I.I. Rabi Scholars Program
FOURTH ANNUAL RESEARCH SYMPOSIUM
12 noon to 2 p.m., Friday, October 2, 2009
Satow Room, Lerner Hall (5th floor)

ABSTRACTS

ALLOCCO, Elizabeth

Mentor: Jennifer Manly (Sergievsky Center, Mailman School of Public Health)

Latent Class Modeling of Cognitive Change in Older Adults

Research on patterns of cognitive change among older adults is highly influenced by the strategy used to recruit the cohort and the domain and psychometric properties of the instruments used to measure cognition. Statistical models used in the analysis of the data collected may obscure patterns associated with normal aging. The current study used a community-based random sample of ethnically diverse older adults and a sensitive instrument of acquisition of verbal information collected at several time points. Individuals were enrolled in the study in 1999, at which time initial anthropometric, socio-demographic, and neuropsychological data were collected. Participants were then reassessed at two subsequent time points, at intervals of 24 months. The total number of words recalled on the six learning trials of the Buschke Selective Reminding test was used as the variable of interest when creating an unconditioned growth model using MPLUS. The mean recall at the initial visit was 35.57 words, with a mean decline of 0.705 words over the two follow-up visits. There was significant variability in this rate of change, suggesting that there may be multiple latent classes of memory decline. Preliminary analysis using a growth mixture model revealed two such groups representing older adults with 1) high baseline recall and stability over time and 2) low baseline recall and more rapid decline over time. The first class consisted of 421 individuals (20% of the total cohort) with a mean baseline recall of 46.97 words and decline at a rate of 0.21 words per visit. The second group consisted of 1708 individuals (80% of the total cohort) whose mean recall at baseline was 33 words and who had a decline of 0.83 words per visit. The existence of these two latent classes suggests that there are different patterns of cognitive decline which may be mediated by non-cognitive factors such as education or socioeconomic status, chronic medical illness, depression, or social support.

ATANASOV, Atanas

Mentors: Henry Adams and Gunnar Carlsson (Stanford University)

Applying the Nudged Elastic Band Method to Topological Data Analysis

There has recently been a surge to employ geometric and topological methods in various applied contexts, and in particular, data analysis. Numerical data can often be presented as a subset X of a finite dimensional Euclidean space R^n, where each point of X corresponds to a single sample, and the individual components stand for collected measurements or inputs. One of the objectives of data analysis is to understand the relations between these separate inputs. It is even more valuable if we can find a subspace $Y \subset R^n$ that parametrizes all possible inputs. Instead of requiring the complete topological

information of Y, one can ask for an invariant which is computable solely from a sample X. For example, persistent homology has been employed to compute $H_*(Y)$ [2, 11, 4]. Another example is given by the mapper algorithm [10] which produces a graph approximating the 1-skeleton Y^1 (provided we assume Y is a CW complex). The main purpose of this exposition is to introduce an alternative approach to recovering skeleta Y^k inspired by the nudged elastic band (NEB) method pertaining to the field of computational chemistry [8, 9, 6]. We have successfully applied these techniques to obtain Y^1 for several instances. None of the datasets we tested exhibited strongly expressed cells in dimensions 2 and above. However, many of the presented techniques generalize to higher dimensions.

The classical NEB method, even with modifications, is only applicable to finding Y^1. We have significantly extended the algorithms so they can handle mesh-like structures, which could represent cells or arbitrary dimensions. Beyond the handling of a single band or mesh, we have formulated concrete procedures that work together to find the higher dimensional skeleta. These methods were applied successfully to two datasets of natural image patches [5, 7], and the known models [3, 1] were recovered.

[1] Henry Adams and Gunnar Carlsson, On the nonlinear statistics of range image patches, SIAM Journal on Imaging Sciences 2 (2009), no. 1, 110–117.

[2] Gunnar Carlsson, Topology and data, Bull. Amer. Math. Soc. (N.S.) 46 (2009), no. 2, 255–308. MR MR2476414

[3] Gunnar Carlsson, Tigran Ishkhanov, Vin Silva, and Afra Zomorodian, On the local behavior of spaces of natural images, Int. J. Comput. Vision 76 (2008), no. 1, 1–12.

[4] Gunnar Carlsson and Afra Zomorodian, The theory of multidimensional persistence, Discrete Comput. Geom. 42 (2009), no. 1, 71–93. MR MR2506738

[5] J. H. Van Hateren and A. Van Der Schaaf, Independent component filters of natural images compared with simple cells in primary visual cortex, 1998.

[6] Graeme Henkelman and Hannes Jonsson, Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points, Journal of Chemical Physics 113 (2000), no. 22, 9978–9985.

[7] Jinggang Huang, Ann B. Lee, and David Mumford, Statistics of range images, CVPR, 2000, 324–331.

[8] Aiichiro Nakano, A space-time-ensemble parallel nudged elastic band algorithm for molecular kinetics simulation, Computer Physics Communications 178 (2008), no. 4, 280 – 289.

[9] Daniel Sheppard, Rye Terrell, and Graeme Henkelman, Optimization methods for finding minimum energy paths, The Journal of Chemical Physics 128 (2008), no. 13, 134106.

[10] Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson, Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition, pp. 91–100.

[11] Afra Zomorodian and Gunnar Carlsson, Computing persistent homology, Discrete Comput. Geom. 33 (2005), no. 2, 249–274. MR MR2121296 (2005j:55004

BECK, Samuel

Mentor: Norman Christ

Finding Eigenvalues of the Dirac Operator with the Lanczos Algorithm

In lattice qcd, a five-dimensional lattice is often used to simulate processes that occur in four-dimensional space. It is widely believed, however, that for low energies, particles in lattice qcd will exhibit four-dimensional behavior. If this is the case, then the lowest eigenvalues of the Dirac operator should have a cubic distribution. Unfortunately, previous methods have only returned a few hundred eigenvalues, hardly enough to see any real distribution. The Lanczos algorithm transforms the Dirac operator into a tridiagonal matrix, meaning that it is zero everywhere except the diagonal and the positions immediately adjacent to the diagonal. Moreover, the Lanczos algorithm is an iterative process, and the eigenvalues of the matrices it generates have been observed to approach the eigenvalues of the original matrix, long before they get to the same size. Using this algorithm, we've so far been able to generate a tridiagonal matrix of dimension more than 800,000 by 800,000, and used another algorithm known as Sturm sequencing to produce the eigenvalues of this matrix. We do not yet have a distribution, since many of the eigenvalues found do not represent eigenvalues of the original matrix, and identifying the spurious eigenvalues is proving problematic.

BELOPOLSKI, Ilya

Mentor: Szabolcs Márka

An Upper Limit on the Effective Rate of Detection of the Gravitational Wave Signatures of Core-Collapse Supernovae

Gravitational waves (GW) promise to provide a new window on the universe by allowing scientists to observe astrophysical phenomena without dependence on light. However, although the existence of gravitational waves has been predicted for nearly one hundred years, these waves have not yet been directly detected because they interact so weakly with matter. Indeed, despite substantial improvements in sensitivity during recent years, detector systems such as the Laser Interferometer Gravitational-Wave Observatory (LIGO) in Louisiana and Washington still output signals corresponding to a mixture of environmental, instrumental, and fundamental noise. One current strategy to detect gravitational waves is to focus search efforts on the GW signatures of core-collapse supernovae. Core-collapse supernovae occur as a result of the internal collapse and explosion of massive stars. Advanced models of gravitational collapse have produced many estimates of the GW emissions produced by such an event. As a result, many efforts are currently underway to use the expected GW signatures of supernovae to serve as a model in the detection of gravitational waves. In particular, improvements to LIGO sensitivity and data analysis can increase the likelihood that GW emissions from such events are detected. However, although supernovae occur with high frequency throughout the universe, models of core-collapse supernovae are not yet able to provide estimates of LIGO detection for the majority of these events. Here we evaluate the restrictions that LIGO sensitivity and current models of supernova waveforms place on the ability of LIGO to detect GW emissions of core-collapse supernovae and the usefulness of directing LIGO efforts towards detection of these events.

We use previously published models of supernova gravitational wave emissions that were produced using computer simulations of core-collapse supernovae. We compare these models to

the current LIGO sensitivity spectrum and possible sensitivity spectra for the Advanced LIGO detector. Our results for these modeled signals suggest that supernova GW emissions within the Milky Way galaxy would be easily visible to the LIGO detector network. However, the rate of such an event is too low to be useful on the timescale for which the Advanced LIGO detector network will be active (approximately ten years). To have a high chance for a successful detection of a supernova GW event, it will be necessary for LIGO to have visibility into our neighboring Andromeda galaxy, which is much larger and where supernovae are more likely to occur in LIGO's lifetime.

To produce a more accurate estimate of LIGO's sensitivity to these supernovae waveforms, we implemented a modified version of the Flare data analysis pipeline. Previously this pipeline was used to search for GW transients in LIGO data segments associated with electromagnetic trigger events picked up by other telescopes. We used this pipeline to inject our simulated waveforms with various strain amplitudes into simulated LIGO noise. For each amplitude injection we determined detection efficiencies for the hypothetical incident GW. By establishing an arbitrary threshold for detection, we were able to calculate a minimum incident strain required for a given waveform. Using existing estimates for GW energy output in supernovae and our minimum strain requirements, we were able to calculate LIGO's maximum visibility for various theoretical supernovae GW emissions.

Early results produced using the Flare pipeline indicate that Advanced LIGO could in principle detect supernova GW emissions from massive asymmetric supernovae in the Andromeda galaxy. However, more refined analysis is needed to understand the difficulties and uncertainties associated with detecting supernova GW either in our own galaxy or in the Andromeda galaxy. Currently, we are continuing to refine our Flare search method to produce more reliable results. Furthermore, questions remain regarding the validity of computer-generated GW waveforms. Successful detection largely depends on the high probability of supernovae events, which are asymmetric. Although computer models attempt to simulate asymmetric core-collapse supernovae, it is unclear how often such asymmetric supernovae occur. Thus, a more robust model may be needed to better understand the difficulties associated with LIGO detection of GW emissions of supernovae.

BREGMAN, Corey

Mentor: Aaron Lauda

Categorification of the Adjoint Representation of sl(n,C)

In recent decades, Category theory has provided a concrete way of clarifying the basic objects of mathematical inquiry and the interrelations of different branches of mathematics. The process of categorification, or replacing set-theoretic notions in various theories with category-theoretic ones has allowed for a richer structure and led to the construction of knot invariants. Fairly recently, it was conjectured that categorification of representations of Hopf algebras will provide a way to construct 4-dimensional topological quantum field theories. One main source of Hopf algebras are quantum groups, or Lie algebras that have been deformed by a continuous parameter q subject to certain relations. In order to categorify representations of these Hopf algebras we proceed in the following way: given a graph G and a vertex set I we construct the Khovanov-Lauda Ring R associated to G by considering formal sums of braid-like diagrams with dots up to planar isotopy. The strands of each braid are indexed by vertices of I, and multiplication is given

by concatenation of braids. It is known from the literature that cyclotomic quotients of R categorify representations of the Lie algebra sln(C) for G the corresponding Dynkin diagram, under the Grothendieck group functor K0. In this case, the number of dots specified in the quotient corresponds to the highest weight of the representation, sequences of vertices correspond to weight spaces, and the induction and restriction functors in R-Mod provide the raising and lowering operators. We prove this categorification explicitly for the adjoint representation of sln(C), thereby finding a basis for the quotient of R for each n. Moreover, using the graded-dimension of Brundan and Kleschev, we compute a closed-form expression for the graded dimension of this quotient.

DHULDHOYA, Jay

Mentors: James Sung and Lily Wu (David Geffen School of Medicine, UCLA)

Constructing Lentiviral Vectors to Study the Role of CCR7 in Prostate Cancer Metastasis

Prostate cancer mortality is commonly the result of metastatic dissemination of the primary tumor to distant sites. Recent evidence suggests that the lymphatic system may be an important conduit for the spread of tumor cells to regional and distant lymph nodes. Furthermore, we have shown that overexpression of the lymphangiogenic factor VEGF-C increases the incidence of prostate cancer metastasis to drainage lymph nodes. Analysis of these lymph node metastases showed upregulation of the chemokine receptor CCR7, implicating its role in the metastatic process. CCR7 and its ligands, CCL19 and CCL21, normally function to traffic immune cells through the lymphatic system. Aberrant expression of these chemokine receptors on tumor cells may allow them to utilize this chemotactic system to traffic to distant sites. In order to study the effects of CCR7 overexpression on lymph node metastasis, we constructed a lentiviral CCR7-Red Fusion vector. First, the red Tomato and CCR7 genes were PCR amplified and inserted into separate pCR2.1 vectors. Then the Tomato gene was restriction digested from pCR2.1 and inserted into pcDNA3.1 using engineered XbaI and ApaI restriction sites. Finally, the CCR7 gene was inserted into pcDNA3.1 with Tomato using NheI and HindIII restriction sites, which ensured that the two proteins were in frame. To test for functionality, we performed an *in vitro* assay in which the CCR7-Tomato fusion construct and a B-arrestin-GFP construct were cotransfected in 293T cells. Activated CCR7 recruits B-arrestin, which conveys the downstream signaling from CCR7 to the nucleus. The cells were stimulated with the ligand CCL21 and fixed at 0, 5, 10, and 20 minute time points. Primary results using confocal microscopy indicated colocalization of the CCR7 and B-arrestin, which suggests that the CCR7 fusion protein is indeed functional. Finally, the construct was cloned into the lentivector pccl-M4, which can be used in the future to study CCR7 overexpression on different prostate cancer cell lines both *in vitro* and *in vivo*.

GEORGESCU, Alexandru

Mentor: Andrei Beloborodov

Modeling the Magnetic Field Lines of a Neutron Star

After most of the hydrogen within a star has been consumed, and there isn't enough fuel to sustain the star in its initial form, it "dies" - with new characteristics that depend on its original size. A star about the size of our Sun turns into a white dwarf. Stars ~3-15 times larger than our own become neutron stars. Still more massive stars turn in to black holes.

My summer project involved determining the distribution of the magnetic field of a pulsar (a very active neutron star, with one of the strongest magnetic fields in the known universe). Close to the star, plasma is distributed along the magnetic field lines, and cannot escape the star's magnetic field. However, at the points where the angular frequency times the radius to the rotation axis equals the speed of light, the speed of the plasma is equal to the speed of light. The points where this is true form a surface called the light cylinder. We thus have to solve the PDE outside and inside the light cylinder, and after a long iteration process that changes the current distribution along field lines by more than 10^4 times, we find a solution that works. This method was first used by Contonopoulos, Kazanas and Fendt in 1999. We illustrate the above with a similar solution. The next step in the research will be to solve the same problem for the case in which "star quakes" emerge. i.e. The magnetic field is so strong that the crust of the neutron star is broken, resulting in the twisting of the magnetic field lines.

This is also an interesting problem from the point of view of classical electrodynamics. Though the theory predicts that there are no magnetic monopoles in nature, the results we've obtained predict that – though dipolar close to the star - the magnetic field lines becomes radial at a few times the radius of the light cylinder.

HANCOCK, Stephen

Mentor: John Parsons

Simulation of Detector Function at the Large Hadron Collider

When two protons collide in a particle accelerator, the energy produced can be converted into new particles via the relation E = mc2. In particular, if the collision is energetic enough, a top/anti-top (tt-bar) quark pair may be created. The short lifetime of these species, however, results in the immediate decay of each into a bottom quark and W boson. The latter then decays into two lighter quarks or two leptons. As the product quarks separate, it soon becomes more energetically favorable for a new quark/anti-quark pair to pop out of the vacuum than to allow a further separation, a phenomenon known as color confinement. Color is a property like charge unique to quarks and gluons that must be zero for any observable particle. The energetic quarks continue this so-called hadronization process such that the detector ends up seeing "jets" of numerous colorless particles coming from the general direction of the initial decay products.

We focused on semi-leptonic events, in which one W decays leptonically and the other hadronically. Thus we expect four jets corresponding to the four quark products (a bottom from each top and two lighter quarks from one of the W's). The problem is that we often see more than four, the additional ones being a result of the extra energy in the event or of a jet "splitting in two" before the detector due to a quark radiating a gluon at large angle. The momenta of these jets are reconstructed from the detected partons using a metric known as anti-KT that takes into account angular distance and energy. The most naïve way to reconstruct the tt-bar mass is simply to add the four momenta of the four jets with highest transverse momentum (pT). However, I found by investigating the truth information of the simulated events that these are not always the correct jets. Moreover, one must sometimes use more than four if the hadronization of one or more quarks was split into two directions by a hard radiation.

Many histograms were produced of the properties of the jets involved in these somewhat anomalous events. I am still searching for a satisfactory combination of observables that would suggest, for instance, that a fifth jet should be added or that a particular high-pT jet is actually

not a product of the tt-bar decay. I hope to develop an effective algorithm that picks out the correct jets to a reasonable degree of accuracy. This research is important because when the Large Hadron Collider becomes operational at CERN, most of what the ATLAS and CMS detectors see will be jets. Reliable reconstruction methods will be needed in order to discover and to determine the properties of new heavy particles like the Higgs boson.

HORLBECK, Max

Mentors: Alenka Copic and Liz Miller

The Role of Sec13p in COPII Vesicle Formation

The Endoplasmic Reticulum (ER) processes newly synthesized proteins within the cell. Proteins are packaged into Coat Protein 2 (COPII) vesicles and transported to cellular destinations. COPII vesicles are formed by the coordinated assembly of coat proteins, culminating with an outer cage of Sec31p and Sec13p. This process is highly conserved in organisms as diverse as humans and budding yeast, our model organism. Each protein is essential for vesicle formation and cell survival. However, if the gene encoding for Sec13p is deleted along with one of a number of genes, the cells remain viable. This indicates that vesicle formation can occur without Sec13p in the absence of certain proteins, named Bypass of Sec Thirteen (BST) proteins. In order to discover more such genes and identify the properties of the BST phenotype, we conducted a Synthetic Genetic Array (SGA) screen, which tests every gene in the yeast genome for the phenotype. In addition, we created mutant versions of Sec31p that are intended to abolish interaction with Sec13p. The mutants are inviable except when BST genes are deleted, demonstrating that the loss of Sec13p presence in the COPII vesicle is responsible for sec13Î" lethality and rescue. We are also conducting the SGA screen with these mutants to verify hits, and performing in vitro budding assays to test the properties of the ER membrane in BST background cells. The ER is implicated in a number of human diseases, including cystic fibrosis.

KANG, Joon Ho

Mentor: Abhay Pasupathy

Learning How to Manufacture Super Thin Graphite

Scanning Tunneling Microscope (STM), first invented by Gerd Binnig and Heinrich Rohrer in 1981, is a powerful tool for viewing conducting surfaces at the atomic level. In this tool, a metallic tip is brought to a few atomic distances from the conducting surface, allowing electrons to tunnel between the tip and the surface. Apart from viewing the surfaces at the microscopic level, STM can measure the atomic scale conductance of a sample, and it is thus a powerful tool for materials research. In my research, we constructed a complete STM apparatus. This included a piezoelectric mechanism to bring a prepared sample a few atomic distances away from the tip and a spring suspension system to reduce vibrational noises of the sample and the tip. The STM microscope was housed in an Ultra High Vacuum System with cryogenic cooling. To test the precision of the STM microscope, prepared copper and gold were imaged in our STM system. After installation, the performance of the STM was improved to observe atomic steps on a copper surface and picometer scale structures on a gold surface. I am currently using the STM to probe new materials such as one sheet of graphite, which forces electrons to move in two dimensions.

KENNELLY, Michael

Mentor: John Parsons

Graviton Analysis in the Dielectron Channel at the ATLAS Experiment

The ATLAS Experiment at the Large Hadron Collider is a multi-purpose particle detector that will be capable of detecting a wide range of decay states from yet to be discovered theoretical particles. One such particle is a 300 GeV Randall-Sundrum Graviton, which could decay to an electron-positron pair. The existence of such Gravitons is predicted in many models, which postulate the existence of additional spacetime dimensions. Using Monte-Carlo data simulation, the acceptance and efficiency of ATLAS to detect these Graviton to ee events was measured. These methods allow us to identify signatures in the ATLAS detector that would result from decays of a Randall-Sundrum Gravitons, and are the foundation for how a Graviton could be detected once the LHC starts producing data. By comparing the shape of the measured ee invariant mass spectrum, which can be calibrated using the peak from decays of Z-bosons to dielectrons, we showed that this method would allow discovery of such a Graviton should it indeed exist.

LAZAR, André

Mentors: Harmen Bussemaker and Todd Riley

Profile Hidden Markov Models Show Significant Improvement over Position Specific Scoring Matrix Techniques in Breast Cancer Prognosis Prediction

The Transcription Factor Database developed by Badis et al. and the Breast Cancer datasets constructed by van Vliet et al. were used in the algorithmic association of transcription factor concentrations to breast cancer prognoses. Based on Spearman Correlation and Pearson parametrics, the intersection of 72 common probes was identified, concentrated in three enriched pathways pertaining to the distinct biochemical pathways of the cell cycle, DNA replication, and DNA proliferation. Previous work using the Jaspar database of transcription factor-associated binding sites had been done, from which Position Specific Scoring Matrices (PSAMs) were constructed. From these data, Profile Hidden Markov Models (PHMMs) were inferred, and a statistical advantage derived from the available data. These PHMMs were used in conjunction with inferred Transcription Factor levels found using the Bussemaker lab MatrixREDUCE algorithm AffinityProfiler, and by using the mRNA expression data collected from the Breast Cancer datasets, a classifier that associated transcription factor concentrations with breast cancer prognosis was developed that had statistical improvement over the previous version derived from PSSMs. Further work on creating a more dynamic algorithm that allows for variations in the DNA caused by random bulk mutations (such as large inserts known as “spacers”) is in progress, and promises to produce an even more accurate classifier.

LOPEZ, Chris

Collaborators: Alex Blumenthal, Karol Koziol, Alexander Moll, Christopher Scaduto and Warren Tai

Mentor: Yakov Kerzhner

Sphere Packing in Odd Dimensional Hyperbolic Space

The aim of sphere packing in a *n*- dimensional space is to fill the space with as many *n*-dimensional non-overlapping spheres of a given radius as possible; an arrangement of these spheres is called a *sphere packing*. More precisely, we want to maximize the *density* of such a packing, which can be determined by looking at the proportion of space taken up by the packing in a sphere of radius *r* and letting *r* go to infinity. Hyperbolic space is locally curved like a saddle, unlike Euclidean space, which is locally flat. Consequently, it has metric properties that are different from that of Euclidean space. For instance, whereas in Euclidean space the maximal density only depends on the dimension of the space, in hyperbolic space the maximal density depends on the radius of the spheres in the packing as well as the dimension of the space.

We sought to compute upper bounds on the maximal density for a ball of a given radius in hyperbolic space of a given dimension. We were able to obtain results for dimension 3 (dimension 1 is trivial). The upper bounds obtained appear to increase with an increase in the sphere radius. Non-trivial (less than 1) upper bounds were obtained only for small radii. For radius 1, the upper bound is 0.82. For radius 5, an upper bound of 0.995 can be extrapolated from our data. We seek to refine the bounds for dimension 3 and obtain bounds for dimension 5.

Sphere-packing was originally considered as people looked to stack spheres against each other, whether they were oranges or cannonballs, and is still useful in this respect. More recently, sphere packing has been applied to error-correcting codes. These correspond to sphere packings, in which denser packings represent codes that are more efficient at detecting and correcting errors.

MIHAILESCU, Ion G.

Mentors: Michael Endres and Norman Christ

The Mass Spectrum of N=1 Supersymmetric Yang-Mills Theory with Domain Wall Fermions

Supersymmetry is the most popular theory beyond the Standard Model of particle physics with the best chance of being confirmed (or confirmed to some degree) at the Large Hadron Collider at CERN. At a theoretical level, supersymmetric quantum chromodynamics is far from being well understood. As a result of the peculiarities of QCD (quantum chromodynamics), such a theory can be tested with respect numerical predictions only through a small number of techniques. Among those is Lattice QCD. In Lattice QCD, one attempts to simulate the physical theory on a discrete set of space-time points (the lattice). However, in discrete space-time many of the symmetries from the continuous space are broken. When the continuous limit is taken, these symmetries and the physical theory must be recovered. Otherwise, such a method would not properly describe the physical world. This requires the tuning of certain parameters in order to obtain the correct results. An essential element of the project in which I am involved is to tune the parameters for the N=1 supersymmetric theory, which we simulate.

MOY, Man-Yu

Mentors: Andrea R. Tan and Clark T. Hung

Optimization of Cloning-Based Approach to Increasing Collagen II Levels in Engineered Cartilage Constructs

An important criterion of engineered cartilage tissue for medical purposes is that the physical properties are similar to those of native cartilage tissue. One related challenge is the consistently low levels of collagen II in constructs. Collagen II is an important protein that provides tensile strength to cartilage. One approach to this problem is through genetic means. It is possible that overexpression of the collagen II gene in cartilage constructs could enhance collagen levels in the engineered tissue and thereby improve physical properties as well. Unfortunately, the process of transfection, or inserting genetic material into cells, is not a simple one. Optimization of this process must first be undertaken in order successfully to introduce genetic material of choice into the chondrocytes (cartilage cells) and to yield the desired phenotypic traits. Also, tests and precautions must be made to ensure that the process of transfection does not ultimately affect other properties of chondrocytes and cartilage tissue negatively.

To test the viability of this transfection approach, a circular DNA vector coding for GFP (Green Fluorescent Protein) was used with nucleofection (a method of transfection where an electric current is passed through cells to form holes in the membrane, thus promoting DNA uptake). The resulting survival of cells and transfection efficiency could thus be monitored over a series of days using fluorescent microscopy. Unfortunately, this method of transfection yielded rather low efficiencies both with and without poloxamer, a reagent added to help seal cell membranes after the nucleofection. Hence, more studies must be taken to find better optimized methods of transfection.

PATEL, Milesh

Mentors: Wan Yang and Brent Stockwell

Finding Novel Cancer Genes and Characterizing PSL1, a Novel Selectively Lethal Compound

Targeted cancer therapy interferes with specific molecules that are needed for carcinogenesis. It is a promising approach to cancer treatment because it is less harmful to healthy cells. However, some oncoproteins and most tumor suppressors are challenging targets because they are mutated or do not even exist in cancer cells. For these challenging targets, we can search for molecules along their signaling pathways as alternative cancer targets. Signaling networks are very complicated. Exploiting the concept of synthetic lethality therefore might be an efficient way to find pathway-targeting drugs systematically.

Erastin, a compound found to kill tumor cells selectively, was found using this concept. Erastin causes an MEK and iron dependent non-apoptotic oxidative cell death by perturbing VDAC2/3 function. Erastin-like molecules allow for discovery of additional cancer genes. To implement this idea, we infected cells with lentivirus containing shRNA expression cassette to knockdown respective target genes. Cells were then treated with Erastin to induce cell death. Any surviving cells from this condition may express shRNAs targeting critical genes involved in erastin-induced cell death. It is likely that these critical genes include oncogenes or components of oncogenic-RAS-signaling because Erastin has selective lethality against oncogenic-RAS. Our

efforts yielded 19 confident hits, seven of which have no known implication in cancer: AKAP10, DGKQ, GSK3A, GTF2H1, ITK, PACSIN3 and PHKG2.

shRNA is prone to off-target binding. Therefore, it is important to ensure that the correct gene was knocked down. shRNA knockdown efficiency of the seven hits was measured using the Western blot and Q-PCR. Only PACSIN3, PHKG2, and AKAP10 showed significant knockdown. To examine the role of the seven genes in transformation process, colony forming assays were carried out with HT1080 cells expressing shRNAs targeting these genes. After three weeks of incubation, shGSK3A, shGTF2H1, and shPHKG2 showed significant inhibition of colony formation. But since the knockdown of GSK3A and GTF2H1 could not be verified, we cannot be sure of their implication in cancer. PHKG2 seems like an interesting gene to further study.

A second part of this study dealt with characterizing a novel lethal compound with selectivity against PTEN-deficient breast cancer cells. Deletions or mutations in the PTEN gene are well known for their tumor-promoting roles in human cancer. A screen of 384,000 compounds showed that PSL1 selectively kills MCF10A cells with a PTEN knockout. We started characterizing PSL1 by performing counter screening with bioactive molecules and then verified hits from the counter screen using the Western blot to detect the activation of certain proteins. The results suggested that PSL1's lethality is JNK, caspase, and RIP3 dependent.

PERRY, Alex

Collaborators: Evan Dummit and Adam Goldberg

Mentor: Ken Ono (University of Wisconsin)

Conjectures on Sums of Twisted Kloosterman Sums

The idea for our work comes from a paper by number theorist Ron Evans, in which he conjectures that certain sums of Kloosterman sums equal the coefficients of certain modular forms. A Kloosterman sum K(x) is a special kind of exponential sum over a finite field, with a parameter x in the field; we considered the twisted sum of a Kloosterman sum, i.e. the sum over x of K(x) times the value of a character at x. Modular forms are basically holomorphic functions on the upper half-plane that transform well with respect to a subgroup of SL(2,Z); as it turns out they have very deep arithmetic properties. We proved one of Evans's conjectures relating these objects – Kloosterman sums and modular forms – and have submitted it to a journal, and we have almost proved a second, harder conjecture. The proof of the first conjecture relies on the fact that there is a certain 3-dimensional complex variety whose points are parametrized by the modular form in question, and the fact that the coefficients of this modular form can be described by Gaussian hypergeometric functions (finite field analogues of hypergeometric functions). The proof of the second relies on interpreting a trace formula on a space of modular forms arithmetically. In the big picture of things, our work concerns the relationship between four objects – varieties, modular forms, Kloosterman sums, and Gaussian hypergeometric functions – which, a priori, have no reason to be related. This is a guiding philosophy in modern number theory: certain geometric objects (varieties) carry arithmetic information, and special functions (e.g., modular forms, Kloosterman sums) count points on them.

PRZYTYCKI, Pawal

Mentors: Anil Raj and Chris Wiggins

Using Protein-Protein Interaction Data to Increase the Significance of Genome Wide Association Studies

A variation in a single nucleotide between members of a species is called a single-nucleotide polymorphism (SNP). In a genome-wide association study (GWAS), thousands of SNPs across the entire genomes of many individuals, who are categorized by observable traits such as race or a disease, are determined. It is hoped that genetic associations can be made with these traits. However, because the entire genome is being examined for indicators, many statistical tests are being performed at the same time, and the problem of multiple hypothesis testing arises. The goal of this project is combine known protein-protein interaction (PPI) information with GWA data in order gain more statistically significant associations. PPI networks have been shown to be modular in nature, meaning that proteins can be grouped into meaningful clusters. Furthermore, mutual information (MI), which is a measure of the dependence of two variables, can be used to quantify the similarity between two SNPs. In order to apply the PPI information to SNPs, each SNP was assigned to a protein based on its respective location in the human genome. This allowed us to create a SNP-SNP interaction network. MI was calculated between a large sample of random SNPs as well as between SNPs that shared edges in the network, with the hope that edge-sharing SNPs would have a higher average MI. Unfortunately, calculations so far have not given results that show, with any significance, that this is the case. With a more powerful computer, a larger selection of SNPs can be tested, and larger scale clustering can be conducted. This may provide better results.

RUBENSTEIN, Mitchell

Mentor: Clifford Stein

Energy Optimization through the Scheduling of Tasks on Multiple Processors

Since the early 1970s, microprocessor power densities have doubled every three years – a growth rate of power consumption that has dwarfed that of battery capacity. This hardware disparity between the growth rates of power consumption and power production technologies poses a demand for algorithmic solutions to the problem of minimizing energy use in the performance of computational tasks.

Seeking an algorithmic solution, we devise an algorithm that intakes a set of processors with given speed limits; intakes a set of tasks with a given start-time, deadline, and workload; and distributes this set of tasks to this set of processors, controlling the speeds at which these processors run as they perform tasks in a way that satisfies the following three conditions: 1) keeps the processor speeds below their speed limits; 2) completes as much task work and/or as many tasks as possible, and/or approximates optimality with respect to these goals; and 3) minimizes the amount of energy the processors expend in executing this optimal schedule of tasks.

The algorithm, which has been programmed and simulated, operates by storing processors in a min-heap that orders the processors by current processing speed, by storing unprocessed tasks in increasing order of deadline, and by repeatedly adding unselected tasks to the processor min-

heap until all tasks have either been completed or expired. Simulations of this algorithm have been run and limitations to its performance have been identified. Moreover, modifications to improve the algorithm's performance have been conjectured.

SCHUBMEHL, Caitlin

Mentor: Dorothy Peteet

The Paleoecology of Croton Marsh

Coastal marshes provide an ecological record of past climate and anthropogenic impacts on both local and watershed scales. The history of ecological and climate change in Croton Marsh was uncovered via samples from a 120-cm core. The core was examined for changes in seeds, foraminifera, charcoal, and organic content as indicators of past climate conditions in the marsh. The rises in charcoal levels indicate drier periods, and the rises in foraminifera indicate periods of higher salinity levels. While radiocarbon dating has not yet been conducted for this core, the rise the in number of the exotic species *Phragmites australis* at 40 cm could indicate the presence of European settlers at this time. Another interesting trend in the data is a sudden rise in *typha sp.* between 40 and 100 cm. The reason for this is unknown; the rise could have been influenced by the relatively low salinity levels, as indicated by the decline in the number of foraminifera found throughout this section of the core. Perhaps the increase in *Phragmites australis* accounts for the decline of *typha sp.* above 40 cm. Therefore, the arrival of European settlers appears to have had a profound impact on the ecology of the marsh.

SCHUDEL, Cecilia

Mentors: Luke Goetzke, Kyungeun Lim and Elena Aprile

Field-Induced Quenching of the Scintillation Light in Liquid Xenon Measured with the XeMini Detector

Data taken with the XeMini detector confirm field-induced quenching occurring in the 1kg dual-phase Xe time projection chamber operated at Nevis Laboratories as one of the XENON related R&D set-ups. With no voltage applied to the cathode, the energy deposited by 662 keV gamma-rays from an external Cs-137 radioactive source, and detected by photomultiplier tubes (PMTs) immersed in the liquid xenon (LXe) volume, is measured as 341 photoelectrons. As the cathode voltage is increased to 1kV, 2kV, and 3kV, the number of photoelectrons observed falls to 225, 184, and 160, respectively. At the maximum applied cathode voltage of 4kV, the number of photoelectrons observed begins to stabilize near 155. Application of a field results in the suppression of the recombination rate between the liberated ionization electrons and parent ions. This results in fewer excitons, which produce the VUV scintillation photons in LXe. Scintillation in liquid noble gases is due to both direct excitation of atoms by the incoming radiation and the recombination process. This reduced recombination rate by the application of an applied electric field in turn reduces the number of scintillation photons, thus decreasing the overall light yield. Decreasing the light yield to a manageable figure allows photons to be detected in an amount proportional to the original number of ionization electrons, so that the original number may be calculated from the amount of detected light.

These data are of interest because they are consistent with the theory of field-induced quenching in liquid xenon. Our new data are crucial to the calibration of the XeMini detector, which is in its final stage of data-collection.

SHAW, Katharina S.

Mentors: Nicholas F. Parrish, Matthias H. Kraus, Julie M. Decker and Beatrice H. Hahn (Departments of Medicine and Microbiology, University of Alabama, Birmingham)

A rev1-vpu polymorphism in HIV-1 Subtype C Impairs Envelope Glycoprotein Expression from rev-vpu-env Cassettes and Reduces Virion Infectivity of env Pseudotypes

The envelope (Env) glycoprotein of Human Immunodeficiency Virus Type 1 (HIV-1) is essential for virus entry into cells and for the transmission and propagation of virus *in vivo*. All transmitted HIV-1 *env* genes and their corresponding glycoprotein products are expected to be biologically functional when assayed *in vitro* and it was thus a surprise when we found a subset of HIV-1 subtype C strains where this was not the case. Biological analysis of HIV-1 Env glycoproteins *in vitro* commonly involves the generation of *trans*-complemented pseudoviruses by co-transfection of *env* expression vectors with an *env*-deficient proviral backbone. Here, we resolve the paradox of ostensibly defective transmitted HIV-1 Env glycoproteins by showing that a nucleotide polymorphism in the intergenic region between the first exon of *rev* (*rev1*) and *vpu* that places these two genes in-frame reduces downstream Env expression from a *rev-vpu-env* gene cassette. We found that disruption of the *rev1-vpu* fusion gene by frameshift mutation, insertion of an in-frame stop codon, or elimination of the *rev* initiation codon (Rev can be supplied *in trans*) restored Env expression and pseudovirion infectivity, while introduction of the mutant *rev1-vpu* fusion gene into wildtype constructs had the opposite effect. We further showed that the phenotypic defect was not due to reduced *env* or *rev* transcription, mRNA stability, or dominant negative effect of the expressed Rev1-Vpu fusion protein. Instead, the altered phenotype resulted from inefficient translation of Env from *rev-vpu-env* transcripts. Finally, we demonstrated that two *rev-vpu-env* cassettes that were phenotypically defective in Env pseudotype assays were functional when expressed in the context of their cognate proviruses, thus explaining the ability of these viruses to be transmitted *in vivo*. These findings support a paradigm of HIV-1 transmission biology (*J Exp Med* 206:1273, 2009) and provide a practical method for analyzing Env function of HIV-1 subtype C in Env-pseudotyping assays.

WANG, Xuran

Collaborators: Rachel Vishnepolsky and Jin Woo Jang

Mentors: Robert Lipshitz and Tim Perutz

Fixed-point Floer Homology

The algebraic notions of a chain complex and its homology groups began in algebraic topology but now pervade mathematics. In the 1980s, Andreas Floer introduced two homology theories based on solutions to certain very special non-linear partial differential equations occurring in geometry. Since then, further theories have been found on the same pattern. Some of these "Floer homology theories" probe the structure of 3- and 4-dimensional manifolds; others relate to symplectic manifolds and their symmetries. Work in progress by Prof. Perutz, affirming a conjecture of Seidel, shows that when a diffeomorphism is presented as a composite of Dehn twists along simple closed curves in a surface, one can reduce the computation of Floer homology *HF* to a situation where the Riemann mapping theorem applies. More explicitly, one can calculate *HF* by computing certain Hochschild complex, HH_*, of the Fukaya category generated by a sequence of curves associated with the Dehn twists. Over the summer my teammates Rachel Vishnepolsky, Jin Woo Jang, and I developed a computer program that

computes the Hochschild complex HH_* from suitable input data. We primarily used Sage, a Python-based mathematic software. We implemented the program to compute several examples, many of which are not algebraically computable by hand. Currently, we are working on a paper that provides the relevant background information as well as an explanation of our algorithm under the supervision of Prof. Lipshitz.

YIO, Jiang

Mentors: Xiuliang Bao and Steven Itzkowitz (Mount Sinai School of Medicine)

Structure and Function of the Trefoil Factor Family

The trefoil factor family consists of three small and structurally-similar proteins made in the gut: TFF1, TFF2, and TFF3. Trefoil factor 1 (TFF1) is a gastric tumor-suppressor, which human gastric cancers typically lack and knockout mice of which easily develop gastric adenomas and carcinomas. Trefoil factor 3 (TFF3), on the other hand, tends to worsen the severity of colon cancers, promoting metastasis and invasion. Both TFF1 and TFF3 serve protective functions in the gut mucosa by preventing apoptosis and allowing migration during wound healing. The purpose of the study was to determine the factors that differentiate these two similar proteins. Several TFF1 mutants exist in nature; two of them, namely A10D and E13K, are single-point mutants that enhance cell growth and invasive potential. Recombinant human TFF1 and TFF3 were compared functionally and structurally; a pathway array is in progress to determine the pathways that the two proteins participate in, and a computational model is being constructed to further visualize the similarities and differences between the proteins. The similarities within the trefoil factor family suggest that TFF1 and TFF3 may have diverged from a common ancestral protein, and the differences suggest that the divergence may have been due to small mutations such as A10D and E13K. Using computational methods to interpolate between TFF1 and TFF3, additional sites of mutagenesis were determined that may help explain the development of the trefoil factor family. These mutants will be prepared and assayed functionally.

ZHANG, Andrew

Mentor: Max Lipyanskiy

Applying the Poincare Theorem to the Elliptic Case

Hyperbolic geometry is a non-Euclidean geometry that rejects the parallel postulate, the statement that, in two dimensional space, for any given line l and point P not on l, there is exactly one line through P that does not intersect l (i.e., that is parallel to l). In hyperbolic geometry there are at least two distinct lines through P that do not intersect l.

A manifold is a mathematical space that on a small enough scale resembles the Euclidean space of a certain dimension, called the dimension of the manifold. The project focused on constructing hyperbolic manifolds based on a generating set of isometries. Poincare's Theorem states that if we start with a polygon P and a collection of side-pairing maps and use these maps to generate a group G, then we can prove that G is discrete and that P is a fundamental domain for G if they satisfy certain conditions.

We present a computer algorithm that proves the discreteness of a subgroup of PSL(2,C). The algorithm proceeds by constructing the Dirichlet domain in H3 of the subgroup, and then checks that the hypotheses of Poincare's Theorem are satisfied. In order to show that the hypotheses are

exactly satisfied, we rely on group theoretical properties resulting from certain geometric conditions of the Dirichlet domain.

Rabi Scholars (2009-10)

Elizabeth Allocco (CC'11, Neuroscience)
Atanas Atanasov (CC'10, Mathematics)
Samuel Beck (CC'11, Physics/Mathematics)
Ilya Belopolski (CC'12, Physics)
*Rohan Bhandari (CC'13, Chemistry)
Corey Bregman (CC'10, Physics)
*Alice Chang (CC'13, Chemistry)
*Vadim Chanyshev (CC'13, Mathematics)
*Kevin Chou (CC'10, Mathematics/ Economics)
*Woo Chang Chung (CC'13, Physics)
Jay Dhuldhoya (CC'11, Biophysics)
*Meagan Gamache (CC'13, Physics)
Alexandru Georgescu (CC'11, Physics)
*Zachary Gray (CC'11, Astrophysics/ Philosophy)
*Lalit Gurnani (CC'11, Biology/Economics)
Stephen Hancock (CC'11, Physics)
Maximilian Horlbeck (CC'11, Biochemistry/Computer Science)
*Jonathan Huggins (CC'12, Computer Science/Mathematics)
Joon Ho Kang (CC'12, Biophysics)
Michael Kennelly (CC'11, Physics)
*Alexis Kurmis (CC'13, Chemical Physics)
Andre Lazar (CC'12, Biophysics/ Biochemistry)
Christopher Lopez (CC'11, Mathematics)
*Dan Margulies (CC'13, Chemical Physics)
Ion Mihailescu (CC'10, Physics/ Mathematics)
Man-Yu Moy (CC'11, Biology)
*James Nugent (CC'13, Mathematics)
*Julia Oktawiec (CC'13, Biochemistry)
*Jungsik Park (CC'13, Physics)
Milesh Patel (CC'12, Biochemistry)
Alexander Perry (CC'11, Mathematics)
Pawel Przytycki (CC'11, Computer Science)
Mitchell Rubenstein (CC'11, Computer Science-Mathematics)
Caitlin Schubmehl (CC'10, Neuroscience)
Cecilia Schudel (CC'13, Physics)
*Jenny Shao (CC'13, Biochemistry/ Chemical Physics)
†*Katharina Shaw (CC'13, Biochemistry/ Biophysics)
*Henry Sohn (CC'13, Physics)
*Svetozar Stankov (CC'13, Mathematics)
*Mohit Tiwari (CC'13, Physics/ Mathematics)
Xuran Wang (CC'10, Mathematics)
Jiang Yio (CC'10, Biology)
Andrew Zhang (CC'11, Mathematics)

*Rabi Scholars not undertaking research projects in summer, 2009.

†Abstract based on Intel project, 2009.