# December 8, 2001

I have studied the CGAP website and I know that I am going to concentrate on breast cancer aberrations, including 11p15, 19p13, 19q13, 1p36, and 8q24. During the next month or so I will obtain the data from CGAP and put it into my excel charts.

# February 19, 2002

# Research, reading, entering data into excel charts.

# March 3, 2002

So far I have analyzed one aberration thoroughly: 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 107 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 107 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 107 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. I will look into stathmin, the cytosolic phosphoprotein. Cytosolic means the protein is found in the cytosol, meaning it is easily accessed by a lipid-soluble drug. A phosphoprotein is made up of certain phosphogroups. | | |  | | |
|  | | |  |  | |
|  | | |  |  | |
|  | | |  |  | |
|  | | |  |  |  |
|  |  | | | | |
|  |  | | | | |

March 23, 2002

Today I moved all my data to my laptop. Now I am going to take a break in the data analysis, and finish what should’ve been my first step: the research.

I’m going to finish my introduction, bibliography, and a continuation in my study of aberrations at 19p13, 19q13, 1p36, 8q24, and continue studying 11p15.

March 30, 2002

Today I am still studying 11p15. I have found that there is another gene 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 exon of this gene were developed by insertion of a retrotransposable element. This gene is thought to be involved in multiple sclerosis. | | |
|  | | |
|  | [**Proteome**](http://www.proteome.com/databases/HumanPD/reports/6888.html) **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 | |
|  | **Locus Type:** | gene with protein product, function known or inferred |
|  | **Product:** | transaldolase 1 |
|  | **Alternate Symbols:** | TAL, TAL-H, TALDOR |
|  | **Alias:** | glycerone transferase  dihydroxyacetone transferase |

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 target.

March 31, 2002

I am going to finish off 11p15 today, and then continue the other aberrations with a concentration on ribosomal genes, which start with RPL (Ribosomal Protein) because ribosomal genes assist in the development of proteins which may be a factor in allowing certain cells to become cancerous.

The gene at 11p15 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 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. This is an interesting topic to pursue because there may be potential drugs that will target ribosomal genes.

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, but I don’t understand why a regulatory gene would be more frequent in cancer than in normal tissue. I am confused with this same situation for TSG101.

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 possibly be a good drug target, but I am not sure.

I think that’s it for 11p15 for now. Tomorrow I will look at Excel spread sheets of other aberrations, with a concentration on ribosomal genes.

April 1, 2002

Today I did much work on my introduction, but unfortunately, I was only able to finish writing up the research of cancer. I still have to explain about cytogenetic bands, chromosome aberrations, and how exactly I go about my data collection on CGAP. I still need to discuss my observations and conclusions. My main observation is of the ribosomal genes. Tomorrow I will do a lot of work!

April 3, 2002

I think I am pretty much done with my introduction. I still need to organize my research a lot though, and write up my conclusion. I am learning a lot and enjoying myself with this project. I finally feel that I have a grasp on the material. Today I explained my procedure thoroughly.

April 4, 2002

I have finally organized my data in a way that I like. My excel spread sheets look organized. Now I will write up my results, and start writing up my observations and conclusion. Then all I will need to do is make sure everything looks nice and turn in the floppy disk on Monday.

April 5, 2002

I finished up my results, and wrote a terse conclusion. Now all I have to do are the “tweaking” steps.

April 6, 2002

Project complete! I revised my conclusion and now it actually seems substantial! Hand in floppy disk on Monday to Mr. Thiel.