite the characteristic amplifications seen. This reduction in 'noise' would be helpful when trying to determine area of real genomic aberration as opposed to outliers caused by poor preservation or cross-linking of DNA.

4.3.3 Discussion

The results of this experiment demonstrating the genomic profile in freshly harvested FFPE tissue compared to our initial archived FFPE tissue allows us to conclude that:

1. There is a gradual progression of accumulated genomic changes from histologically normal margin specimens to histologically normal adjacent gastric mucosa to overt carcinoma.
2. There is a distinct improvement in 'noise' reduction when comparing the newly processed specimens compared to the archival tissues. This may be related to the time from harvesting to fixation, the duration of formalin fixation or perhaps even the age of the block itself.

Chapter 5
Final Analysis and Discussion

5.1 Summary of initial results

The results of the first 57 archived formalin-fixed paraffin-embedded (FFPE) specimens from 17 patients were characterized by a relatively 'noisy' genomic signature despite the use of smoothing algorithms such as the LOWESS technique.

Nevertheless, once the thresholds for each specimen were defined using the Douglas *et al.* method (43), a pattern was discernible. The expected genomic amplifications and deletions in tumor tissue were seen. There were also a satisfyingly large number of changes in dysplastic and metaplastic tissue that mirrored the aberrations in tumor tissue. What was unexpected however was the quantity and magnitude of changes in adjacent histologically normal gastric mucosa from these cancer patients.

These aberrations in the adjacent normal mucosa were further analyzed by comparing them with the tumor specimens in all 17 patients. There was a marked similarity in the genomic signature of adjacent normal tissue with tumor tissue on visual inspection of the data using our new ACAVIS software. This was further confirmed when the data was represented on bar charts. An unsupervised clustering of the tumor and adjacent normal samples failed to detect any pattern of segregation between the 34 samples (17 tumors and 17 adjacent normals) with the only conclusion being that all 34 were abnormal compared to our control hybridizations.

Looking more closely at the regions of similarity, it was discovered that 92 cytobands which were amplified or deleted in at least 50% of both adjacent normal

and tumor specimens. These were too many to characterize as there was no practical method available to determine which were more significant than others.

The surprising results also raised disturbing questions as to the possibility of cross contamination or bias arising from experimental error. This was despite a fairly rigorous process during which we had established controls for the reference DNA and minimum DNA quantities before hybridization. The controls had been procured from similar FFPE sources and self versus self hybridization of these controls had revealed no discernible error within the hybridization process.

Additional experiments were designed to confirm or refute our initial findings. A decision was also made to focus on adjacent mucosa and tumors rather than intestinal metaplasia or dysplasia since it was assumed that this would serve to accentuate the significance of any findings if the histological types were far removed from each other along the pathway of the Correa hypothesis.

5.2 Summary of results from further experiments

The first additional experiment that was performed was on FFPE gastric mucosa from non-cancer patients. As expected, the non-cancer genomic signature was similar to that of our spleen versus spleen reference control, and completely different from the tumors or adjacent normals. This experiment served to confirm that our bench work processes were not the source of the unexpected initial results.

The second additional experiment examined the margin blocks from the first 17 patients. As only 8 such blocks were available from the pathology archives, our analysis was confined to these alone. Nevertheless, the results were highly significant

showing that the margins do not share many of the genomic abnormalities of the adjacent normals. Although the genomic signature of the margins themselves were relatively ‘noisy’ much like the other 57 original samples, they were clearly less aberrant in terms of significant copy number changes and as such all 8 samples clustered away from their corresponding tumors and adjacent normals (see section 4.2.2). The conclusion served to confirm the suspicion that the adjacent normals themselves, while histologically normal, harbored extensive genomic aberrations.

The final additional experiment involved the collection of fresh cancer specimens which were then processed with formalin in our own laboratory. The results confirmed the expected progression of changes from distant normal mucosa to adjacent normal mucosa to tumor, which was the logical conclusion of the earlier experiments.

The results from the 3 experiments also demonstrated that the ‘noise’ from the older archived pathology blocks was significantly greater than the ‘noise’ seen in the specimens processed in our laboratory. This was manifested by a more widely spread out distribution of outliers than in the more recent specimens.

5.3 Field Cancerization

It is universally recognized that histopathology is the 'gold standard' for diagnosis of cancer. Therefore it was unexpected that so many significant changes were found in non-cancer mucosa in our study. These histologically normal adjacent regions harbored many of the same changes that were also found in their corresponding tumors.

The most likely explanation for our findings is the concept of a field change in the gastric mucosa. This concept was first proposed in 1953 (50) and it explains why the changes are less pronounced or even absent at the distant margins of the gastrectomy specimens. The general pathogenesis of a field defect can be seen in the diagram on the next page. The theory is that chronic exposure to a DNA-damaging agent leads to the clonal expansion of inappropriate cell types that exhibit genetic instability. This premalignant state would eventually lead to transformation into overt carcinoma. When compared to the Correa hypothesis, it is clear that gastric carcinoma falls neatly into this process. The initiator for the field defect would be some sort of injury such as chronic gastritis secondary to *Helicobacter pylori* infection triggering the progressive sequence of gastric atrophy, intestinal metaplasia, dysplasia and finally carcinoma.

Another potential trigger for field cancerization in the stomach may be injury to the stomach mucosa by bile acids and this is the theory that has been advanced to explain the known phenomenon of higher rates of gastric cancer in patients with previous partial gastrectomies for peptic ulcer disease. The recent dramatic rise in proximal gastric or cardio-oesophageal carcinomas is also supported by this theory of

cancerization in which the presence of Barrett's esophagus serves as an intermediate entity in carcinogenesis.

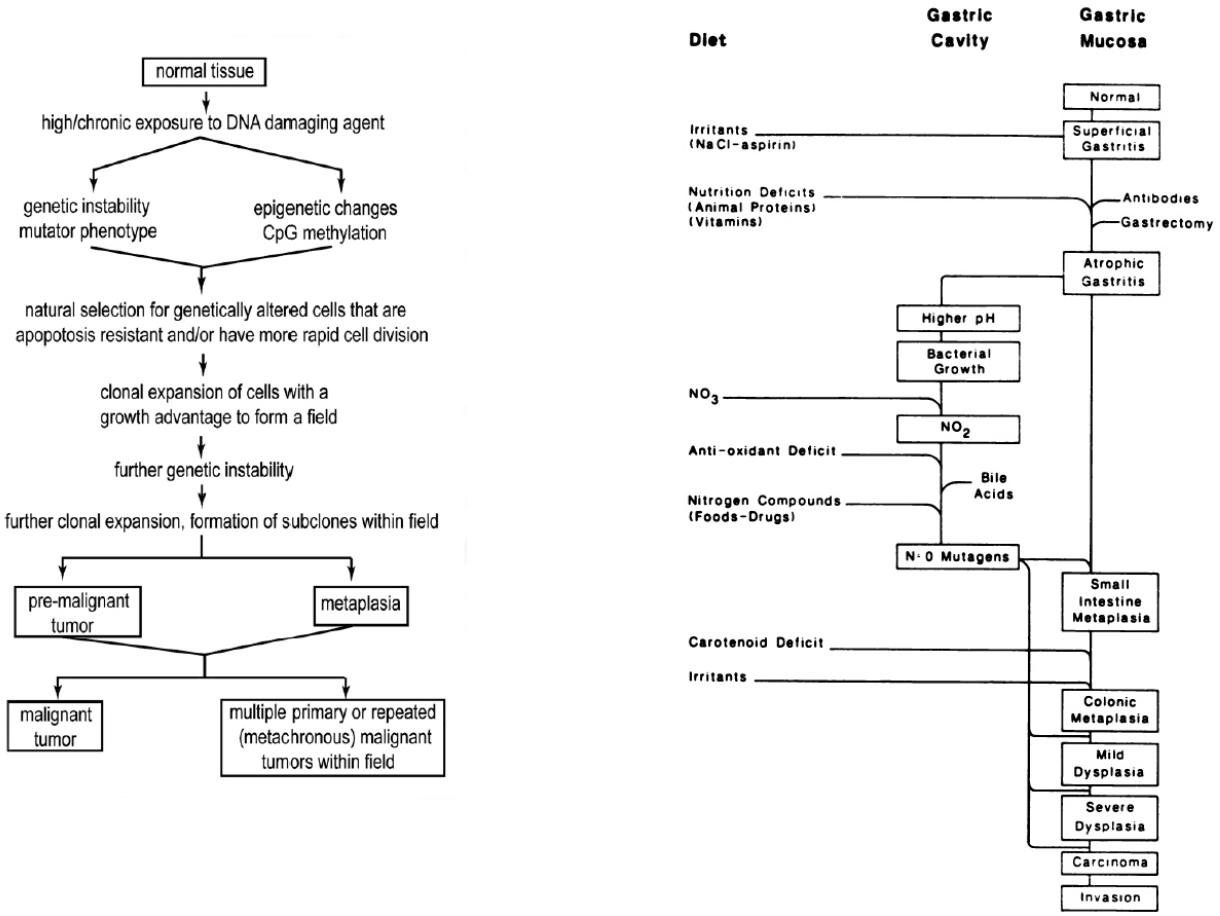


Fig 44. General pathway for the development of a field defect (adapted from Bernstein) (51) on the left and the Correa hypothesis on the right.

The concept of field cancerization and our discovery that histologically normal gastric mucosa harbors many similar changes to carcinoma lends credence to the old surgical maxim that the resection margin should be at least 5 cm away from the tumor. While it was previously believed that this was to allow for the possibility of submucosal microscopic spread of tumor cells, it can now be attributed to the propensity of adjacent mucosa to develop cancer.

The ability to detect these genomic changes may potentially allow a more sensitive method for intraoperative decision-making on the extent of resection. This role is currently occupied by frozen section histopathology. Given the superior sensitivity of genomic analysis, should a rapid test be available one day, it would undoubtedly supplant frozen section not only in gastric cancer but for any malignancy that has an element of field cancerization (e.g. head and neck squamous cell carcinomas).

Other cancers that have had reported genetic or structural changes in the absence of histopathological evidence of malignancy include colon (52) (53), prostate, breast, esophagus (54) and the upper aerodigestive tract (55) (56).

The evidence for colon cancer was first reported in 2004 when it was found that histologically normal adjacent mucosa had altered gene expression in mice and in human cancer patients.

Proteomic analysis of morphologically normal mucosa in patients with colorectal malignancies further confirmed that there were field-wide changes in protein expression (57).

Further evidence for field cancerization is provided by the recent finding that there are nanoscale cellular changes in histologically normal mucosa in colon cancer, pancreatic cancer and lung cancer (58) (59). It was found that partial wave spectroscopy could quantify statistical properties of nanoscale cell structures (59). The disorder strength of the nanoscale architecture was reduced in both tumor cells as well as microscopically normal cells adjacent to the tumor.

A study of gene expression in prostate cancer and normal-appearing adjacent tissue found that both were fundamentally different from prostatic tissue in cancer-

free organ donors (60). Studies in the breast have also reported genomic instability in histologically normal tissues (61) (62).

Although no reports have yet emerged on genome-wide copy number aberrations in histologically normal stomach mucosa, there have been some reports of genetic changes in adjacent normal gastric epithelium involving the hMSH2 gene (63) and the RUNX3 gene (64).

5.4 Regions of interest

A systematic review of the genomic alterations in gastrointestinal cancers published last year (65) noted that in 45 published reports of CGH, the most frequent alterations found in gastric cancer were +20q13 (38.9%), +8q23 (31.7%), -19p13 (20.9%) and +17q21 (20.5%). All 4 of these aberrations were found in our study population (see section 3.3.2) in both tumor and adjacent normal samples. In the further subset analysis of 8 sets of samples in section 4.2.2, it was noted that +20q13 and +17q21 were present in both adjacent normals and tumors but not in proximal margin samples.

20q13 contains a region encoding for the PTP-RT gene (Protein tyrosine phosphatase, receptor type, T). PTP's are known to be signaling molecules that regulate cellular processes such as cell growth, cell differentiation, mitosis, and oncogenic transformation. PTP expression has previously been correlated to gastric cancer progression (66). 17q21.33 contains genes such as NGFR, NXPH3, SPOP, SLC35B1 and FAM117A. Unlike PTPRT, there are as yet no reports linking the gene products to gastric cancer.

Examples of other cytobands that have been reported to be involved in gastric carcinogenesis include 7p12, 8q22 and 15q22-q25 (67). These were also found in our cohort of patients as can be seen in the tables in chapters 3 and 4.

Although the gene pathways correlating these regions of genomic abnormality may not be well understood yet, the discovery of these regions can have an immediate impact on the way we manage gastric cancer. For example, aberrations on chromosome 8 have been suggested as a diagnostic marker while chromosome 19 abnormalities have been associated with younger patients and gains in chromosome 17 have been linked to rapid tumor progression and poor prognosis (68).

5.5 Issues with FFPE tissue

A recent report suggested that FFPE tissues display abnormally large numbers of spurious copy number changes when used for the purpose of array CGH as compared to fresh tissue (69). This is certainly consistent with our experience. It has been suggested that the presence of necrosis in a tissue specimen has an adverse effect on the quality of array CGH as well (70).

It was unfortunate that the quality of the genomic DNA in the formalin-fixed paraffin-embedded tissue in our hospital archives was suboptimal. The results from the few prospective specimens processed in our laboratory were significantly cleaner. This may have been because of the shorter fixation times since it has been reported that fixation times of less than 20 hours do not impact on array CGH results (71). In retrospect, in addition to looking at the size of the DNA fragments within our initial sample set, it might have been possible to evaluate the DNA quality using more recently described methods such as those techniques involving PCR (72) or isothermal

whole genome amplification (73) prior to performing array CGH. However, if the samples had not passed these qualifying tests, we may have had to use them anyway as there was a paucity of specimens available that satisfied our primary inclusion criteria.

The root of the problem however, appears to lie with the cross-linking action of formalin on nucleic acids (42). Some alternative methods of fixation involving new fixatives such as methacarn, RCL2 (42), HOPE (74) and FineFix (75) have been suggested. However the problem remains that while they may be ideal for a research laboratory setting, most hospital pathology departments continue to use formalin because it is more economical yet maintains consistency with world-wide standards for histopathological diagnosis. The potential requirement for molecular or genomic analysis is unfortunately not part of the cost structure of most clinical institutions.

5.6 Further studies

With the experience from this study, it would be a natural extension to consider a more detailed study of freshly harvested tissue processed in our own laboratory with one of the new fixatives. Laser Capture Microdissection (LCM) if available would be ideal as the sampling method. Using an accurate method of isothermal whole genome amplification described one of our laboratory colleagues (38), we could then proceed to look at the genomic signatures using a newer array such as the 32,000-BAC array, the 500,000-SNP Affymetrix platform or Molecular Inversion Probe (MIP) microarrays.

Despite our stated aim to study intestinal pathway of carcinogenesis, we were only able to acquire 6 complete sets comprising 4 tissue types each. We were also

hindered by the similarities and the 'noise' inherent in our archival specimens. Should a set of freshly harvested tissues be available, this would be ideal to pursue our original intention.

One other group of patients that would be interesting to study would be non-cancer patients. If we could acquire a library of non-cancer gastric tissues, it would be possible to study their genomic profile in comparison with the margins of gastrectomy specimens to determine if there are any subtle differences.

5.7 Conclusion

The study of the human genome is an exploding field exemplified by the surge in research effort and publications in recent years. Gastric carcinoma is one of the major killers in our society and this study confirms that field cancerization is an important concept for this malignancy.

In addition to explaining recurrences and the etiology of gastric cancer, the concept of field cancerization holds the potential for accurate and sensitive genomic diagnosis of 'premalignant' gastric mucosa that may appear histologically normal. It is also likely to be a key area of research in the future as initiators for carcinogenesis are more likely to be apparent in premalignant regions than in areas of full-blown malignancy.

References

1. Brenner H, Rothenbacher D, Arndt V. Epidemiology of stomach cancer. Methods Mol. Biol. 2009;472:467-477.

2. Griffin MS, Raimes S. Oesophagogastric Surgery: A Companion to Specialist Surgical Practice. 3rd ed. Saunders Ltd.; 2006.

3. Roder DM. The epidemiology of gastric cancer. Gastric Cancer. 2002;5 Suppl 1:5-11.

4. Lee HP. Singapore Cancer Registry Interim Report: Trends in Cancer Incidence in Singapore 2002-2006. National Registry of Diseases Office, Singapore; 2008.

5. Yeoh K. How do we improve outcomes for gastric cancer? J Gastroenterol Hepatol. 2007 7;22(7):970-972.

6. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol Microbiol Scand. 1965;64:31-49.

7. Correa P. A human model of gastric carcinogenesis. Cancer Res. 1988 Jul 1;48(13):3554-3560.

8. Rugge M, Correa P, Dixon MF, Hattori T, Leandro G, Lewin K, et al. Gastric dysplasia: the Padova international classification. Am. J. Surg. Pathol. 2000 Feb;24(2):167-176.

9. Schlemper R, Riddell R, Kato Y, Borchard F, Cooper H, Dawsey S, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut. 2000 Aug;47(2):251–255.

10. Lee HS, Chang MS, Yang H, Lee BL, Kim WH. Epstein-barr virus-positive gastric carcinoma has a distinct protein expression profile in comparison with epstein-barr virus-negative carcinoma. Clin. Cancer Res. 2004 Mar 1;10(5):1698-1705.

11. Gutiérrez-González L, Wright NA. Biology of intestinal metaplasia in 2008: more than a simple phenotypic alteration. Dig Liver Dis. 2008 Jul;40(7):510-522.

12. Hamilton JP, Meltzer SJ. A review of the genomics of gastric cancer. Clin. Gastroenterol. Hepatol. 2006 Apr;4(4):416-425.

13. Gumucio DL, Fagoonee S, Qiao XT, Liebert M, Merchant JL, Altruda F, et al. Tissue stem cells and cancer stem cells: potential implications for gastric cancer. Panminerva Med. 2008 Mar;50(1):65-71.

14. Takaishi S, Okumura T, Wang TC. Gastric cancer stem cells. J. Clin. Oncol. 2008 Jun 10;26(17):2876-2882.

15. Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998 Mar 26;392(6674):402-405.

16. Tamura G, Yin J, Wang S, Fleisher AS, Zou T, Abraham JM, et al. E-Cadherin gene promoter hypermethylation in primary human gastric carcinomas. J. Natl. Cancer Inst. 2000 Apr 5;92(7):569-573.

17. Correa P, Haenszel W, Cuello C, Tannenbaum S, Archer M. A model for gastric cancer epidemiology. The Lancet. 1975 Jul 12;306(7924):58-60.

18. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N. Engl. J. Med. 1988 Sep 1;319(9):525-532.

19. Hisamichi S. Screening for gastric cancer. World J Surg. 1989 Feb;13(1):31-37.

20. Kamangar F, Dores GM, Anderson WF. 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. 2006 May 10;24(14):2137-2150.

21. Marrelli D, Stefano AD, Manzoni GD, Morgagni P, Leo AD, Roviello F. Prediction of Recurrence After Radical Surgery for Gastric Cancer: A Scoring System Obtained From a Prospective Multicenter Study. Ann Surg. 2005 Feb;241(2):247–255.

22. Lee S, Lee J, Hwang N, Kim Y, Rhee P, Kim J, et al. The role of follow-up endoscopy after total gastrectomy for gastric cancer. European Journal of Surgical Oncology. 2005 Apr;31(3):265-269.

23. Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, et al. Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene. 2003 Mar 27;22(12):1872-1879.

24. Tay ST, Leong SH, Yu K, Aggarwal A, Tan SY, Lee CH, et al. A Combined Comparative Genomic Hybridization and Expression Microarray Analysis of Gastric Cancer Reveals Novel Molecular Subtypes. Cancer Res. 2003 Jun 15;63(12):3309-3316.

25. Anderson C, Nijagal A, Kim J. Molecular markers for gastric adenocarcinoma: an update. Mol Diagn Ther. 2006;10(6):345-352.

26. Han S, Oh D, Im S, Park SR, Lee K, Song HS, et al. Phase II study and biomarker analysis of cetuximab combined with modified FOLFOX6 in advanced gastric cancer. Br. J. Cancer. 2009 Jan 27;100(2):298-304.

27. Kaminishi M, Takubo K, Mafune K. The Diversity of Gastric Carcinoma: Pathogenesis, Diagnosis and Therapy. 1st ed. Springer; 2005.

28. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992 Oct 30;258(5083):818-821.

29. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Döhner H, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer. 1997 Dec;20(4):399-407.

30. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci U S A. 1992 Sep 15;89(18):8794–8797.

31. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser capture microdissection. Science. 1996 Nov 8;274(5289):998-1001.

32. Paris PL, Albertson DG, Alers JC, Andaya A, Carroll P, Fridlyand J, et al. High-Resolution Analysis of Paraffin-Embedded and Formalin-Fixed Prostate Tumors Using Comparative Genomic Hybridization to Genomic Microarrays. Am J Pathol. 2003 Mar 1;162(3):763-770.

33. Paris PL, Andaya A, Fridlyand J, Jain AN, Weinberg V, Kowbel D, et al. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. Hum. Mol. Genet. 2004 Jul 1;13(13):1303-1313.

34. Lage JM, Leamon JH, Pejovic T, Hamann S, Lacey M, Dillon D, et al. Whole Genome Analysis of Genetic Alterations in Small DNA Samples Using Hyperbranched Strand Displacement Amplification and Array–CGH. Genome Research. 2003 Feb 1;13(2):294-307.

35. Telenius H, Carter NP, Bebb CE, Nordenskjo¨ld M, Ponder BA, Tunnacliffe A. Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. Genomics. 1992 Jul;13(3):718-725.

36. Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmüller G. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proceedings of the National Academy of Sciences of the United States of America. 1999 Apr 13;96(8):4494-4499.

37. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: implications for genetic analysis. Proc. Natl. Acad. Sci. U.S.A. 1992 Jul 1;89(13):5847-5851.

38. Lee CIP, Leong SH, Png AEH, Choo KW, Syn C, Lim DTH, et al. An isothermal primer extension method for whole genome amplification of fresh and degraded DNA: applications in comparative genomic hybridization, genotyping and mutation screening. Nat. Protocols. 2006 Dec;1(5):2185-2194.

39. Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, et al. Assembly of microarrays for genome-wide measurement of DNA copy number.

Nat Genet. 2001 Nov;29(3):263-264.

40. Jain AN, Tokuyasu TA, Snijders AM, Segraves R, Albertson DG, Pinkel D. Fully Automatic Quantification of Microarray Image Data. Genome Res. 2002 Feb;12(2):325–332.

41. Srinivasan M, Sedmak D, Jewell S. Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids. Am J Pathol. 2002 Dec;161(6):1961–1971.

42. Delfour C, Roger P, Bret C, Berthe M, Rochaix P, Kalfa N, et al. RCL2, a New Fixative, Preserves Morphology and Nucleic Acid Integrity in Paraffin-Embedded Breast Carcinoma and Microdissected Breast Tumor Cells. J Mol Diagn. 2006 May;8(2):157–169.

43. Douglas EJ, Fiegler H, Rowan A, Halford S, Bicknell DC, Bodmer W, et al. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. AACR; 2004.

44. Kolosionek E, Savai R, Ghofrani HA, Weissmann N, Guenther A, Grimminger F, et al. Expression and Activity of Phosphodiesterase Isoforms during Epithelial Mesenchymal Transition: The Role of Phosphodiesterase 4. Mol. Biol. Cell. 2009 Nov 15;20(22):4751-4765.

45. Falchetti M, Saieva C, Lupi R, Masala G, Rizzolo P, Zanna I, et al. Gastric cancer with high-level microsatellite instability: target gene mutations, clinicopathologic features, and long-term survival. Hum. Pathol. 2008 Jun;39(6):925-932.

46. Friedenson B. BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. MedGenMed. 2005;7(2):60.

47. Liede A, Karlan BY, Narod SA. Cancer risks for male carriers of germline mutations in BRCA1 or BRCA2: a review of the literature. J. Clin. Oncol. 2004 Feb 15;22(4):735-742.

48. Jakubowska A, Nej K, Huzarski T, Scott RJ, Lubiński J. BRCA2 gene mutations in families with aggregations of breast and stomach cancers. Br. J. Cancer. 2002 Oct 7;87(8):888-891.

49. Takahata M, Inoue Y, Tsuda H, Imoto I, Koinuma D, Hayashi M, et al. SKI and MEL1 cooperate to inhibit transforming growth factor-beta signal in gastric cancer cells. J. Biol. Chem. 2009 Jan 30;284(5):3334-3344.

50. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer. 1953 Sep;6(5):963-968.

51. Bernstein C, Bernstein H, Payne CM, Dvorak K, Garewal H. Field defects in progression to gastrointestinal tract cancers. Cancer Lett. 2008 Feb 18;260(1-2):1-10.

52. Chen L, Hao C, Chiu YSY, Wong P, Melnick JS, Brotman M, et al. Alteration of gene expression in normal-appearing colon mucosa of APC(min) mice and human cancer patients. Cancer Res. 2004 May 15;64(10):3694-3700.

53. Hao C, Moore DH, Chiu YSY, Wong P, Bennington JL, Smith AP, et al. Altered gene expression in normal colonic mucosa of individuals with polyps of the colon. Dis. Colon Rectum. 2005 Dec;48(12):2329-2335.

54. Brabender J, Marjoram P, Lord RV, Metzger R, Salonga D, Vallbohmer D, et al. The Molecular Signature of Normal Squamous Esophageal Epithelium Identifies the Presence of a Field Effect and Can Discriminate between Patients with Barrett's Esophagus and Patients with Barrett's-Associated Adenocarcinoma. Cancer Epidemiol Biomarkers Prev. 2005 Sep 1;14(9):2113-2117.

55. Strong MS, Incze J, Vaughan CW. Field cancerization in the aerodigestive tract--its etiology, manifestation, and significance. J Otolaryngol. 1984 Feb;13(1):1-6.

56. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. Cancer Prev Res (Phila Pa). 2008 Nov;1(6):396-403.

57. Polley AC, Mulholland F, Pin C, Williams EA, Bradburn DM, Mills SJ, et al. Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia. Cancer Research. 2006;66(13):6553.

58. Subramanian H, Roy HK, Pradhan P, Goldberg MJ, Muldoon J, Brand RE, et al. Nanoscale Cellular Changes in Field Carcinogenesis Detected by Partial Wave Spectroscopy. Cancer Res. 2009 Jul 1;69(13):5357-5363.

59. Subramanian H, Pradhan P, Liu Y, Capoglu IR, Li X, Rogers JD, et al. Optical methodology for detecting histologically unapparent nanoscale consequences of genetic alterations in biological cells. Proceedings of the National Academy of Sciences. 2008 Dec 23;105(51):20118-20123.

60. Chandran UR, Dhir R, Ma C, Michalopoulos G, Becich M, Gilbertson J. Differences in gene expression in prostate cancer, normal appearing prostate tissue adjacent to cancer and prostate tissue from cancer free organ donors. BMC Cancer. 2005;5:45.

61. Ellsworth DL, Ellsworth RE, Liebman MN, Hooke JA, Shriver CD. Genomic instability in histologically normal breast tissues: implications for carcinogenesis. Lancet Oncol. 2004 Dec;5(12):753-758.

62. Lakhani SR, Chaggar R, Davies S, Jones C, Collins N, Odel C, et al. Genetic alterations in 'normal' luminal and myoepithelial cells of the breast. J. Pathol. 1999 Dec;189(4):496-503.

63. Sud R, Wells D, Talbot IC, Delhanty JD. Genetic alterations in gastric cancers from British patients. Cancer Genet. Cytogenet. 2001 Apr 15;126(2):111-119.

64. Nakase Y, Sakakura C, Miyagawa K, Kin S, Fukuda K, Yanagisawa A, et al. Frequent loss of RUNX3 gene expression in remnant stomach cancer and adjacent mucosa with special reference to topography. Br. J. Cancer. 2005 Feb 14;92(3):562-569.

65. Nishimura T. Total number of genome alterations in sporadic gastrointestinal cancer inferred from pooled analyses in the literature. Tumour Biol. 2008;29(6):343-350.

66. Wu C, Kao H, Li AF, Chi C, Lin W. Protein tyrosine-phosphatase expression profiling in gastric cancer tissues. Cancer Lett. 2006 Oct 8;242(1):95-103.

67. Buffart TE, Carvalho B, Hopmans E, Brehm V, Kranenbarg EK, Schaaij-Visser TBM, et al. Gastric cancers in young and elderly patients show different genomic profiles. J. Pathol. 2007 Jan;211(1):45-51.

68. Zhang YJ, Fang JY. Molecular staging of gastric cancer. J. Gastroenterol. Hepatol. 2008 Jun;23(6):856-860.

69. Mc Sherry EA, Mc Goldrick A, Kay EW, Hopkins AM, Gallagher WM, Dervan PA. Formalin-fixed paraffin-embedded clinical tissues show spurious copy number changes in array-CGH profiles. Clin. Genet. 2007 Nov;72(5):441-447.

70. Johnson NA, Hamoudi RA, Ichimura K, Liu L, Pearson DM, Collins VP, et al. Application of array CGH on archival formalin-fixed paraffin-embedded tissues including small numbers of microdissected cells. Lab. Invest. 2006 Sep;86(9):968-978.

71. Ghazani AA, Arneson NCR, Warren K, Done SJ. Limited tissue fixation times and whole genomic amplification do not impact array CGH profiles. J. Clin. Pathol. 2006 Mar;59(3):311-315.

72. Jacobs S, Thompson ER, Nannya Y, Yamamoto G, Pillai R, Ogawa S, et al. Genome-wide, high-resolution detection of copy number, loss of heterozygosity, and genotypes from formalin-fixed, paraffin-embedded tumor tissue using microarrays. Cancer Res. 2007 Mar 15;67(6):2544-2551.

73. Buffart TE, Tijssen M, Krugers T, Carvalho B, Smeets SJ, Brakenhoff RH, et al. DNA quality assessment for array CGH by isothermal whole genome amplification. Cell. Oncol. 2007;29(4):351-359.

74. Vollmer E, Galle J, Lang DS, Loeschke S, Schultz H, Goldmann T. The HOPE technique opens up a multitude of new possibilities in pathology. Rom J Morphol Embryol. 2006;47(1):15-19.

75. Stanta G, Mucelli SP, Petrera F, Bonin S, Bussolati G. A novel fixative improves opportunities of nucleic acids and proteomic analysis in human archive's tissues. Diagn. Mol. Pathol. 2006 Jun;15(2):115-123.

Appendices

Appendix 1: Protocol of DNA extraction from FFPE tissue

PUREGENE® DNA Purification Kit

DNA Purification From 0.5-2.0 mg Paraffin-embedded Tissue

Sample De-paraffinization

1. Place 0.5-2.0 mg (0.0005-0.002 g) of finely minced tissue into a 1.5 ml tube. Add 100 µl **Xylene** or **Hemo-De** (Non-toxic alternative / Scientific Safety Solvents, catalog number HD150A) and incubate 5 minutes with constant mixing at room temperature.
2. Centrifuge at 13,000-16,000 x g for 1-3 minutes to pellet the tissue. Discard the xylene or Hemo-De.
3. Repeat steps 1 and 2, twice (for a total of three washes).
4. Add 100 µl of **100% Ethanol** to the tube and incubate 5 minutes with constant mixing at room temperature.
5. Centrifuge at 13,000-16,000 x g for 1-3 minutes to pellet the tissue. Discard the ethanol.
6. Repeat steps 4 and 5 (for a total of two ethanol washes).

Cell Lysis

1. Add 100 µl **Cell Lysis Solution**, and homogenize using 30-50 strokes with a microfuge tube pestle.
2. Incubate lysate at 65°C for 15-60 minutes.
3. If maximum yield is required, 0.5 µl **Proteinase K Solution** (20 mg/ml) may be added to the lysate. Mix by inverting 25 times and incubate at 55°C until tissue particulates have dissolved (3 hours to overnight). If possible, invert tube periodically during the incubation.

RNase Treatment

1. Add 0.5 µl **RNase A Solution** (4 mg/ml) to the cell lysate.
2. Mix the sample by inverting the tube 25 times and incubate at 37°C for 15-60 minutes.

Protein Precipitation

1. Cool sample to room temperature.
2. Add 33 µl **Protein Precipitation Solution** to the RNase A-treated cell lysate.
3. Vortex vigorously at high speed for 20 seconds to mix the **Protein Precipitation Solution** uniformly with the cell lysate. Place sample on ice for 5 minutes.
4. Centrifuge at 13,000-16,000 x g for 3 minutes. The precipitated proteins will form a tight pellet. If the protein pellet is not visible, repeat Step 3 followed by incubation on ice for 5 minutes, then repeat Step 4.

DNA Precipitation

1. Pour the supernatant containing the DNA (leaving behind the precipitated protein pellet) into a clean 1.5 ml centrifuge tube containing 100 µl **100% Isopropanol** (2-propanol). If DNA yield is expected to be low (<1 µg), add 0.5 µl **Gentra Glycogen Solution** (20 mg/ml) to the Isopropanol.
2. Mix the sample by inverting gently 50 times.
3. Centrifuge at 13,000-16,000 x g for 5 minutes.
4. Pour off supernatant and drain tube on clean absorbent paper. Add 100 µl **70%-Ethanol** and invert tube several times to wash the DNA pellet.
5. Centrifuge at 13,000-16,000 x g for 1 minute. Carefully pour off the ethanol. *Pellet may be loose so pour slowly and watch pellet.*
6. Invert and drain the tube on clean absorbent paper and allow to air dry 10-15 minutes.

DNA Hydration

1. Add 20 µl **DNA Hydration Solution** (20 µl will give a concentration of 50 ng/µl if the total yield is 1 µg DNA).
2. Rehydrate DNA by incubating sample 1 hour at 65°C and/or overnight at room temperature. If possible, tap tube periodically to aid in dispersing the DNA.
3. Store DNA at 4°C. For long-term storage, store at -20°C or -80°C.

Printed in USA • 2/03• 00608Rev C

13355 10th Avenue North • Suite 120 • Minneapolis, Minnesota 55441
ph. (763) 543 0678 • USA (800) 866 3039 • fax (763)543 0699 • www.gentra.com

Appendix 2: Protocol of Random Primer Labeling

Random Primer Labeling

	gDNA Digestion	Test Sample 1 = _____ Test Sample 2 = _____			
1		Test 1	Reference 1	Test 2	Reference 2
	Initial DNA conc (ng/μL)				
	DNA vol for ____ ng (μL)				
	Sterile H_20 (μL)	(to 24)	(to 24)	(to 24)	(to 24)
	10x Dpn II buffer (μL)	_____(2.7)	_____(2.7)	_____(2.7)	_____(2.7)
	Dpn II 5 units (μL)	_____(0.55)	_____(0.55)	_____(0.55)	_____(0.55)
	Total	____(27.25)	____(27.25)	____(27.25)	____(27.25)
2	Mix contents and flash spin < 3sec				
3	Incubate 37°C for 5hrs to overnight				
4	Can store digested DNA at -20°C				
	Qiaquick PCR purification kit (Qiagen) Reservoir capacity 800uL, maximum 2.5ug/125uL DNA solution per column				
1	Add 5 vol. PB buffer to samples (for 25 μL digested sample add 125 μL PB)				
2	Vortex; flash spin (3 sec each)				
3	Transfer sample + PB buffer to column				
4	Centrifuge 1min at 16,100 x g				
5	Discard flow-thru and return column to same collection tube				
6	Wash column with 750 μL PE buffer (EtOH added)				
7	Centrifuge 1min at 16,100 x g				
8	Discard flow-thru & return column to same collection tube				
9	Centrifuge 1min at 16,100 x g to remove any remaining wash buffer				
10	Place column onto a new 1.5ml collection tube				
11	Add 44uL EB buffer (0.3μg DNA)/ 88μL EB buffer (0.6μg DNA) to elute DNA				
12	Wait 2 min				
13	Centrifuge 2 min at 16,100 xg				
14	Can store DNA at -20°C				
	Random Primer Labeling				
1	Put into 200 μL tube: gDNA(from above) 44μL 2.5x random primer solution 80μL Total 124μL				
2	Denature DNA by heating mixture at 99°C in a thermocycler for 10min				
3	Snap-cool on ice immediately. Wait 15min to chill. Flash spin ***(SWITCH OFF LIGHTS !)***				
4	Add (on ice): 10x dNTP mixture 20μL Cy3 or Cy5-dUTP 8μL Klenow DNA polymerase 4μL Total volume 156μL		Cy 3 pink soln 1_____ green spots Cy 5 blue soln 1_____ red spots Cy 3 pink soln 2_____ green spots Cy 5 blue soln 2_____ red spots		
5	Mix well and flash spin				
6	Place tube in thermocycler at 37°C for 4hrs to overnight.				
	DNA purification				
1	Place a Microcon YM-30 column into a new tube				
2	Pipette **52 μL** of labeled probes from sample into the centre of the column				
3	Centrifuge at 14,000g for 12min at room temperature				
4	Transfer column into to a new collection tube				
5	Pipette 50uL of sterile water into centre of column → Gently agitate column for 30sec				
6	**Invert** column to collect eluate(use a scissor to cut away caps before spinning)				
7	Centrifuge 2min (14,000g) at room temperature to obtain the purified DNA in 50uL solution				

Appendix 3: Protocol of BAC array hybridization

Array slide hybridization

(Pre-stratalinked at UCSF)

Preparation of samples for hybridization

			Array near End (E)		Array near Label (L)	
			Test 1	Ref 1	Test 2	Ref 2
1	a	DNA concentration (ng/μL)				
	b	DNA vol (μL) [max equal mass test & Ref]	(50)	(50)	(50)	(50)
	c	Human Cot-1 (35 μg) volume (μL)	70.0		70.0	
	d	Ice-cold 100% EtoH **[2.5 vol b+c]**	(425)		(425)	
	e	3M pH 5.2 NaAcetate **[0.1 vol b+c]**	(25)		(25)	
		Total (μL)	(620)		(620)	

2. Place at -20°C for 1 hour
3. Collect the precipitate by centrifugation at 16,100 rpm for 60 minutes at 4°C
4. Carefully aspirate & discard supernatant. Wipe excess liquid from tube with Kim-wipe
5. Air-dry the pellet for approximately 10 minutes
6. Prepare pre-hybridisation solution in a new 1.5mL eppendorf:

Hybridisation mixture(Note 3)	175 μL
20% SDS	50μL
dH2O	25μL
Total	250μL *(keep in thermomixer 37°C)*

7. Dissolve the pellet in 60μL pre-hybridization solution
8. Incubate at room temp for an hour to re-suspend pellet *(keep in thermomixer 37°C)*
9. Pre-warm waterbath to 73°C

Hybridization

1. Denature the DNA sample solution in a water bath at 73°C for 12 minutes
2. Incubate at 37°C in thermomixer for 60 minutes (Cot1 anneals to repetitive sequences)
3. Mark out array boundary under phase-contrast microscopy
4. Apply Easi-Seal & remove all plastic
5. Place the array on slide warmer at 37°C for 10 minutes
6. Add ~ 200 μL Washing solution to bottom of slide box. Pre-warm in 37°C incubator
7. Apply 50 μL of the warmed pre-hybridization solution to each array inside the 'frame'
8. Tilt to wet the whole slide
9. Aspirate ~40 μL of pre-hybridization mixture from the array
10. Apply the DNA sample onto array and spread evenly ***(SWITCH OFF LIGHTS !)***
11. Put in slide box. Watch out sticky Easi-Seal
12. Close carefully keeping it horizontal. Seal with parafilm
13. Incubate 37°C for 48-68 hours on slow rocker
14. Turn slide box 90° every 24 hours

Slide no. _____
Array near end **E** _____
Array near label **L** _____

Post-hybridization wash ***(SWITCH OFF LIGHTS !)***

1. Pre-warm washing solution in a 50°C water bath for 20min
2. Wash the slides once in Washing solution for 15 minutes at 50°C. Keep Easi-Seal on
3. Wash 15 minute in PN buffer at room temperature
4. Remove Easi-Seal while keeping array moist with PN buffer
5. Rinse in 2x SSC briefly
6. Rinse in 70% ethanol, then 85%, then 100% each for 2 minutes
7. Spin dry at 800rpm (Eppendorf Centrifuge) for 2min

→ **Imaging**