Swedish University of Agricultural Sciences
The Faculty of Natural Resources and Agricultural Sciences
Uppsala BioCenter
Department of Microbiology

Copy number variations as potential biomarkers for disease development in breast cancer

Kancherla Reddy Prakash

Master´s thesis • 30 hec • Second cycle, A2E
Biotechnology • ISRN: SLU-MIKRO-EX - 11/13-SE: 2011:13 • ISSN 1101-8151
Uppsala 2011

Copy number variations as potential biomarkers for disease development in breast cancer

Kancherla Reddy Prakash

Supervisor: Tomas Linder, Swedish University of Agricultural Sciences, Department of Microbiology

Assistant Supervisors: Jan P. Dumanski, Uppsala University, Department of Immunology, Genetics and Pathology, Rudbeck Laboratory

Hamid Reza Razzaghian, Uppsala University, Department of Immunology, Genetics and Pathology, Rudbeck Laboratory

Examiner: Volkmar Passoth, Swedish University of Agricultural Sciences, Department of Microbiology

Credits: 30 hec
Level: Second cycle, A2E
Course title: Independent project in Biology - Master's thesis
Course code: EX0565
Programme/education: Biotechnology

Place of publication: Uppsala
Year of publication: 2011
Title of series: no: ISRN: SLU-MIKRO-EX - 11/13-SE: 2011:13
ISSN: 1101-8151
Online publication: http://stud.epsilon.slu.se

Key Words: Breast cancer, Copy number variation, IFNAR1, VNTR

SLU

Swedish University of Agricultural Sciences
The Faculty of Natural Resources and Agricultural Sciences
Uppsala BioCenter
Department of Microbiology

ABSTRACT

Breast cancer is second foremost cause of deaths in women caused by malignant tumor. Comparing genetic profiles of cells surrounding the tumor and blood from same patient may reveal *de-novo* somatic aberrations that may predispose normal cells to cancer cells. Two major types of genetic variations were discovered in human genome: single nucleotide polymorphisms (SNPs) and more recently discovered copy number variations which are chromosomal segments ranging from kilobases to megabases. It is assumed that somatic cells are genetically identical. But it is unknown if CNVs arise in somatic cells. Somatic mosaicism is defined as genetic differences in the cells of a single individual which were developed from one fertilized egg. The purpose of this study was to identify and compare copy number variants in *IFNAR*1 loci of healthy uninvolved margin of breast tissue surrounding primary tumor and blood derived-DNA of breast cancer patients by using Illumina SNPs array-Human660W-Quad.

So far, long and very rigorous research on breast cancer explains only about 10% of all cases. This may signify unusual complexity of cancer and discovery of factors responsible for progression of cancer from healthy cells.

Illumina SNPs genotyping of blood and cells surrounding tumor indicated the occurrence of changes in the region connected to *IFNAR*1 gene loci. Variations in *IFNAR*1 loci formed the base for detailed investigation of expected aberrations. The amplification of DNA fragments and deeper investigations were performed to identify and confirm the differences in genotype between blood and cells surrounding the tumor by cloning, restriction digestion and Sanger's sequencing. The results noticeably confirmed the presence of somatic mosaicism. We have identified aberrations occurring as variable number of tandem repeats (VNTRs) with length of 32 nucleotides.

Consequently, the significant progress in research on cause of cancer carries development of applied technologies (next generation sequencing) that allow examining more number of samples in fewer instances with better precision.

ABBREVIATIONS

BAF	B Allele Frequency
BRCA1	Breast cancer susceptibility protein 1
BRCA2	Breast cancer susceptibility 2
CNV	Copy Number Variation
FISH	Fluorescent In Situ Hybridization
IFNAR	Interferon-α/β receptor
INF	Interferon
JAK	Janus Family Kinases
MMBIR	Microhomology-mediated break-induced replication
PCR	Polymerase Chain Reaction
SNP	Single-Nucleotide Polymorphism
VNTR	Variable Number Tandem Repeat
WGG	Whole Genome Genotyping

CONTENTS **PAGE NO**

INTRODUCTION

Breast cancer is an uncontrolled growth of breast cells forming a malignant tumor. Cancer occurs due to mutations or abnormal changes in the genes which are responsible for growth of the cells. The genes in the nucleus of each cell act as a control room and regulate the growth of the cells [1,2]. Cell division is a process in which new cells develop as the old cells die. When it comes to cell division, cancer cells break all the rules. Over time, mutations can "turn off or turn on'' certain genes in a cell. These cells gain the ability to divide without control or order, producing a high number of cells and forming tumor. Tumors can be benign or malignant. Benign tumors are not cancerous and grow slowly are considered as normal cells and do not spread to other parts of the body. Malignant tumors are cancerous and spread to other parts of the body [3].

Breast cancer is a malignant tumor either beginning in the cells of lobules or ducts. In few cases, breast cancer begins in the stromal tissues which includes fatty and fibrous connective tissues of the breast (Figure 1) [4].

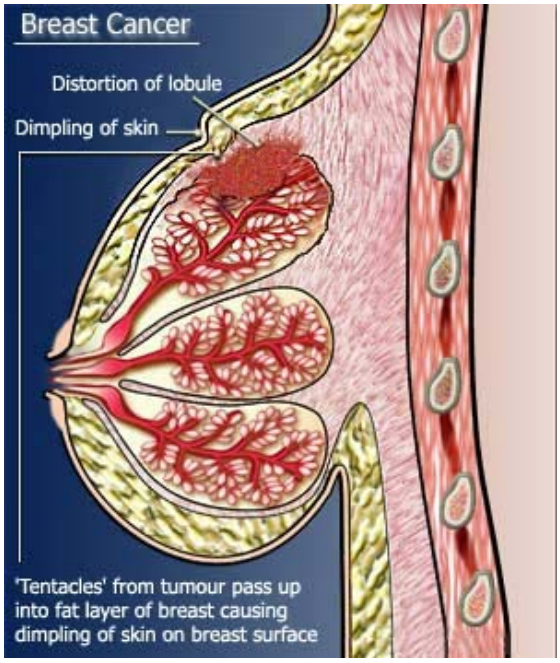

Figure 1. Schematic diagram of breast with cancer (adopted from http://www.healthspablog.org/tag/breast-cancer/ Posted in cancer on 1 May 2009)

Only 5-10% breast cancer is caused due to abnormality inherited from parents and 90% is due to genetic abnormality that occurs due to the aging process. A primary tumor is a tumor growing at the site where tumor progression begins and proceeded to yield a mass of cancerous cells. Breast cancer tissue that has metastasized to the bone is still called as breast cancer. Margins, also known as margins of resection, mean the distance between the tumor and healthy tissue also known as uninvolved margin [5].

One major issue is to find connections of these changes to a certain diseases. The major challenge in human genetics is to identify the DNA sequence differences that distinguish

individuals. These differences were thought to be mostly single nucleotide polymorphism (SNP) changes, structural variation including duplications, deletions and inversions of large blocks of DNA sequence in the human genome [6, 7]. Structural variants have recently been linked to diseases such as asthma and analyzing these variants counts an important step in determining the genetic basis of diseases [8]. Approximately 5-10 % of mutations contributing to breast cancer are caused by the inactivation of autosomal dominant genes: the breast cancer susceptibility genes *BRCA1* and *BRCA2* [9,10]

Copy number variation (**CNV**) is the number of copies of a particular gene or several genes in the genotype of an individual. These segments may range from one kilobase to several megabases in size. Humans being diploid have two copies of each autosomal region in autosomal chromosomes. This number could increase or decrease due to deletion or duplication [11].

CNVs occur due to *de novo* mutation or may be inherited. Fork stalling and template switching are replication missteps that recently have been proposed as mechanism for the cause of some CNVs [12]. However, this model was subsequently superseded by microhomology-mediated break-induced replication (MMBIR) [13].

Some genomic rearrangements such as deletions, duplications, inversions and translocations cause CNVs [14]. Some region-specific repeated sequences like low copy repeats are susceptible to such genomic rearrangements which results in CNVs [15]. CVNs can be identified by using cytogenetic techniques like fluorescent in situ hybridization (FISH), array comparative genomic hybridization and virtual karyotyping with SNP arrays.

After the completion of the human genome project the fact that DNA copy number variation is a widespread and common phenomenon among humans was first uncovered. With respect to copy number approximately 0.4% of genomes of unrelated people typically differ. *De novo* CNVs have been observed in identical twins who otherwise have similar genomes[16].

IFNAR 1

Interferon-α/β receptor (IFNAR) is a cell surface receptor composed of one chain with two subunits IFNAR1 and IFNAR2. IFNAR binds type 1 interferons together with Interferon α and β [17]. Type I interferons (IFNs) are cytokines with multifunctional activity that direct antiviral innate immunity virtually in all nucleated cells [18]. All type I interferons regulate their biological activity through a common receptor which is composed of the transmembrane proteins IFNAR1 and IFNAR2 (Figure 2). After the formation of a ternary complex, the interferon signals are transduced by receptor-associated Janus family kinases (JAK), which activate the STAT proteins. These in turn form homo and heterodimers that translocate to nucleus and directly regulate transcription of interferon responsive genes [19].

Figure 2: Schematic view of the chromosome 21q22.11 class II cytokine receptor cluster: IFNAR2 (interferon-α receptor-2), IL10RB (interleukin-10 receptor B), *IFNAR*1 (interferon-α receptor-1), IFNGR2 (interferon-γ receptor-2). Arrows indicate the direction of transcription and numbers indicate the distances between adjacent genes. Adapted and modified from Hardy et al. 2002

Type I interferons (IFNs) play an important role in defense against viral infections and in modulating the adaptive immune system. In addition, they are capable of inducing a noticed antiproliferative activity. Type I interferon encoding genes, show polymorphisms in many diseases [20]. Protection or susceptibility against hepatitis B and C viruses, cerebral malaria and HIV are influenced by IFNAR polymorphisms [21,22,23]. Polymorphism of IFNAR1 gene can cause impaired INFs signaling which may contribute in tumor progression and may lead to immune dysfunction which is widespread in cancer [24]. Two type I interferons Hu-IFN-αB2 and Hu-IFN-αF have been identified in Chinese hamster ovary cells which interact with splice variants [11, 25].

IFNAR polymorphism is due to Variable Number Tandem Repeat (VNTR). VNTR is a region in the genome where a short nucleotide sequence is repeated tandemly (Figure.3). VNTRs are found in many chromosomes and vary in length between individuals. Analysis of VNTRs is another DNA amplification-based typing method. The number of tandem repeats of the analyzed locus can be enumerated by PCR amplification using primers targeting the flanking regions and determination of resultant fragment length [26]. This method is applied in different organisms from nonculturable bacteria to humans [27].

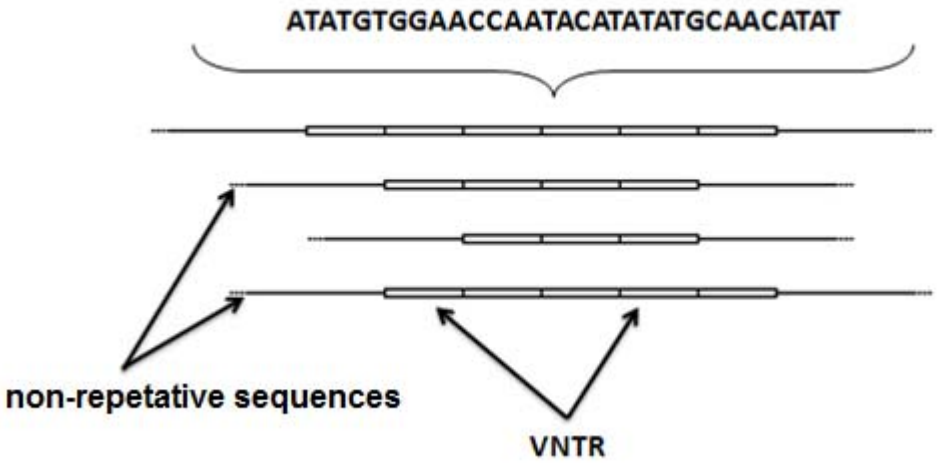

Figure 3: Schematic diagram of VNTR with sequence of 32 nucleotides flanked by non-repetitive sequence. The lines symbolize non-repetitive sequence surrounding VNTR. Repetitive sequences are shown with rectangles. Sequence of each repeated fragment is the same as sequence shown on the top.

By recombination or replication errors individual repeats in VNTR can be removed or added. Non-repetitive sequences are the segments flanking to the repeats which allow the extraction of VNTR blocks by using restriction enzymes and amplification by using Polymerase Chain Reaction (PCR). The size of the each VNTR blocks can be verified by gel electrophoresis. There are two main classes of VNTRs: Microsatellites sequence repeats which are less than 5 base pairs and Minisatellites which are with longer blocks up to 100 base pairs. The analyzed VNTR in the current study approximately consist of 32bp repetitions and located in putative enhancer of *IFNAR*1 gene [28]. The putative enhancer is present approximately 4kbp upstream of the *IFNAR*1 gene.

METHODOLOGICAL BACKGROUND

Illumina SNP Genotyping

Illumina SNP Genotyping arrays like Human660W-Quad® (660K beadchip), are used for detection of variations in DNA copy number (chromosomal aberration) across the genome, characteristically found in cancer [29]. The chromosomal aberrations in DNA copy number can inactivate tumor suppressor genes or activate oncogenes. Based on this, SNP genotyping can indicate deletions, duplications and amplifications which can be associated with cancer. Illumina SNPs array- Human660W-Quad® detects copy number changes by comparison of normalized intensity of signal (R) of a subject sample and a pool of reference samples. The basis for detecting chromosomal aberrations are genomic plots of the $\log_2$ (R subject / R reference) called log R ratio, and so called B allele frequency (BAF) value. Loss or gain of chromosomal material can be identified by measuring the intensities of signal in test DNA versus reference DNA which indicates the DNA copy number.

POLYMERASE CHAIN REACTION

PCR is a rapid and versatile in vitro method for amplifying a single or few copies of DNA to numerous copies of desired DNA fragments. This method includes thermal cycling which consists of repeated heating and cooling of DNA for melting and enzymatic activity of DNA polymerase in reaction mixture. PCR mixture consists of template DNA, $MgCl_2$, PCR buffer, dNTPs (G, C, A, T), a pair of primers (forward and reverse) which enables selective and repeated amplification of template DNA and heat-stable DNA polymerase. dNTPs are molecules when joined together make up structural unit of DNA.

Polymerase is an enzyme that catalyzes polymerization of dNTPs into a DNA strand and primers are short fragments that are complementary to template DNA and are designed so that they anneal at both ends of the DNA fragment that should be amplified during PCR. One primer is complementary to 5’ end and other primer is complementary to 3’ end of template DNA. PCR consists of three steps i) denaturation- in which double strand DNA become a two single strands ii) annealing- each primer anneals to one of the strands iii) elongation- this step depends on DNA polymerase, which synthesizes a new DNA strand complementary to template DNA by adding dNTPs in 5’ to 3’ direction.

CLONING AND TRANSFORMATION

Cloning is a process in which multiple copies of desired DNA fragments can be generated. The DNA fragment is inserted into vector by ligation, which is then linearized by using restriction enzyme. The DNA fragment is then incubated under suitable conditions with DNA ligase. The plasmid contains an antibiotic resistance gene (Figure.4). Following ligation, vector with insert is transformed into bacteria cells (*E.coli*) by transformation. The transfected cells are cultured in a medium containing antibiotics. The plasmid replicates along with

bacteria. The bacterial colonies are selected and the purified plasmid can be used for further studies. Bacteria cells are cultured on medium with antibiotics. This is a fact that only bacteria with antibiotic resistance gene will be able to grow.

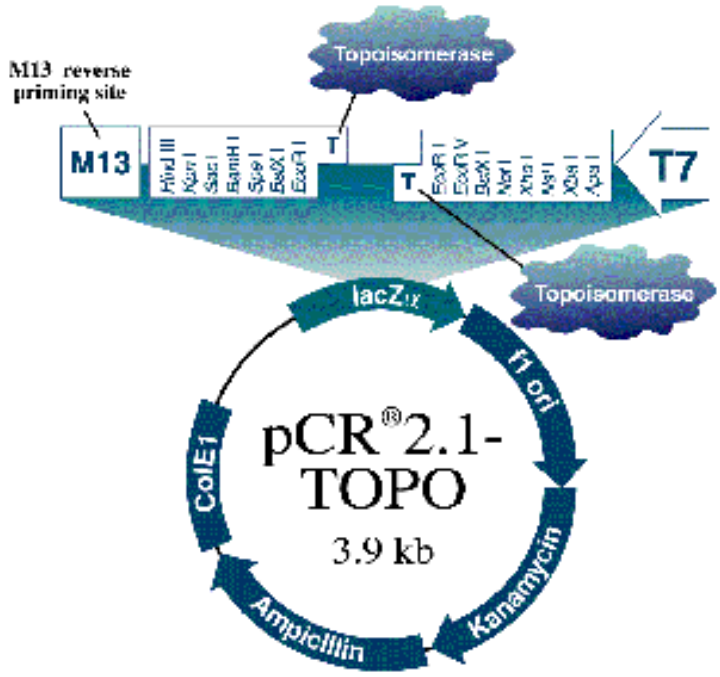

Figure 4: Map of plasmid from Invitrogene TOPO® TA® cloning kit (pCR®2.1-TOPO®) showing multiple cloning sites and with marked place of PCR-product ligation. The linearized vector supplied with in this kit has single overhanging 3' deoxythymidine (T) residues. This allows the insert to ligate with the vector. (Adopted from http://f1000scientist.com/article/display/18015/)

DIGESTION WITH RESTRICTION ENZYME

The process of cutting DNA molecules into smaller pieces at the recognition site with specific enzymes like restriction enzymes or restriction endonucleases is called restriction digestion. These enzymes are able to cleave DNA molecules at specific positions (Figure.5).

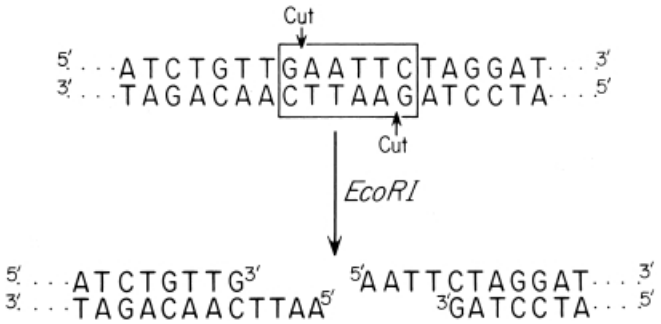

Figure 5: Specific restriction site of *Eco*RI enzyme restriction enzyme. The nucleotide sequence in the DNA contains recognition site (GAATTC) for *Eco*RI. *Eco*RI cuts DNA in both strands between the indicated nucleotides resulting in 5′ end overhangs (adopted from BC Decker Inc, 2000).

Restriction enzymes are classified into three general groups Types I, II, III and IV based on requirements of enzyme cofactor, composition, and nature of their target sequence, position of their DNA cleavage site relative to target sequence [30]. Restriction enzymes first were isolated from bacteria and now there are several types of synthesized restriction endonuleases which cut DNA differently [31]. For example *Eco*RI enzyme recognizes and cuts sequence GAATTC which is a palindrome as a complementary sequence CTTAAG. It creates sticky ends that have T (Thymine) overhangs at their 5' ends and bind to the DNA products that have A (Adenine) overhangs at their 3' ends (Figure.5). Sites where enzyme cuts are called restriction sites and they are typically four, six, eight, ten or twelve nucleotides long. Plasmids contain dozens of restriction sites. This allows inserting of any DNA fragment into plasmid vector which can be later replicated with bacteria cells. After restriction digestion length of DNA fragment can be analyzed by gel electrophoresis.

SANGER SEQUENCING

The sequencing determines order of nucleotides in DNA or RNA molecule. In 1974, two methods "Chemical cleavage protocol" by Maxam and "dideoxy sequencing or chain termination" by Sanger were developed [32]. Sanger's method became standard method because of its practical similarity to natural DNA replication process [33]. The dideoxy sequencing or chain termination is based upon use of ddNTPs in addition to normal dNTPs present in the DNA. The ddNTPs are similar to normal nucleotides but differ in the 3' carbon and they contain a hydrogen group instead of hydroxyl group (Figure 6). During synthesis of DNA the dideoxynucleotide is added to new DNA strand, next nucleotide cannot be added because ddNTPs are lacking 3' hydroxyl group on their deoxyribose sugar. Due to this reason ddNTPs are called as chain terminators.

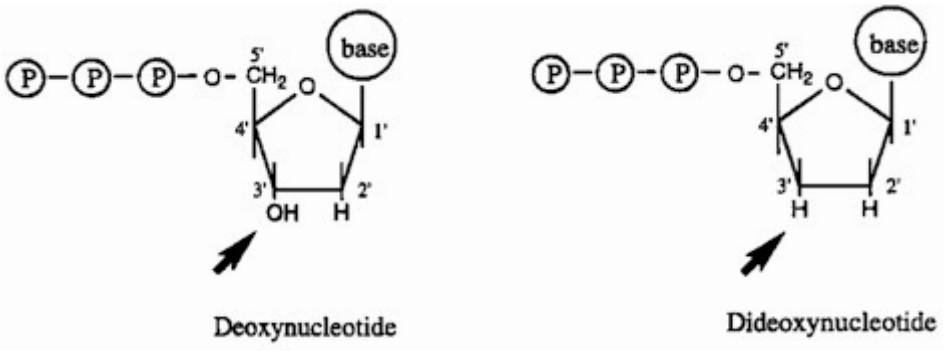

Figure 6: Schematic diagram of deoxynucleotide and dideoxynucleotide. Arrows indicates the single difference in the structures of the two molecules in 3' carbon.

In Sanger sequencing one primer which is complementary to the template DNA, four dNTPs and four ddNTPs in lower concentration than normal dNTPs concentration are used. Before sequencing, DNA has to be initially denatured. To one of template strand the primer is being annealed. The primer or nucleotides are fluorescently or radioactively labeled as final product has to be detected on gel or capillary.

AIM

Identification and comparison of copy number variants in IFNAR1 loci of healthy uninvolved margin of breast tissue surrounding primary tumour and blood derived-DNA from the same patient by using Illumina SNPs array-Human660W-Quad®. To confirm this we have carried out PCR, cloning and sequencing. The long term aim of the project is to investigate genetic variations occurring in enhancer region 4kb upstream of IFNAR1 loci whether these factors influence the development of cancer and if there is link between the predicted enhancer region mutation and disease susceptibility.

MATERIALS AND METHODS

Sample acquisition

DNA samples from the blood and healthy uninvolved margin were acquired from breast cancer patients via collaboration with Oncology Center in Bydgoszcz, Poland. These samples were genotyped using Illumina SNPs array- Human660W-Quad® at University of Alabama at Birmingham. The raw data was analyzed using the Nexus software (version 4).

PCR

PCR amplification was carried out in 25 µl containing final concentrations of template DNA (5 ng/µl), (0.4 µM) of each primer, (0.04 U/µl) of Platinum Taq DNA polymerase (Invitrogen), dNTPs (0.04nM) (Fermentas), 2.5 µl Stofell Buffer (Invitrogen), $MgCl_2$ (2 mM) (Invitrogen) and H_2O.

PCR was carried out using the following program: Initial denaturation at 95°C for 4 min; amplification for 35 cycles for 30 sec at 95°C, 30 sec at 64°C and 1 min at 72°C; and a final extension for 10 min at 72°C in a programmable thermal controller (MJ Research Inc). The sequence of primers is shown in table1.

Primer	Sequence (5´- 3´)	Product length (bp)
Fwd blue primer (STSG607079 L)	CCTAACAGCTGGATAGATTGCC	416
Rev blue primer (STSG607079 R)	CCATGCGTGTATATTCCATACG	416

Table. 1: Primes used for PCR

Gel Electrophoresis

The PCR products were analyzed on 1% agarose gel with 1xTAE to measure their size. Approximate 8 µl of PCR product and 2 µl of loading buffer were added to each well. The electrophoresis was set at 120 voltes for 1 hr. The band size was verified with 1kb ladder (GeneRuler™).

Cloning of PCR Products with plasmid pCR®2.1-TOPO® vector and Transformation Using One Shot® MAX Efficiency® DH5α-T1R Chemically Competent cells

The PCR fragments were cloned using One Shot® DH5α Competent Cells (Invitrogen) which has maximum efficiency. The following steps were carried out for cloning.

To carry out subcloing 2 µl of PCR product was taken to setup the reaction mixture (see table 2).

Reagents	Amount (µl)
PCR product	2 µl
Salt solution	2 µl
Sterile H2O	1 µl
TOPO vector	1 µl
Total	**6 µl**

Table 2. Composition of reaction mixture according to Invitrogen TOPO-TA Cloning kit manual.

The reaction mixture was gently mixed by pipetting up and down several times and incubated at room temperature for 26 min. After incubation the reaction mixture was kept on ice.

2 µl of reaction mixture was added to One Shot® MAX Efficiency® DH5α-T1R Chemically Competent *E. coli* cells (Invitrogen). The reaction mixture was incubated on ice for 20 min followed by heat shock for 30 sec at 42 °C in a water bath and immediate incubation on ice for 2 min. 250 µl of SOC medium was added to reaction mixture which was then placed on a shaker at 160 rpm for 1 hr at 37 °C.

To each LB-agar plate containing the antibiotic kanamycin (Sigma) 50 µg/ml, 83 µl of transformation mixture was added and spread with a sterile L-shaped plastic rod and incubated for overnight at 37°C. 40 single selected colonies were picked from LB-agar plates with a tip. Each clone was added to 5 ml of LB with kanamycin and incubated in a shaker for overnight at 37°C at 250 rpm.

Plasmid DNA purification

Prior to purification of plasmid 1 ml of culture was mixed with 200 µl of glycerol stored at -70°C for further use. Plasmid DNA was purified by using the QIAprep Spin Miniprep Kit, according to manufactures protocol (QIAGEN). The culture medium was added to 2 ml tubes and centrifuged at 10000 rpm for 3 min in micro centrifuge. Supernatant was discarded; to the pellet 250 µl of P1 buffer (Resuspension Buffer) was added and resuspended by vortexing. Then 250 µl P2 buffer (Lysis buffer) was added and mixed thoroughly by inverting the tubes for several times. Buffer P2 was used to lyse bacterial cell membrane to release DNA. 350 µl of N3 (Neutralization Buffer) was added and mixed thoroughly by inverting the tubes 4-6 times until the solution became viscous and cloudy. The tubes were centrifuged at 17400 rpm for 10 min. Supernatant was loaded to QIAprep spin column by pipetting and centrifuged for 1 min and follow through was discarded. The column was washed with 750 µl PE buffer (Wash Buffer) and centrifuged for 1 min, flow through was discarded and centrifuged for additional 1 min to remove the residual wash buffer. Later, 120 µl of water was added to column and let it stand for 1 min. Elution was performed by centrifuging for 1 min. The concentration of extracted DNA was measured by spectrophotometer (NanoDrop).

Confirmation of insert existence and Gel electrophoresis

By using *EcoR*1 restriction enzyme, insert was cut out from plasmid. The composition of reaction mixture was as shown in table 3.

Components	Volume (µl)
EcoR-1 buffer	1 µl
EcoR-1 enzyme	1 µl
Plasmid DNA	8 µl
Total	**10 µl**

Table 3. Composition of restriction digestion reaction mixture.

500 ng of plasmid DNA with 20 000 U/ml of *EcoR*1 enzyme (Bio Labs) and 1 µl of 10x *EcoR*1 Reaction Buffer (Bio Labs) was incubated in a thermocycler to carry out restriction digestion and program was set at 37°C 1 h; 65°C 20 min; 8°C ∞.

Digested plasmid DNA was run on 1% agarose gel at 120V for 2 hrs by using xylene loading dye. The fragments with different insert length were chosen for sequencing.

Sanger Sequencing

Plasmids with different insert lengths were chosen for sequencing. Each plasmid was sequenced in triplicates with eight primes (Table 4).

Primer	Sequence (5´- 3´)
ORANGE_F	GAATCGCTTGAACCCGGAAGG
BLACK_F	CCTGGGTAACACAGCGGAAATCC
M13_R	CAGGAAACAGCTATGAC
LF4_F	GTTGTGGTGAGCCGAGATCG
JD3_F	CAGGAGAATCGCTTGAAC
JD4_R	AATATATACGTATGGAATATA
M13_F	CAGGAAACAGCTATGAC
LF3_R	AGTATGGAATATATACGTAT

Table 4. Primes used for sequencing reaction.

To 96 well plates, 12.5 ng/µl template DNA and 1.6 pmol/µl of each primer in triplicates was added. Plate was spin down for few seconds at 700 rpm followed by initial denaturation at 95^{o}C for 5 minutes in thermocycler. Meanwhile BigDye termination reaction mixture was prepared as follows shown in table 5.

Reagents	Volume (µl)
BigDye Terminator 3.1	4 µl
SD-buffer	2 µl
MQ H_2O	1 µl
Total	**7** µl

Table 5. Composition of sequencing reaction mixture according to Applied Biosystems BigDye terminator 3.1 Cycle Sequencing kit manual.

After initial denaturation of plasmid DNA, sample was cooled on ice and to each well 7 µl of BigDye terminator reaction mixture was added. The Plate was spin down for a few seconds and was placed in thermocycler with a standard sequencing program as shown in table 6.

Temp ^oC	Time	No. Cycles
95	30 sec	
94	25 sec	
50	15 sec	35
60	2 min	
8	Forever	

Table 6. Sequencing program

After sequencing program, samples were precipitated as follows, 1 µl of 125 mM EDTA pH 8 and 1 µl of 3 M sodium acetate pH 5,4 were added to each well. The plate was spin down for a few seconds to ensure that solutions get to bottom of wells. Later 29 µl of 95% ethanol was added to each well; plate was vortexed briefly and incubated in room temperature for 15-minutes in dark. After incubation samples were centrifuged at 1750 g for 45 minutes at 4^{o}C. After precipitation the reaction mixture was removed by inverting the plate upside down and spin down at 300 rpm for 30 sec. Later 70 µl of 70% ethanol was added to each well to remove residual impurities. The plate was vortexed briefly and centrifuged at 1750 g for 15 minutes in 4^{o}C. Ethanol was removed from wells by spinning the plate in inverted position at 700 rpm for 30 seconds and the plate was left for 15 minutes at room temperature to air-dry. As ddNTPs present in BigDye Terminator 3.1 are light sensitive, plate was protected against light during whole procedure. Pellets of sequencing reaction were solved in 10 µl of HiDi Formamide (Applied Biosystems P/N 4311320) and submitted to Uppsala Genome Center (UGC) core facility in order to conduct capillary electrophoresis of samples. The results of capillary electrophoresis was assembled and aligned with CodonCode Aligner software.

RESULTS

Illumina

DNA samples from the blood and healthy uninvolved margin that was acquired from breast cancer patients was genotyped by Illumina SNPs array- Human660W- Quad® at University of Alabama at Birmingham. The results from Illumina and data analysis from Nexus software for blood and uninvolved margin of breast tissue show deletion in healthy uninvolved margin of putative enhancer region 4kb upstream of *IFNAR1* gene of two samples SK58B, KJ42B surrounding primary tumor when compared to SK58K, KJ42K of blood derived DNA. (Figure 6)

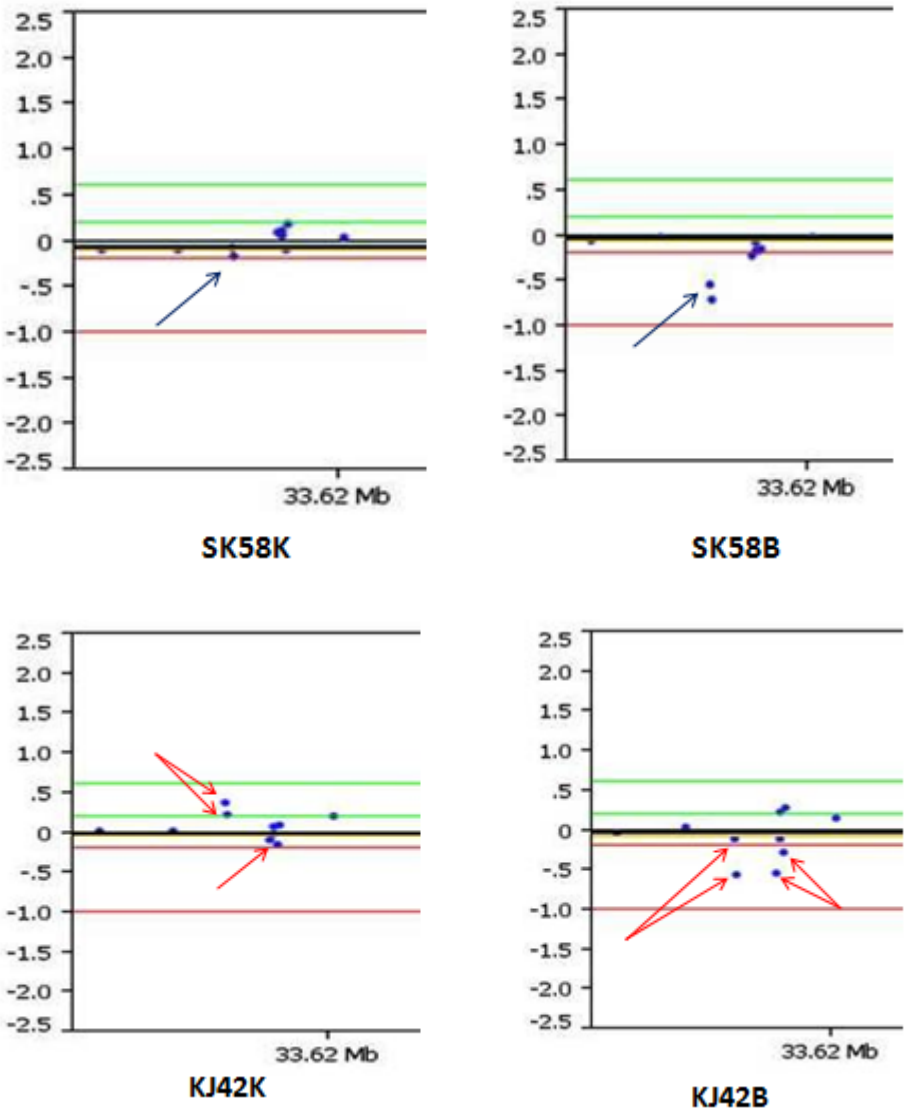

Figure 6: The data analyzed from Illumina SNPs 660W array using Nexus Software of two SNPs in KJ42B and KJ42K Breast cancer patients. On *y* axis Log R ratio value, indicates measured intensity of signal in reference to signal from pool of reference samples and on *x* axis SNPs places on chromosome shown in megabase pairs. In both patients we can see deletion of some SNPs in B samples compare to the same SNPs in samples K, which was used as a reference, marked with arrows. B refers to DNA derived from Healthy uninvolved margin and K refers to DNA derived from blood.

These results suggest there is difference between CNVs in blood and uninvolved margin of breast tissue surrounding primary tumor. These results were confirmed later by cloning and sequencing.

Gel electrophoresis

Based upon the results from Illumina the samples (KJ42K, KJ42B and SK58K, SK48B) was amplified by using primes (table.6) and gel image shows an expected band around 1.4kb in the upstream of IFNAR1 loci (Figure 7).

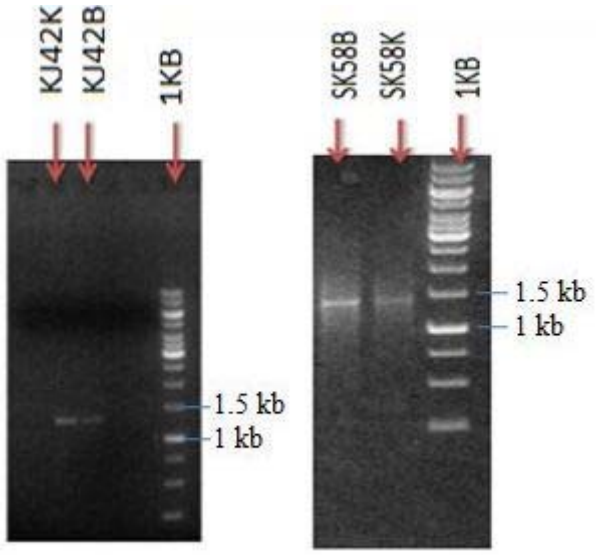

Figure 7: the amplified PCR products of (KJ42K & KJ42B), (SK58K & SK58B) show a band of fragment length around 1.4kb

Restriction digestion and gel electrophoresis

The results from the Illumina confirmed the presence of variation in the samples SK58B and KJ42B DNA derived from the uninvolved margin compared to SK58K and KJ42K DNA derived from the blood of the same patient (Figure 8 & 9).

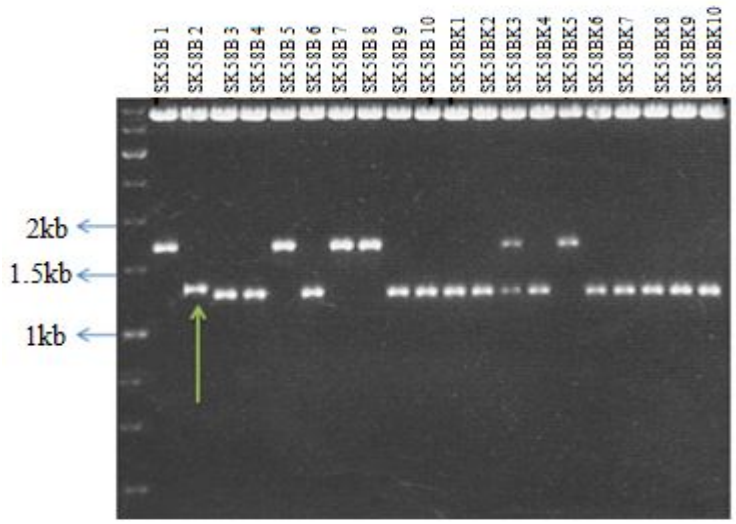

Figure 8: Gel image showing first ten clones on the left are from sample SK58B and the next ten clones are from SK58K. In sample SK58B clone number 2 indicated by arrow shows variation in fragment length which is not presented in any clone of SK58K.

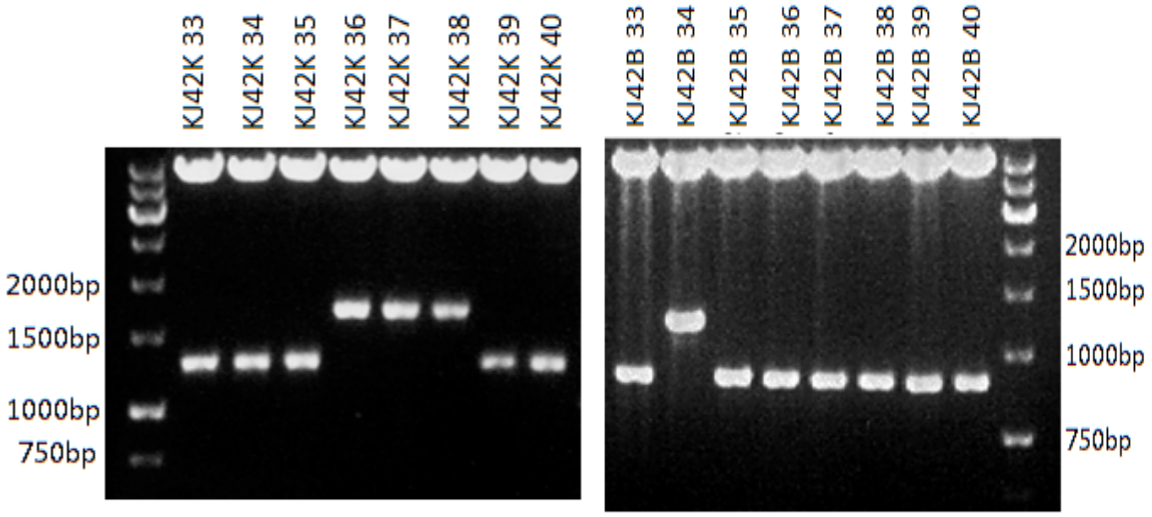

Figure 9: Gel image shows clones numbered from 33 to 40 on the left are from sample KJ42K, showing two variations in length, clone number 33, 34, 35, 39 & 40 around 1400kb and in KJ42B clone number 34 shows variation in size around 1400bp compared to other clones 33,35,36,37,38,39 & 40.

Sequencing

The sequencing results showed presence of three or two different variants in clones from sample SK58B and SK58K respectively (Table6). The fragment length was analyzed by codoncode analyzer for SK58B and SK58K confirmed presence of different fragments lengths in SK58B and two different fragments in SK58K in which one clone is similar to SK58B as shown in the table (7).

Sample	Clone No.	Length in bp (by codoncode aligner)
SK58B	1	1151
	2	710
	3	633
SK58K	1	633
	5	1150

Table 7: Length of clones analyzed with codoncode software. Sample SK58B showed three clones with different sizes and in sample SK58K two clones with different sizes.

The results after sequencing reaction of clones from samples SK58K and SK58B was concordant with gel images. The sequencing reaction was carried out for clone number 33 and 38 from sample KJ42K which shows different sizes on the gel. The obtained result confirms the presence of two different fragments (Table 8).

Sample	Clone No.	Length in bp (by codoncode aligner)
KJ42K	33	1119
	38	633

Table 8: Size of the clones from based on condoncode aligner. Clone 33 and 38 showed different sizes.

As per the results observed, fragments analyzed in the study show variation in size due to presence of VNTRs of 32 repeats with the sequence **ATATGTGGAACCAATACATATAT GCAACATAT.** It also confirms the presence of CNVs in the IFNAR1 loci of uninvolved margin of the breast tissue surrounding the primary tumor. These variations in the gene may lead to the development of the disease.

DISCUSSION

The results from few decades in progression of breast cancer explain <10% of all cases [34, 35]. In this study in a set of few breast cancer samples, a remarkable variation was found in the region connected to IFNAR1 gene by using Illumina SNPs array- Human660W-Quad®. This variation in putative enhancer region of IFNAR1 loci may lead to conflict in gene expression and can guide to disease development [36,37]. To assess if this polymorphism in IFNAR1 gene was related to breast cancer, region was studied by using PCR, subcloning, restriction digestion and sequencing. The genotyping data, gel images of the digested fragments and sequencing showed same results, which indicates presence of CVNs in uninvolved margin of breast tissue surrounding primary tumor and blood of same patient in the study. But results do not specify that the variation in VNTR region leads to development of breast cancer development.

Characterizing changes in DNA copy number is of huge importance in understanding tumorigenesis. While investigating genomes of breast cancer patients from Illumina analysis, numerous genetic variations were identified in the form of tandem repeats in few numbers of samples. The study shows that polymorphism in variable number of tandem repeats of insulin gene in promoter region plays an important role in modulating insulin transcription and diabetes susceptibility [38]. Other findings also recognized role of insulin gene polymorphism in pancreatic carcinogenesis as well as in tumor staging [39]. These studies stress that pathogenic potential of VNTRs not only in size but also in its sequence.

The genetic modifications identified in breast cancer might be associated with IFNAR1 gene. There is no much published data about VNTR polymorphism of IFNAR1 gene related to breast cancer. In this study genetic alteration was identified approximately 4kb upstream of IFNAR1 gene. But this does not appear in all investigated samples in correlation with breast cancer development. However there is genomic imbalance in this region because these aberrations seem to be frequent in numerous cohorts of cancer patients.

Nevertheless, presence of somatic mosaicism was found. The presence of more than one genetically distinct lineage of somatic cells in a single organism is called somatic mosaicism. These variations arise *de-novo* and thus will occur typically only in a few cells. Indeed, present hypothesis in origin of cancer state that, accumulation of somatic mutations leads to unregulated cell division [40].

It is to be expected that rate of discovery for short somatic variations (i.e. CNVs) that are present in small proportion of studied cells will continue to accelerate and studies on genetic heterogeneity of breast malignancies create great opportunities for finding novel biomarkers for the disease progression in breast cancer.

FUTURE PERSPECTIVES

A large number of samples, showing polymorphism in putative enhancer region of *IFNAR*1 gene of uninvolved margin of breast tissue surrounding primary tumor consent to validate expression level of the gene. These results may show if there is any variation in this region which affects the features of the gene.

Development of new methods and already existing techniques allowing analyzing more samples in shorter period of time may contribute to investigations at larger scale (e.g. parallel sequencing) and less cost effect. Several studies demonstrated greater number of aberrations found in primary tumors than in matched metastases [41]. However, these findings demand further investigation.

Studies related to metastasis research in breast cancer are not well developed. Investigations that could compare genetic profiles of primary tumor and uninvolved margin may lead to identification of genomic abreactions in relation to cancer development. On other hand widening the studies on mechanism of metastasis in breast cancer may help in development of new drugs and early diagnostic tests.

ACKNOWLEDGEMENTS

I am heartily thankful to my supervisor, Prof. Jan P. Dumanski, whose encouragement, supervision and support from the preliminary to the concluding level enabled me to develop an understanding of the subject.

Lastly, I offer my regards to Hamid, Lars, Chiara and Geeta who supported me in any respect during the completion of the project.

Last but not the least; I would like to thank my parents, for giving birth to me at the second place and supporting me spiritually throughout my life.

REFERENCES

1. Chiang AC, Massagué J (December 2008). "Molecular basis of metastasis". The New England Journal of Medicine 359 (26): 2814–23.
2. Klein CA (September 2008). "Cancer. The metastasis cascade". Science (New York, N.Y.
3. Chiang AC, Massagué J (December 2008). "Molecular basis of metastasis". The New England Journal of Medicine 359 (26): 2814–23.
4. Wiseman, B.S.W., Z. (2002) Stromal effects on mammary gland development and breast cancer. Science 296, 1046-9.
5. Grange JM, S.J., Stanford CA (2002). "Campbell De Morgan's 'Observations on cancer', and their relevance today". Journal of the Royal Society of Medicine 95 (6): 296–9.
6. Researchers Produce First Sequence Map of Large-Scale Structural Variation in the Human Genome. Retrieved 2009-05-31 genome.gov.
7. Nature. "Mapping and sequencing of structural variation from eight human genomes : Abstract". Retrieved 2009-05-31. Nature. doi:10.1038/nature06862.
8. Edwards, A.C.a.C., Structural variation in the named bronchi of the left lung. A morphometric study. 1983. British Journal of Diseases of the Chest, Volume 77, : pp. 344-348.
9. Baynes, J.W.a.M.H.D., Medical Biochemistry. Second Edition ed. 2005: Elsevier Mosby.
10. Benetkiewicz, M., et al., Chromosome 22 array-CGH profiling of breast cancer delimited minimal common regions of genomic imbalances and revealed frequent intra-tumoral genetic heterogeneity. 2006. Int J Oncol, 29(4): pp. 935-45.
11. Cook EH, Scherer SW (2008). "Copy-number variations associated with neuropsychiatric conditions". Nature 455 (7215): 919–23.
12. Lee JA, C.C., Lupski JR. (2007). "A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders". Cell 131 (7): 1235–47. doi:10.1016/j.cell.2007.11.037. PMID 18160035. reported in "Copy Number Variation May Stem From Replication Misstep". ScienceDaily. 2008-01-04.
13. Hastings PJ, L.J., Rosenberg SM, Ira G. (2009). "Mechanisms of change in gene copy number". Nature Reviews Genetics 10: 551-564.
14. Kidd JM, C.G., Donahue WF, et al. (May 2008). "Mapping and sequencing of structural variation from eight human genomes". Nature 453 (7191): 56–64.
15. Lee J.A., Lupski J.R. Genomic Rearrangements and Gene Copy-Number Alterations as a Cause of Nervous System Disorders (2006) Neuron, 52 (1), pp. 103-121.
16. Human Genetic Variation Fact Sheet". July 2008. National Institute of General Medical Sciences (NIH). Retrieved 2008-08-16.
17. Uzé G, S.G., Piehler J, Pellegrini S (2007). "The receptor of the type I interferon family". Curr. Top. Microbiol. Immunol. 316: 71–95. doi:10.1007/978-3-540-71329-6_5. PMID 17969444.
18. Klaus, W., Gsell, B., Labhardt, A. M., Wipf, B., and Senn, H. (1997) J. Mol. and Biol. 274.
19. Stark, G.R., Kerr, I. M., Williams, B. R., Silverman, R. H., and Schreiber, and R. D. (1998) Annu. Rev. Biochem. 67.
20. Leyva, L., et al., IFNAR1 and IFNAR2 polymorphisms confer susceptibility to multiple sclerosis but not to interferon-beta treatment response. J Neuroimmunol, 2005. 163(1-2): p. 165-71.
21. Diop, G., et al., 2006. Exhaustive genotyping of the interferon alpha receptor 1 (IFNAR1) gene and association of an IFNAR1 protein variant with AIDS progression or susceptibility to HIV-1 infection in a French AIDS cohort. Biomed Pharmacother, 60(9): p. 569-77.

22. Saito, T., et al., 2004. Genetic variations in humans associated with differences in the course of hepatitis C. Biochem Biophys Res Commun. 317(2): p. 335-41.

23. Aucan, C., et al., 2003. Interferon-alpha receptor-1 (IFNAR1) variants are associated with protection against cerebral malaria in the Gambia. Genes Immun. 4(4): p. 275-82.

24. Critchley-Thorne, R.J., et al., 2009. Impaired interferon signaling is a common immune defect in human cancer. Proc Natl Acad Sci U S A. 106(22): p. 9010-5.

25. Abramovich, C., Ratovitski, E., Lundgren, E., and Revel, M. (1994) FEBS Lett. 338, 295–300.

26. Krishna K. Gopaul, T.J.B., Andrea L. Gibson, and a.F.A.D. Malcolm D. Yates, July 2006, Progression Toward an Improved DNA Amplification-Based Typing Technique in the Study of Mycobacterium tuberculosis Epidemiology. Journal of Clinical Microbiology, pp. 2492–2498.

27. Goguet de la Salmoniere, Y., H. M. Li, G. Torrea, A. Bunschoten, J. van Embden, and B. Gicquel. 1997. Evaluation of spoligotyping in a study of the transmission of Mycobacterium tuberculosis. J. Clin. Microbiol. 35:2210–2214. 29.

28. Palin, K., J. Taipale, and E. Ukkonen, 2006. Locating potential enhancer elements by comparative genomics using the EEL software. Nat Protoc, 1(1): pp. 368-74.

29. Conrad, D.F., et al., 2009. epub Oct 7. Origins and functional impact of copy number variation in the human genome. Nature.

30. Bickle TA, Krüger DH (June 1993). "Biology of DNA restriction" Microbiol. Rev. 57 (2): 434–50.

31. Danna K, Nathans D (December 1971). "Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae"

32. Maxam AM, Gilbert W (February 1977). "A new method for sequencing DNA". Proc. Natl. Acad. Sci. U.S.A. 74 (2): 560–4.

33. Speed, Terrence. Sanger's Method. (Feb. 10th, 2003).

34. Duncan, J.A., J.R. Reeves, and T.G. Cooke, 1998. BRCA1 and BRCA2 proteins: roles in health and disease. Mol Pathol. 51(5): pp. 237-47.

35. Hynes, N.E. and J.H. Dey, Potential for targeting the fibroblast growth factor receptors in breast cancer. Cancer Res. 70(13): pp. 5199-202.

36. Paquette, J., et al., 1998. The INS 5' variable number of tandem repeats is associated with IGF2 expression in humans. J Biol Chem, 273(23): pp. 14158-64.

37. McGeary, J., 2009. The DRD4 exon 3 VNTR polymorphism and addiction-related phenotypes: a review. Pharmacol Biochem Behav. 93(3): pp. 222-9.

38. Krechler, T., et al., 2009. Polymorphism -23HPhI in the promoter of insulin gene and pancreatic cancer: a pilot study. Neoplasma. 56(1): pp. 26-32.

39. Jackson, A L & Loeb, L.A. (1998). The mutation rate and cancer. Genetics 148. 1483- 1490.

40, Stadler, Z.K. and S.E. Come, 2009. Review of gene-expression profiling and its clinical use in breast cancer. Crit Rev Oncol Hematol. 69(1): pp. 1-11.

41. Torres, L., et al., 2007. Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. Breast Cancer Res Treat, 102(2): pp. 143-55.