|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Results My first step of the project was to identify chromosomal aberrations for breast cancer. At CGAP, I was able to find a long list of many chromosomal aberrations in breast cancer. The list is at http://cgap.nci.nih.gov/Chromosomes/Mitel\_Search  Here are the first 50 chromosomal aberration loci:  The numbers on the left specify the cytogenetic location, and the numbers on the far right specify the number of breast cancer cases in which these aberrations were found.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Band | Locus | Cancer Type | Tissue | Cases | | 1q10 | I(1)(q10) | Adenocarcinoma | Breast | 56 | | 1q10 | der(1;16)(q10;p10) | Adenocarcinoma | Breast | 48 | | 16p10 | der(1;16)(q10;p10) | Adenocarcinoma | Breast | 48 | | 12q24 | add(12)(q24) | Adenocarcinoma | Breast | 25 | | 1p13 | del(1)(p13) | Adenocarcinoma | Breast | 23 | | 1p22 | del(1)(p22) | Adenocarcinoma | Breast | 21 | | 1p36 | add(1)(p36) | Adenocarcinoma | Breast | 21 | | 11p15 | add(11)(p15) | Adenocarcinoma | Breast | 21 | | 1p11 | add(1)(p11) | Adenocarcinoma | Breast | 19 | | 8q10 | i(8)(q10) | Adenocarcinoma | Breast | 19 | | 6q25 | del(6)(q25) | Benign epithelial tumor, NOS | Breast | 18 | | 3p12 | del(3)(p12) | Adenocarcinoma | Breast | 14 | | 7p22 | add(7)(p22) | Adenocarcinoma | Breast | 14 | | 17p11 | add(17)(p11) | Adenocarcinoma | Breast | 14 | | 19q13 | add(19)(q13) | Adenocarcinoma | Breast | 14 | | 1q12 | del(1)(q12) | Adenocarcinoma | Breast | 13 | | 3p12 | del(3)(p12p14) | Adenocarcinoma | Breast | 13 | | 3p14 | del(3)(p12p14) | Adenocarcinoma | Breast | 13 | | 12p13 | add(12)(p13) | Adenocarcinoma | Breast | 13 | | 8q24 | add(8)(q24) | Adenocarcinoma | Breast | 12 | | 1q11 | del(1)(q11) | Adenocarcinoma | Breast | 11 | | 3p13 | del(3)(p13) | Adenocarcinoma | Breast | 11 | | 6q21 | del(6)(q21) | Adenocarcinoma | Breast | 11 | | 15p11 | add(15)(p11) | Adenocarcinoma | Breast | 11 | | 2q37 | add(2)(q37) | Adenocarcinoma | Breast | 10 | | 11q23 | del(11)(q23) | Adenocarcinoma | Breast | 10 | | 17q10 | i(17)(q10) | Adenocarcinoma | Breast | 10 | | 1p21 | del(1)(p21) | Adenocarcinoma | Breast | 9 | | 3p13 | del(3)(p13p14) | Adenocarcinoma | Breast | 9 | | 3p14 | del(3)(p13p14) | Adenocarcinoma | Breast | 9 | | 3q13 | del(3)(q13) | Adenocarcinoma | Breast | 9 | | 6p10 | i(6)(p10) | Adenocarcinoma | Breast | 9 | | 6q16 | del(6)(q16) | Adenocarcinoma | Breast | 9 | | 6q23 | del(6)(q23) | Adenocarcinoma | Breast | 9 | | 6q24 | del(6)(q24) | Carcinoma, NOS | Breast | 9 | | 7q32 | del(7)(q32) | Adenocarcinoma | Breast | 9 | | 11p11 | add(11)(p11) | Adenocarcinoma | Breast | 9 | | 14p11 | add(14)(p11) | Adenocarcinoma | Breast | 9 | | 16p13 | add(16)(p13) | Adenocarcinoma | Breast | 9 | | 1p13 | add(1)(p13) | Adenocarcinoma | Breast | 8 | | 3p11 | add(3)(p11) | Adenocarcinoma | Breast | 8 | | 3p13 | add(3)(p13) | Adenocarcinoma | Breast | 8 | | 3p14 | del(3)(p14) | Adenocarcinoma | Breast | 8 | | 5p15 | add(5)(p15) | Adenocarcinoma | Breast | 8 | | 6q25 | del(6)(q25) | Carcinoma, NOS | Breast | 8 | | 9p11 | add(9)(p11) | Adenocarcinoma | Breast | 8 | | 11q14 | del(11)(q14) | Adenocarcinoma | Breast | 8 | | 19p13 | add(19)(p13) | Adenocarcinoma | Breast | 8 | | 20q13 | add(20)(q13) | Adenocarcinoma | Breast | 8 | | 1p12 | add(1)(p12) | Adenocarcinoma | Breast | 7 |   There were 358 aberrant loci on the original list. I then sorted by frequency of occurrence. I first picked 11p15 as a test case. Then I studied 1p36 because it was also high on the list and not a centromeric region, like 1q10 which doesn�t have many genes. My dad suggested 19p13 and 19q13 because these regions have genes that his company is studying in prostate cancer. Finally, I picked 8q24 because it is mentioned in a science paper, Science 294, pg. 1343, "A Phosphatase Associated with Metastasis of Colorectal Cancer". If metastasis is related to 8q24 in colorectal cancer, then it is possibly for breast cancer as well.  This is the ideogram (diagram of a chromosome) of chromosome 11. Note that 11p15 has five subbands and is far from the centromeric location.  11p15  I first compiled data and did a statistical analysis on 11p15. At this band, the DNA tends to be amplified, (meaning there may be duplicate or triplet copies of some of the genes there) in cancer according to the Cancer Genome Anatomy Project website. I downloaded the list of genes that have been mapped to this cytogenetic band. For each of the 100 genes, I opened up Virtual Northern and copied by hand the expression data into an excel spread sheet. For the mammary gland, there were a total of 36,030 normal tissue EST�s sequenced, and 61,513 cancer tissues sequenced. These totals represent the number of messages expressed from all the genes active in the mammary gland. A single gene might express just one message or a few hundred. For each of the 100 genes I entered the number of messages made by both the normal tissue and the cancer tissue. I then used a chi square test to determine if the gene was significantly differentially expressed between normal and cancer. After doing all 100 genes, I sorted the spread sheet on the basis of the P values, smallest to highest. I studied with the smallest P values to determine which genes may play a significant role at the 11p15 location in cancer. My next step is to determine which genes seem to be interesting to study further. Certain genes can be targeted by drugs. Unfortunately, many genes are difficult to target with drugs because the genes also serve vital functions in other parts of the body. For example, gene HBB, is for a beta hemoglobin which is important for all blood cells. It has a Pvalue of (2.49 \* 10^-6) %. The level of expression of this gene went from 15 out of 36,030 in normal tissue to 1 message out of 61,513 in cancer tissue representing a twenty-five-fold decrease in gene expression. This may have been from a contamination of blood cells in the sample.  Since the aberration showed an amplification, I am looking for increases in gene expressions, not decreases.  What is the significance of this? What other genes are important? Back to the spread sheet:  I noticed TSG101, a tumor susceptibility gene. The level of expression was 0 times out of 36,030 for normal tissue and 7 times out of 61,513 for cancer tissue. This is an addition in expression for cancer, which is exactly what I am looking for in the 11p15 cytogenetic band. I went to Locus Link, a website that contains information about many genes. Here is the summary of TSG101:  ***Summary:*** *The protein encoded by this gene belongs to a group of apparently inactive homologs of ubiquitin-conjugating enzymes. The gene product contains a coiled-coil domain that interacts with stathmin, a cytosolic phosphoprotein implicated in tumorigenesis. The protein may play a role in cell growth and differentiation and act as a negative growth regulator. In vitro steady-state expression of this tumor susceptibility gene appears to be important for maintenance of genomic stability and cell cycle regulation. Mutations and alternative splicing in this gene occur in high frequency in breast cancer and suggest that defects occur during breast cancer tumorigenesis and/or progression.*  This gene is a winner. It mentions breast cancer tumorigenesis. This means that TSG101 has already been discovered to be a negative growth regulator. It is known as a tumor susceptibility gene, so if mutated, its absence will likely contribute to the growth of a tumor. The fact that this gene makes a cytosolic phosphoprotein is also good news. Cytosolic means the protein is found in the cytosol, meaning it is easily accessed by a lipid-soluble drug.  I have found that there is another gene located at 11p15, called TALDO1, or transaldolase 1, which is a tumor susceptibility gene. The level of expression was 1 out of 36030 for normal tissue, and 9 out of 61504 for cancerous tissue. The pvalue is 0.07. I went to Locus Link, and here is what they say about it:  *Transaldolase 1 is a key enzyme of the nonoxidative pentose phosphate pathway providing ribose-5-phosphate for nucleic acid synthesis and NADPH for lipid biosynthesis. This pathway can also maintain glutathione at a reduced state and thus protect sulfhydryl groups and cellular integrity from oxygen radicals. The functional gene of transaldolase 1 is located on chromosome 11 and a pseudogene is identified on chromosome 1 but there are conflicting map locations. The second and third exons of this gene were developed by insertion of a retrotransposable element. This gene is thought to be involved in multiple sclerosis.* ***Summary:*** *Protein related to transaldolase; catalyzes formation of fructose-6-P and erythrose-4-P from sedoheptulose-7-P and glyceraldehyde-3-P in the pentose phosphate pathway, may transfer aldol unit from sedoheptulose-7-P to glyceraldehyde-3-P*  This description shows me that this is not a very good candidate for a drug target, because it is involved with nucleic acid synthesis and respiration so this is a vital part of cell life. This wouldn�t make a good drug target because it would hurt many cells.  The gene at the top of the 11p15 spread sheet with the lowest p-value is LDHA, lactate dehydrogenase. This is an enzyme that breaks down the sugar lactose. I am assuming that the only reason it is so abundant in cancer cells is that they are developing very rapidly, and so generate more lactate dehydrogenase. This would make a bad drug target because it would hurt the normal cells by destroying their normal amount of lactose.  The gene with the second lowest p-value is HBB, a beta hemoglobin protein. This gene may be present only because of a contamination of a blood cell into the data. A drug target for hemoglobin would also be a bad idea because it would target all blood cells.  The two genes with the lowest p values are most likely errors! Oh well, back to the spread sheet:  The next two genes that interest me are the ribosomal genes. On the excel spread sheet I found two ribosomal genes, S13 and L27a, that were both highly expressed in cancer. The reason they are overly expressed is because the cancer cells are dividing rapidly and many ribosomes are making more mRNA than usual.  The 14th lowest p-value is TSSC1, a tumor suppressing gene. Its official name is tumor suppressing subtransferable candidate 1. The LocusLink report says that it assists in regulating cell division.  [Here is an example of Virtual Northern for TSSC1](http://cgap.nci.nih.gov/Tissues/VirtualNorthern?TEXT=0&ORG=Hs&CID=4992), I scroll to mammary gland to obtain the EST data (I did not use the SAGE data). Notice that the p value at "mammary gland" is red to show that it is differentially expressed:  Note: When I did this in December, CGAP did not automatically list the p values and numbers, and that is why I had to enter the data into Excel to do the calculations. Possibly our request for the chi squared equation test prompted CGAP to include this calculation here.  Then I went back to the spread sheet:  Two proteasome genes came up as 18th and 19th on my list. According to the Molecular Cell Biology Text book*,*  *"a proteasome is a large multifunctional protease complex in the cytosol that degrades intracellular proteins marked for destruction by attachment of multiple ubiquitin molecules. Ubiquitin is a small highly conserved protein that becomes covalently linked to lysine residues in other intracellular proteins. Proteins to which a chain of ubiquitin molecules is added usually are degraded in a proteasome."*  Proteasomes may be a good target because they are also found in the cytosol.  After I did this more thorough analysis of 11p15, I decided that it would be more useful to find patterns between all five aberration loci, instead of looking at individual genes for each locus. Right away I noticed that every aberration spread sheet had ribosomal genes with very low p values. This was good news to me because I had noticed the 11p15 ribosomal genes early on in the project. I continued my project with a concentration only on genes that start with RPL, meaning it is a ribosomal gene according to Locus Link.  First I analyzed 19p13, which I�m glad has some good results because there were 134 genes to statistically analyze!  Gene RPL18A is the second on the excel spread sheet with a P-value of 0.0000. Here is what Locus Link has to say about it: (I have copied the whole webpage so you can see what it is like to enter a gene into Locus Link.) All that is needed is to type the gene name, like RPL18A into the Query, press go, and you get this summary.  [Locus Link](http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=6142) - The actual page where I located gene summaries.  My next cytogenetic location I studied is 1p36. RPL11 came up fourth on the list and RPL22 came up twenty-ninth.  RPL11�s summary is  **Summary:** Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L5P family of ribosomal proteins. It is located in the cytoplasm. The protein probably associates with the 5S rRNA. Alternative splice variants encoding different isoforms may exist, but they have not been fully characterized. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome.  RPL22�s summary is:  Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a cytoplasmic ribosomal protein that is a component of the 60S subunit. The protein belongs to the L22E family of ribosomal proteins. Its initiating methionine residue is post-translationally removed. The protein can bind specifically to Epstein-Barr virus-encoded RNAs (EBERs) 1 and 2. The mouse protein has been shown to be capable of binding to heparin. Transcript variants utilizing alternative polyA signals exist. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome. It was previously thought that this gene mapped to 3q26 and that it was fused to the acute myeloid leukemia 1 (AML1) gene located at 21q22 in some therapy-related myelodysplastic syndrome patients with 3;21 translocations; however, these fusions actually involve a ribosomal protein L22 pseudogene located at 3q26, and this gene actually maps to 1p36.3-p36.2.  For Chromosome Aberration 8q24, the gene on the top of the excel spread sheet with the lowest p value is RPL8. I have to find out the significance of these ribosomal genes. Here is Locus Link�s summary:  Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L2P family of ribosomal proteins. It is located in the cytoplasm. In rat, the protein associates with the 5.8S rRNA, very likely participates in the binding of aminoacyl-tRNA, and is a constituent of the elongation factor 2-binding site at the ribosomal subunit interface. Alternatively spliced transcript variants encoding the same protein exist. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome.  For Chromosome aberration 19q13, there is not a ribosomal gene until the 48th row. I will still do an analysis of it:  Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L13P family of ribosomal proteins. It is located in the cytoplasm. Transcript variants utilizing alternative polyA signals have been observed. This gene is co-transcribed with the small nucleolar RNA genes U32, U33, U34, and U35, which are located in its second, fourth, fifth, and sixth introns, respectively. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome. |
|  |