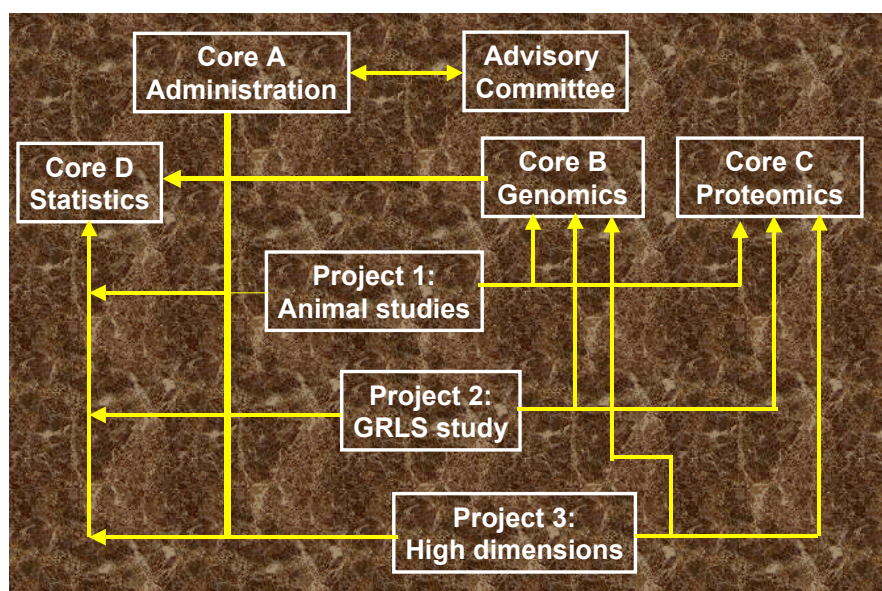The UAB-NCCC Center for Nutrient-Gene Interaction (CNGI) in Cancer Prevention was established in 2002 with a planning grant from the NCI to develop multidisciplinary approaches to investigate the changes in gene and protein expression/abundance in the mammary gland and prostate that occur during the early stages of puberty and are associated with the risk of cancer. In particular, a hypothesis was created that specific dietary components, the polyphenols, alter gene expression and protein abundances and forms in the pathways of biosynthesis and metabolism of estrogenic steroids. A Center proposal was submitted to NCI in July, 2002 and subsequently a U54 award was made. Work in the Center began in September, 2003. The Center consists of project leaders from the UAB School of Medicine (Coral Lamartiniere), UAB School of Public Health (David Allison), and from the Northern California Cancer Center (Pam Horn-Ross). Besides UAB and NCCC, statistical support for the Center is provided by Dr. Gary Gadbury, University of Missouri-Rolla. Individual UAB departmental units involved in CNGI are the Departments of Biostatistics (Section on Statistical Genetics), Environmental Health Sciences, Medicine (Divisions of Clinical Immunology and Arthritis and Genetics and Translational Medicine), Microbiology, and Pharmacology & Toxicology.

Three main polyphenols are the basis of the investigations, genistein from soy, epigallocatechin-gallate (EGCG) from green tea, and resveratrol from red grapes. In the first 18 months of the Center experiments have been carried out to determine which of these polyphenols administered pre-pubertally are chemopreventive using a rat model of breast cancer (**project 1**). Separate studies have been carried out to more specifically investigate the effects of these polyphenols on gene expression and protein abundance in the mammary gland using microarray analysis and two dimensional isoelectric focusing/SDS-PAGE, peptide mass fingerprinting and nanoLC-tandem mass spectrometry. To examine the effect of soy isoflavones on puberty, an observational cohort study is being conducted that involves following young girls going through puberty who are either low or high soy consumers (**project 2**). The effects of polymorphisms in genes involved in estrogen biosynthesis and metabolism will also be examined. Recruitment of the cohort began in April 2005. A characteristic of each of these studies is the highly dimensional nature of the data with a strong tendency towards a large false discovery rate. **Project 3** carefully examines the validity of the statistical methods that are widely used to analyze high dimensional data.

The UAB-NCCC CNGI consists of three main projects carrying out basic research and four cores that support it. **Core A** organizes all the interdisciplinary aspects of CNGI and supports a seminar program, pilot/feasibility studies and career development. An external committee provides advisory opinions on the progress and directions of CNGI.

Core A

Administration
Stephen Barnes, PhD, Director and Center Director
sbarnes@uab.edu
Clinton J Grubbs, PhD, Assistant Director
clintongrubbs@uab.edu
Rose Johnson, Office Associate
rosej@uab.edu

CORE A PROGRESS AND PLANS

Accomplishments
The Core has held regular monthly meetings of project leaders, core directors, pilot project awardees and other investigators. It also has organized a quantitative pathways group.

Year 02 Pilot/feasibility project recipients

- Ada Elgavish "*Prostate cancer prevention by epigallocatechin-3-gallate*"
- Grier Page "*Are there common mechanisms of action for phytoestrogens?*
- Mark Garfinkel (2nd year funding) "*Dietary polyphenol effects on sirtuin epigenetics in Drosophila''*

Seminar Program
Ray Rodriguez (UC-Davis) "*Nutritional Genomics: Towards Individualized Nutrition and Genome-Based Dietary Recommendations*"

Heather Kleiner (LSU) – "*Effects of the Tropical Ginger Compound, 1' -acetoxychavicol Acetate, on Human Breast Cancer Cells*"

Keith Dawson (UC-Davis) "*Reducing the complexity of complex datasets in Nutrigenomics''*

Jim Kaput (UC-Davis) "*Diet X Gene Interactions: An Experimental Paradigm*"

Marilyn Morris (U. Buffalo) "*Flavonoid-Drug Interactions: Effects of flavonoids on ABC transporters*"

Tom Badger (U. Arkansas) "*Soy Infant Formula: What Are The Long-term Health Effects?*"

Xiao-Ou Shu (Vanderbilt University) "*Soy Food and women's health*"

External advisory Committee

The External Advisory Committee performs an important consultative function for CNGI and most recently met on April 27th, 2005 at UAB. The 2005 committee consisted of Dr. Kei Cheung (Chair, Yale University - bioinformatics), Dr. Marc Goodman (Cancer Center of Hawaii – molecular epidemiology), Dr. Jose Russo (Foxchase Medical Center – animal models of mammary cancer), Dr. Clive Slaughter (St. Jude's Hospital – proteomics), and Dr. Hongyou Zhao (Yale - microarray analysis). Two UAB faculty, Dr. Don Hill (Cancer Chemoprevention) and Dr. Joe Messina (Endocrinology) were also members of the committee. Dr. John Milner represented NCI.

Core B

Genomics
Lisa Guay-Woodford, MD, Director
lgw@uab.edu
Michael Crowley, PhD, Assistant Director
mcrowley@uab.edu
Haitao Li, MD, Bioinformatics
haitao@uab.edu

Microarray analysis
The new high-density rat expression arrays developed by Affymetrix (230 2.0) provide the most robust platform for the analysis of gene expression for the experiments in Project 1. The UAB Cancer Center core facility has the current Affymetrix scanner and software and the microarray analysis was outsourced to this facility.

Genotyping
DNA has been isolated from buccal swabs obtained from the subjects enrolled in **Project 2**. Whole genome amplification (WGA) has been tested due to the limiting amount of DNA isolated from buccal swabs and the amount of initial SNPs (43) to be analyzed. Amplification

was done using the GenomiPhi DNA Amplification system from Amersham which yielded up to a 5000 fold increase in DNA. Sequence tag site (STS) markers were selected for each chromosomal arm that an estrogen biosynthetic gene resides to analyze amplification bias. The markers chosen were heterogeneous in size from sample to sample but were not heterozygous in the samples analyzed. Therefore amplification bias could not be ascertained with these markers. New markers will be analyzed for heterozygosity and WGA DNA will be screened for representational amplification. Meanwhile, several SNPs that comprise halpotype blocks have been identified for each of the genes that comprise the estrogen biosynthetic pathway. Currently **Core B** is evaluating SNPs in the cytochrome p450, family 17 A-1 gene by pyrosequencing.

Core C

Proteomics
Helen Kim, PhD, Director
helenkim@uab.edu
Stephen Barnes, PhD, Co-Director
sbarnes@uab.edu

Optimization of 2D-gel parameters for analysis of the rat mammary tissues:

The standard combination of 150 μg protein load, pH 4-7 gradient and 2nd dimension 8-16% acrylamide gradient gel parameters generated good resolution of multiple protein spots in the rat mammary tissues. However, there was detectable "piling up" of proteins at the basic end, and noticeable horizontal streaking of abundant proteins, probably albumin, in the 80 kDa range. To minimize the loss of proteins at the basic end, particularly since several components of the steroid synthesis and metabolism pathway have basic pI's, the broader 3-10 pH range would be more appropriate for resolution of proteins in mammary tissues. Overall, 2D-analysis of mammary tissues will require some additional cleanup steps to deplete the major proteins including albumin.

Analysis of 21-day and 50-day rat mammary tissue samples:

Mammary tissues were harvested in **Project 1** from rats maintained on vehicle, genistein, resveratrol, or EGCG from birth. **Core D** randomized the entire set of 21 and 50 day samples, so that on any given day, 12 IPG strips or 2nd dimension gels were run. All gels were imaged using the Perkin Elmer ProExpress imager. Image analysis was carried out using Progenesis Discovery. Some difficulties were encountered using the Progenesis

Estrogen pathways bioinformatics tools

In a collaborative effort with the **Biostatistics-Bioinformatics Core D**, the **Genomics Core** has developed a web-based schema of the estrogen synthesis and metabolism pathways that incorporates gene-specific SNP content. This resource is designed to be iteratively updated as new SNP information becomes available from the NCBI, Santa Cruz, and Celera databases and will provide a "real-time" SNP compendium to support **Project 2** as well as relevant pilot projects as they are developed. When fully operational, this resource will be made available to the greater research community via a web-based interface accessible through the CNGI website.

CORE C SPECIFIC AIMS:

- Perform 2D-IEF/SDS-PAGE separation of proteins and identify them using mass spectrometry from:
  – mammary glands from rat experiments in project 1
  – urine samples from project 2
- Carry out evaluation of new proteomics methods and their reproducibility with Core D including image analysis

software and ongoing systematic experiments are being carried out with **Core D**.

SDS-PAGE analysis of the IEF buffer-insoluble fraction of rat mammary proteome:

This was carried out, following up suggestions by the External Advisory Committee in the –01 year. Several major bands were observed; however, peptide mass fingerprinting and nanoLC-tandem mass spectrometry revealed that these were nuclear or cytoskeletal proteins and not those from membranes. Alternative methods based on the isolation of specific membrane fractions and those to select from membrane protein complexes are being studied.

2D gels of human urines:
To analyze urines from **Project 2**, it was important to determine several parameters. These were: (1) which collection time gives the most consistent proteomic pattern of urine proteins among different donors; (2) what is the stability of a given urine sample, stored for different lengths of time at 4°C; (3) for a given donor, what is the variation in proteome pattern between different time points. These are issues that impact on time of collection by **Project 2**, whether more than one collection will be required, whether the different times of collection will be pooled, or analyzed separately, and whether a given urine sample

can be left in the refrigerator for extended periods of time.

Core D

Biostatistics/Bioinformatics

David Allison, PhD, Director
dallison@uab.edu
Grier Page, PhD, Co-Director
gpage@uab.edu
Renee Desmond, PhD, statistician
desmond@uab.edu
Ann Loraine, PhD, bioinformaticist
aloraine@uab.edu
Mahyar Sabripour, PhD
msabri@uab.edu
Haitao Li, MD, bioinformatics
haitao@uab.edu
Sreelath Meleth, PhD, Statistician
dhanya@uab.edu
Dung-tsa Chen, PhD, Statistician
dtchen@uab.edu

CORE D SPECIFIC AIMS:

- Provide statistical expertise in study design;
- Provide existing analytical techniques for microarray data, including design, extraction of data, correction for background, normalization/standardization, and statistical analysis;
- Implement novel statistical methodologies developed in Project 3 for the analyses of microarray data;
- Provide statistical analyses for the individual cell-culture or animal model CNGI projects using contemporary statistical methodologies;
- Develop and manage a data management, dissemination, and retrieval systems for all CNGI data;
- Design and implement web interfaces to the databases and provide bioinformatics support such as sequence analysis and search of NCBI and Celera databases as needed;
- Provide the computing resources (software and hardware) to support the CNGI

Project 1:

We have taken a number of actions to assess our ability to extract reliable and meaningful information from the 2D gels for subsequent formal statistical analysis. To date, results have been troubling and we have unearthed substantial limitations in expensive commercially available software. We have developed a thorough plan to work around this and have a project underway to carefully evaluate our measurement procedures on a subset of 32 gels. Once this established (expected in $\leq$ 45 days), we will formally analyze the first major dataset containing 96 gels. We also plan to host Dr. Mike Rogers from the Department of Medical Imaging in Manchester, UK to visit UAB from July 18-22 to serve as a consultant for proteomic issues.

Specimens for 21 day and 50 day experiments from Project 1 have been run following a randomization plan in the laboratory of Dr. Lihong Teng on the Affymetrix platform. Data were delivered to **Core D** for analysis. Preprocessing included examination of 3' to 5' ratios using Bioconductor and flagging outliers. HDBSTAT (http://www.soph.uab.edu/ssg_content.asp?id=1164) was utilized for analysis. A listing of significant genes in each treatment group compared to the control chip was provided to Dr. Lamartinere. Additional analyses were conducted and are ongoing.

Project 2:

The **Biostatistics Core** has been working closely with Dr. Pam Horn-Ross in development of a database for the laboratory data (proteomics/micro-array) that can be linked with the demographic/dietary exposure data housed at NCCC. The standard nomenclature developed by the Bioinformatics Core will be adapted for use in this longitudinal study. We have worked with **Core B** to determine the format of the SNP data that will be forthcoming from this project (buccal swab analyses) and with project 3 to prepare for and evaluate methods for Structured Association Testing (SAT) that will be needed to control for potential confounding by admixture.

Due to the recognized multiple extra-biologic source of variation in conducting proteomic and microarray assays, the **Biostatistics Core** has worked with **Cores B** and **C** to develop experimental designs to allow for controlling this variability, as well as analyzing it. The **Biostatistics Core** has provided statistical support in experimental design. Specifically, we used randomization procedures to control variation among 2D gels. In addition, we designed a pilot experiment to access the variation of gene arrays and other sources of laboratory variation.

Bioinformatic Support Services

The CNGI website has been evaluated by Jelai Wang and several upgrades have been proposed to make it more user friendly and functional. Revisions have been made to the estrogen biosynthesis pathway tools (http://www.heflingenetics.uab.edu/ESP/pathway.html). With regard to standardized databases, **Core D** is recommending use of GEO

dissemination.

Project 1

Animals and chemoprevention
Coral A Lamartiniere, PhD, Leader
helenkim@uab.edu
Isom El-Toum, MD, investigator
sbarnes@uab.edu
Jun Wang, MD, investigator

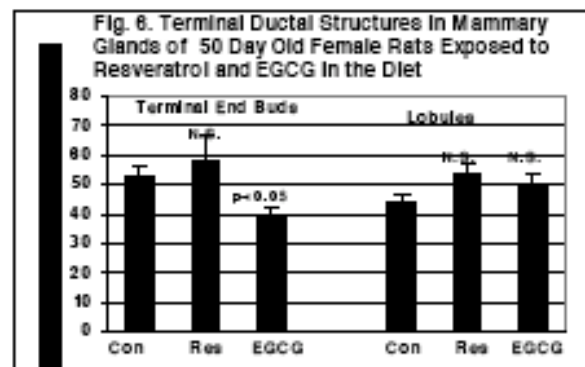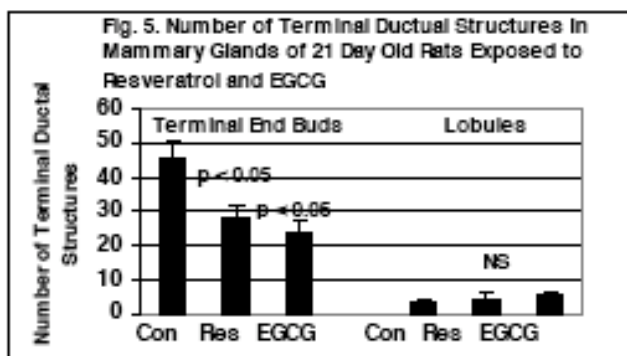
Project 1 SPECIFIC AIMS for –02 year:

- To examine the chemoprevention properties of resveratrol and EGCG administered in the diet of animals treated with the carcinogen DMBA
- To evaluate the effects of genistein, resveratrol and EGCG on mammary gland cell proliferation and differentiation at 21 and 50 days of age
- To provide mammary gland samples for DNA microarray analysis nd proteomics analysis by Cores B and C
- To design a chemoprevention experiment using a combination of polyphenols

---

Chemoprevention studies

Resveratrol, but not EGCG in the diet was able to significantly reduce the number of tumors per rat ($p < 0.05$). Resveratrol treatment resulted in significantly greater latency in tumor development over the controls with respect to time to first, second and third tumor detections. On the other hand, EGCG resulted in significantly earlier time to first tumor development. There was no significant difference to second and third tumor development with EGCG. Histopathology showed that every rat with mammary tumors had adenocarcinomas. We conclude that pure preparations of *resveratrol*, and not EGCG, suppress DMBA-induced mammary tumor development in rats.

significant difference in number of lobules (Figure 6). *Resveratrol* did not alter numbers of terminal end buds or lobules in mammary glands of 50-day old rats.

Analysis following 2 h of BrdU incorporation demonstrated that *resveratrol* in the diet resulted in reduced cell proliferation in terminal end buds, terminal ducts and lobules of mammary glands of 50-day old rats. There was no change in BrdU incorporation in mammary terminal ductal structures of 21 day old rats treated with *resveratrol* (data not shown). In contrast, EGCG reduced cell proliferation in terminal end buds of 21-day old rats only. Hence, there is good correlation between reduced cell proliferation at time of carcinogen exposure and reduced mammary tumorigenesis following *resveratrol*

Fig. 5. Number of Terminal Ductual Structures In Mammary Glands of 21 Day Old Rats Exposed to Resveratrol and EGCG

Fig. 6. Terminal Ductal Structures In Mammary Glands of 50 Day Old Female Rats Exposed to Resveratrol and EGCG In the Diet

treatment.

Mammary gland differentiation and cell proliferation

Analysis of whole mounts of mammary glands from 21-day old rats exposed from birth onward to these polyphenols showed that *resveratrol* and EGCG reduced the number of terminal end buds and caused no difference in the number of lobules (Figure 5). From mammary glands of 50-day old rats, we determined that EGCG resulted in fewer terminal end buds, but no

***Resveratrol* and EGCG serum concentrations** were measured from blood of 50 day old rats exposed to 1000 mg *resveratrol*/kg diet or 0.065% EGCG in drinking water since birth whereby they ingested approximately 75 mg *resveratrol*/kg BW or 68 mg EGCG/kg BW daily. For *resveratrol*- and control- treated rats, we found the

total *resveratrol* concentrations to be 2.00 ± 0.64 μM and 0.01 ± 0.01 μM, respectively. For EGCG- and control-treated rats, we found the total EGCG concentrations to be 9.71 ± 0.67 nM. There was no detectable EGCG in the control serum. We conclude that there is a large difference (186-fold) in availability of these two purified compounds to the blood in rats, suggesting reduced mammary tissue bioavailability of EGCG compared to *resveratrol* and a potential reason for the lack of chemoprotective effect of EGCG.

Genomics and proteomics analyses

Ten mammary glands from 21- and 50-day old rats from EGCG, *resveratrol, genistein* and *control* treatments (total of 80) were sent to the Genomics Core for analysis. RNA ware recovered from the samples and processed for analysis using high density Affymetrix RAE 230 version 2.0 microarrays. Genes have been identified that are differentially expressed by *genistein*, *resveratrol* and EGCG dietary treatments compared to *controls*. Similarly, samples (total of 80) were submitted to the Proteomic Core for analysis. The Proteomics Core has run the 2-D gels; however, there have been problems of performance of the gel imaging and analysis software. These issues are discussed in the Proteomics Core (**Core C**) section.

Design of Phase II tumorigenesis studies These are designed to investigate combinational lower doses of the bioactive polyphenols that show promise as mammary cancer chemoprevention agents. Since *genistein* and *resveratrol*, but not EGCG, were demonstrated to suppress DMBA-induced mammary cancer, we have chosen two lower doses of *genistein* and *resveratrol* with the intent to see additive or synergistic chemopreventive effects. The second phase of chemoprevention studies will include the following groups (30 rats each):

- Group 1) AIN-76A diet (Controls) + DMBA
- Group 2) 83 mg *Genistein* + DMBA
- Group 3) 333 mg *Resveratrol* per kg AIN-76A diet + DMBA
- Group 4) 83 mg *Genistein* + 333 mg *Resveratrol* per kg AIN-76A diet + DMBA
- Group 5) 25 mg *Genistein* + 333 mg *Resveratrol* per kg AIN-76A diet + DMBA
- Group 6) 83 mg *Genistein* + 100 mg *Resveratrol* per kg AIN-76A diet + DMBA

Project 2

GRLS Study
Pamela L. Horn-Ross, PhD, Leader
        pam@nccc.org Christine Collins, MS, Program Manager
        ccollins@nccc.org

Project 2 SPECIFIC AIMS for –02 year:

- To prepare dietary and other questionnaires and develop biospecimen collection and processing protocols
- To obtain IRB approval to begin recruitment
- To carry-out a pilot study
- Begin recruitment of the cohort

In –02 year, a pilot study was conducted with 11 girls to test all aspects of the procedures and to finalize changes to the study instruments before entering the field for the full study. IRB approval was obtained for all materials, allowing us to begin the full-scale study. In conjunction with **Cores B** and **C**, written protocols/procedures for the collection of biological specimens were developed and finalized. The pilot study suggested that obtaining blood samples may be difficult with only 40% of girls agreeing to participate in that component of the study.

.

Recruitment of the cohort began in late April 2005. As of May 25th, 2005, screening has been conducted for 31 girls. Twenty-eight are eligible and 16 have been interviewed. All 16 of these girls have provided buccal and urine specimens. Thirteen were willing to provide blood samples, a much higher proportion than suggested by the pilot study data. Blood collection was delayed for several reasons, but will begin shortly. Interviews with the other 12 eligible girls have been scheduled.

Project 3

High dimensional data analysis
David B. Allison, PhD, Leader
dallison@uab.edu
Grier Page, PhD, co-leader
gpage@uab.edu

Project 3 SPECIFIC AIMS for –02 year:

- To Develop a quality control metric based upon Geography
- To assess the measurement properties of Affymetrix microarrays.
- To Compare the performance of a variety of non-hierarchical cluster algorithms
- Develop methods for combing data from multiple sources
- Continue the development of software to analyze microarray and proteomic data

Progress in **project 3** has been highly productive. Several papers have been published and several are under review. Below is a description of some of the projects we have been working on. We have also being working in cluster stability, Bayesian analysis, quality control, and model building as well as review articles on microarrays and proteomics.

We have been developing software for microarray data analysis that implements several methods for processing and normalizing single and dual channel data. It is has variety of inferential statistical procedures including T-test, ANOVA, Chebby Checker, and Bootstrapping. In addition false discovery rate and family wise error rate methods have been implemented. In addition, we have implemented a variety of quality control procedures. Finally, we have implemented our method for the novel power analysis of microarray studies. This tool, entitled *HDBStat!*, is available at www.ssg.uab.edu and described in a recent publication.

We published a new method for power analysis in high dimensional biology such as microarray or proteomic research when many hypotheses are being tested simultaneously.

Progress has also occurred in the area of designed experiments. We designed an experiment that would yield high-quality information about the measurement characteristics of the microarray and proteomic techniques available in our laboratories. An experiment has been completed in which eight (8) mammary gland samples were separately labeled and spotted onto Affymetrix microarrays with labeling and technical replication (24 total arrays). A variety of gene chip image processing algorithms were using the same chip images. These algorithms themselves were found to give rise to different interpretations of the same raw images. In contrast, repeated labeling, either on the same day or on different days, created little to the overall variation. As expected, *biological differences* between the *outbred rats* animals used were the principal source of variation.

A publication is in revisions for BMC Bioinformatics. A similar paper is being developed for the cDNA arrays.

In a third development, a procedure was developed to test the quality of data from individual chips. This is important since the manufacturing process has not yet reached the point where it can be assumed that every chip will give rise to uniform data quality. The method is based on examination of the distribution of deleted residuals both within a given experiment and for every group in the experiment. This method was applied to 169 sets of available published data. Software (HDBStat!) is available from the Section on Statistical Genetics website (http://www.soph.uab.edu/ssg_content.asp?id=1164) for investigators to evaluate the quality of their microarray data. This is very important for analysis of microarray data since homogeneity of variance is essential for statistical interpretations of array and proteomics data. A publication on the method is being developed.

Fourthly, we have derived via both analytic solutions and computer simulations, the optimal allocation of costs when the number arrays that can be allocated to a measurement evaluation experiment is fixed to achieve the most efficient design. The question considered is whether it is better to run more subjects for a fewer number of replicates or more replicates for a fewer number of subjects when evaluating the reproducibility of a microarray experiment. *We showed that the optimally efficient design is one in which the largest number of cases are used with only two replicates per case*.

As high-density microarray technology is increasingly applied to study gene expression levels on a large scale, assessment of array's quality is an important task in development of microarray technology as well as for gene expression studies. Microarray experiments rely on several critical steps that may introduce error and uncertainty in the analyses. These steps include: mRNA sample extraction, amplification and labeling, hybridization, and scanning. In some cases, this may be

manifested as systematic spatial variation on the surface of the microarray in which the mean expression level within an individual microarray may vary as a function of geographic position on the array surface. We hypothesized that an index of the degree of spatiality of gene expression measurements associated with their physical geographic locations on arrays could serve as an indicator of the overall summary of reliability of the microarray. The objective of this study is to introduce a novel idea how to formulate this index using a regression analysis approach in which gene expression intensity measurements are regressed on a polynomial response surface of the microarray's Cartesian coordinates. We apply this method to real data in an effort to illustrate the method's utility.

The emphasis in the balance of –02 year will be the analysis of the experiments developed to determine evaluate the optimal analysis platform and the algorithms for analysis of microarray image data. Similar approaches will be applied to the optimum analysis of proteomics data. Experiments based on empirical biological data as well as using simulations will be conducted to examine the properties of emerging statistical procedures for high dimensional analysis both derived by the CNGI team and by other investigators. Work is also underway on methods for human disequilibrium and haplotype mapping and methods for dealing with admixture that will be useful to **project 2**.

Publications from the CNGI team in the –02 year

1. Barnes, S. and Allison, D.B. Excitement and realities in microarray analysis of the biological effects of polyphenols (2004). *Pharm Biol*, **42**(suppl): 94-101.
2. Barnes, S., Prasain, J.K. (2005) Current Progress in the use of Traditional Medicines and Nutraceuticals. *Current Opin. Plant Biol.*, **8**: 324-328.
3. Barnes, S., Allison, D.B., Page, G.P., Carpenter, D.M., Gadbury, G.L., Meleth, S., Horn_Ross, P., Kim, H. and Lamartiniere, C.A. (2005) Genistein and polyphenols in the study of cancer prevention: chemistry, biology, statistics and experimental design In: "*Discovering the Path to Personalized Nutrition*", J. Kaput, R. Rodriguez, Wiley Press.
4. Beasley, T., Page, G., Brand, J., Gadbury, G., Mountz, J. and Allison, D. (2004) Chebyshev's inequality for non-parametric testing with small n and a in microarray research. *Applied Statistics*, **53**, 95-108.
5. Deshane, J., Chaves, L., Sarilonda, K., Isbell, S., Wilson, L., Kirk, M., Clinton, G., Barnes, S., Meleth, S. and Kim, H. (2005) Proteomics Analysis of Rat Brain Protein Modulations by Grape Seed Extract. *J. Agric. Food Chem.* **52**: 7872-7883.
6. Gadbury, G., Page, G., Edwards, J., Weindruch, R., Permana, P.A., Mountz, J. and Allison, D.B. (2004) Power Analysis and Sample Size Estimation in the Age of High Dimensional Biology: A Parametric Bootstrap Approach and Examples from Microarray Research. *Stat.Meth.Med.Res*, **13**, 1.
7. Garge, N., Page, G., Sprague, A., Gorman, B. and Allison, D. (2005) Reproducible Clusters from Microarray Research: Whither? *BMC Bioinformatics*, In press.
8. Kim, K, Zakharkin, S.O., Loraine, A. and Allison, D. (2004) Picking the most likely candidates for further development: Novel intersection-union tests for addressing multi-component hypotheses in comparative genomics. *Nutrition* **20**:1.
9. Kim, H., Hall, P., Smith, M., Kirk, M., Prasain, J., Barnes, S. and Grubbs, C. (2004) Chemoprevention by Grape Seed Extract and Genistein in Carcinogen-induced Mammary Cancer in Rats Is Diet Dependent. *J. Nutr.* **134**: 3445S-3452S.
10. Prasain, J.K, Wang, C.C, Barnes, S. (2004) Mass spectrometric methods for the determination of flavonoids in biological samples. Free Rad Biol Med **37**: 1324-1350.
11. Ruden, D.M., Cui, X., Loraine, A., Ye., J., Bynum, K., Kim, N.C., De Luca, M., Garfinkel, M.D. and X. Lu (2005) Methods for nutrigenomics and longevity studies in *Drosophila*: Effects of diets high in sucrose, palmitic acid, soy or beef. In: "Methods in Nutrigenomics," edited by T.O. Tollefsbol. Methods in Molecular Biology series, edited by J. Walker. Humana Press, Totowa, NJ.
12. Ruden, D.M., Garfinkel, M.D., Liao, X. and X. Lu (2005) Epigenetic regulation of trinucleotide repeat expansions and contractions and the "biased embryo" hypothesis for rapid morphological evolution. *Current Genomics* **6**, in press.
13. Ruden, D.M., De Luca, M., Garfinkel, M.D., Bynum, K. and X. Lu (2005) *Drosophila* nutrigenomics can provide clues to human gene-nutrient interactions. *Annual Review of Nutrition* **25**, in press.
14. Ruden, D.M., Liao, X., Garfinkel, M.D. and X. Lu (2005) Hsp90 and environmental impacts on epigenetic states: a model for the trans-generational effects of diethystilbesterol on uterine development and cancer. *Hum. Mol. Genet.* **14**: R149-R155.
15. Trivedi, P., Edwards, J.W., Wang, J., Gadbury, G.L., Srinivasasainagendra,V., Zakharkin, S.O., Kim, K., Mehta, T., Brand, J.P., Patki, A. *et al.* (2005) HDBStat!: A platform-independent software suite for statistical analysis of high dimensional biology data. *BMC Bioinformatics.*, **6**, 86.
16. Yang, D., Zakharkin, S.O., Page, G.P., Brand, J.P., Edwards, J.W., Bartolucci, A.A. and Allison, D.B. (2004) Applications of bayesian statistical methods in microarray data analysis. *Am. J. Pharmacogenomics.*, **4**, 53-62.