Final Progress Report for Research Projects Funded by Health Research Grants

Instructions: Please complete all of the items as instructed. Do not delete instructions. Do not leave any items blank; responses must be provided for all items. If your response to an item is "None", please specify "None" as your response. "Not applicable" is not an acceptable response for any of the items. There is no limit to the length of your response to any question. Responses should be single-spaced, no smaller than 12-point type. The report **must be completed using MS Word**. Submitted reports must be Word documents; they should not be converted to pdf format. Questions? Contact Health Research Program staff at 717-783-2548.

1. **Grantee Institution:** Carnegie Mellon University

2. **Reporting Period (start and end date of grant award period):** 1/1/10 – 12/31/13

3. **Grant Contact Person (First Name, M.I., Last Name, Degrees):** Jim Osborn

4. **Grant Contact Person's Telephone Number:** 412 268-6553

5. **Grant SAP Number:** 410005089

6. **Project Number and Title of Research Project:**

   Project 3 - Computational Immunology for Toleragenic Composite Tissue and Solid-Organ Transplantation

7. **Start and End Date of Research Project:** 1/1/10 – 12/31/13

8. **Name of Principal Investigator for the Research Project:** Jaime Carbonell, PhD

9. **Research Project Expenses**.

   9(A) Please provide the total amount of health research grant funds spent on this project for the entire duration of the grant, including indirect costs and any interest earned that was spent:

   $ 74,568.27

   9(B) Provide the last names (include first initial if multiple individuals with the same last name are listed) of **all** persons who worked on this research project and were supported with health research funds. Include position titles (Principal Investigator, Graduate Assistant, Post-doctoral Fellow, etc.), percent of effort on project and total health research funds expended for the position. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name, First Name	Position Title	% of Effort on Project	Cost
Starzl, Ravi	Researcher	58.33% Yr 1; 45.83% Yr2	$59,608.50

9(C) Provide the names of **all** persons who worked on this research project, but who *were not* supported with health research funds. Include position titles (Research Assistant, Administrative Assistant, etc.) and percent of effort on project. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

Last Name, First Name	Position Title	% of Effort on Project
Carbonell, Jaime	PI	0%; advisory role only

9(D) Provide a list of **all** scientific equipment purchased as part of this research grant, a short description of the value (benefit) derived by the institution from this equipment, and the cost of the equipment.

Type of Scientific Equipment	Value Derived	Cost

10. Co-funding of Research Project during Health Research Grant Award Period. Did this research project receive funding from any other source during the project period when it was supported by the health research grant?

Yes_____ No____x_____

If yes, please indicate the source and amount of other funds:

11. Leveraging of Additional Funds

11(A) As a result of the health research funds provided for this research project, were you able to apply for and/or obtain funding from other sources to continue or expand the research?

Yes_____ No____x_____

If yes, please list the applications submitted (column A), the funding agency (National Institutes of Health—NIH, or other source in column B), the month and year when the application was submitted (column C), and the amount of funds requested (column D). If you have received a notice that the grant will be funded, please indicate the amount of funds to be awarded (column E). If the grant was not funded, insert "not funded" in column E.

Do not include funding from your own institution or from CURE (tobacco settlement funds). Do not include grants submitted prior to the start date of the grant as shown in Question 2. If you list grants submitted within 1-6 months of the start date of this grant, add a statement below the table indicating how the data/results from this project were used to secure that grant.

A. Title of research project on grant application	B. Funding agency (check those that apply)	C. Month and Year Submitted	D. Amount of funds requested:	E. Amount of funds to be awarded:
	☐NIH ☐ Other federal (specify:_____ _____) ☐ Nonfederal source (specify: _____)		$	$
	☐NIH ☐ Other federal (specify:_____ _____) ☐ Nonfederal source (specify: _____)		$	$

11(B) Are you planning to apply for additional funding in the future to continue or expand the research?

Yes____X_____ No_____

If yes, please describe your plans:

We intend to apply for NIH and DOD grants to gather additional samples, improve these methods, and investigate identified genes of interest.

12. **Future of Research Project.** What are the future plans for this research project?
The models, methods, and findings of this research has spawned several subsequent research projects that are already in various stages of progress with other collaborating institutions, including projects to:
    - Determine the disease state of a heart transplant by imaging the migration and localization of macrophages and lymphocytes
    - Monitor the rejection state of an allograft using the algorithms developed in this project with non-invasive tape-stripping sample collection
    - Predict the survival in tissue reperfusion scenarios with greater accuracy than is currently possible
    - Investigate how the inflammasome assembles on the basis of complex and specific cytokine milieus, and how the inflammasome affects the local immunologic process
    - Develop fundamental principles that accurately explain the ability of the immune system is able to rapidly transmit information across large distances and affect complex outcomes with a high degree of consistency and effectiveness.

13. **New Investigator Training and Development**. Did students participate in project supported internships or graduate or post-graduate training for at least one semester or one summer?

Yes_____x____ No_____

If yes, how many students? Please specify in the tables below:

	Undergraduate	Masters	Pre-doc	Post-doc
Male			1	
Female				
Unknown				
Total			**1**	

	Undergraduate	Masters	Pre-doc	Post-doc
Hispanic				
Non-Hispanic			1	
Unknown				
Total			**1**	

	Undergraduate	Masters	Pre-doc	Post-doc
White			1	
Black				
Asian				
Other				
Unknown				
Total			**1**	

14. Recruitment of Out-of–State Researchers. Did you bring researchers into Pennsylvania to carry out this research project?

Yes_____ No____x_____

If yes, please list the name and degree of each researcher and his/her previous affiliation:

15. Impact on Research Capacity and Quality. Did the health research project enhance the quality and/or capacity of research at your institution?

Yes_____x____ No_____

If yes, describe how improvements in infrastructure, the addition of new investigators, and other resources have led to more and better research.

This research built cross-institutional bridges with other world-class centers of investigation like Johns Hopkins Medical Institute, the Pittsburgh NMR center, the Cleveland Clinic, and the University of Pittsburgh. Several collaborators from around the world participated in the research and continue to carry forward the ideas developed in this project. Through these collaborations, access to high-tesla MRI machines, sophisticated laboratory equipment (such as Luminex machines and gel electrophoresis columns), surgical expertise, and much more has been made available to Carnegie Mellon University researchers, which has promoted and enabled additional collaborations beyond the scope of this project alone.

16. Collaboration, business and community involvement.

16(A) Did the health research funds lead to collaboration with research partners outside of your institution (e.g., entire university, entire hospital system)?

Yes____x_____ No_____

If yes, please describe the collaborations:

The raw data gathered was through ongoing collaborative projects with the University of Pittsburgh Departments of Immunology, Plastic and Reconstructive Surgery, The

McGowan Institute for Regenerative Medicine, The Pittsburgh Nuclear Magnetic Resonance Center, and the University of Innsbruck School of Medicine.

16(B) Did the research project result in commercial development of any research products?

Yes_____ No____x_____

If yes, please describe commercial development activities that resulted from the research project:

16(C) Did the research lead to new involvement with the community?

Yes_____ No_____x_____

If yes, please describe involvement with community groups that resulted from the research project:

17. Progress in Achieving Research Goals, Objectives and Aims.

List the project goals, objectives and specific aims (as contained in the grant agreement). Summarize the progress made in achieving these goals, objectives and aims for the period that the project was funded (i.e., from project start date through end date). Indicate whether or not each goal/objective/aim was achieved; if something was not achieved, note the reasons why. Describe the methods used. If changes were made to the research goals/objectives/aims, methods, design or timeline since the original grant application was submitted, please describe the changes. Provide detailed results of the project. Include evidence of the data that was generated and analyzed, and provide tables, graphs, and figures of the data. List published abstracts, poster presentations and scientific meeting presentations at the end of the summary of progress; peer-reviewed publications should be listed under item 20.

This response should be a DETAILED report of the methods and findings. It is not sufficient to state that the work was completed. Insufficient information may result in an unfavorable performance review, which may jeopardize future funding. If research findings are pending publication you must still include enough detail for the expert peer reviewers to evaluate the progress during the course of the project.

Health research grants funded under the Tobacco Settlement Act will be evaluated via a performance review by an expert panel of researchers and clinicians who will assess project work using this Final Progress Report, all project Annual Reports and the project's strategic plan. After the final performance review of each project is complete, approximately 12-16 months after the end of the grant, this Final Progress Report, as well as the Final Performance Review Report containing the comments of the expert review panel, and the grantee's written response to the Final Performance Review Report, will be posted on the CURE Web site.

There is no limit to the length of your response. Responses must be single-spaced below, no smaller than 12-point type. If you cut and paste text from a publication, be sure symbols print properly, e.g., the Greek symbol for alpha (α) and beta (ß) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.

Specific aim 1:
Isolate genes specific to rejection in skin and muscle in composite tissue transplantation
Enumerate genes for confirmation by collaborating centers
Progress made on aim 1:
Specific aim 1 was achieved in a rat model sample set. Genes that are activated only under conditions of allograft transplant or isograft transplant have been identified are discussed beginning page 14. These results have been shared with collaborators at the University of Pittsburgh, Johns Hopkins Medical Institute, the Cleveland Clinic, and the University of Innsbruck for further investigation.

Specific aim 2: Develop tissue-specific rejection early warning models comprised of:
1) Proteomic data classifiers for skin, muscle, heart and lung rejection from both composite tissue and solid organ transplants in animal models
2) Genetic data classifiers for skin and muscle rejection utilizing data from composite tissue transplants in animal models
Progress made on aim 2:
Specific aim 2 part 1 was achieved. Discussion of the model begins on page 10. The final results are discussed on pages 22 and 23.
Specific aim 2 part 2 was achieved to the extent that available data allowed. Specific discussion of the use of a k-means clustering algorithm to identify gene activations unique to allografts and isografts is discussed beginning on page 15.

Specific aim 3: Develop gene-cytokine-cell system models enumerating the processes that initiate, propagate and potentially down regulate rejection.
Progress made on aim 3:
The initial foundations for a potential system that is able to enumerate and infer the complex regulatory activity of an immune signaling network (from protein to cell) called a 'cellulograph' is discussed on page 18.

Specific aim 4: Translate successful models from animal to human data utilizing samples from transplant programs at the University of Innsbruck and the University of Pittsburgh.
Progress made on aim 4:
Translation into human samples is currently being pursued by the investigators in collaboration with teams from Johns Hopkins Medical Institute, the University of Pittsburgh, and the University of Innsbruck. Access to human samples is in the process of being requested or evaluated by the institutional review boards of these centers.

In this final report for the project *Computational Immunology for Toleragenic Composite Tissue and Solid-Organ Transplantation* we detail the progress we have made to elucidate the tissue specific biological and informational principles of the complex immunological mechanisms and systems that drive rejection and inflammation in solid or composite tissue transplants.

Both solid organ and composite tissue transplantation has traditionally required extensive life-long immunosuppression therapy to prevent rejection of the graft. It is this long-term high-dose immunosuppression regimen that is the limiting factor for both the quality of organs that can be used for a solid organ transplant and for the widespread adoption of composite tissue transplantation. Further, the present clinical method for detecting rejection in solid organ transplant is fluctuation of organ function (caused by damage from rejection), or skin discoloration in the case of composite tissue rejection.

The recent paradigm of transplant tolerance, or chimerism, has revealed a conceptual framework where it may be possible to detect the onset of rejection before tissue damage occurs, adjust immunosuppression dosage to optimal levels, and potentially wean patients to very low or no maintenance doses of immunosuppression.

Through advanced computational methods, machine learning and other high dimensional analytic techniques, a deeper understanding of the patterns and governing principles of the tolerance paradigm, as well as the immune system in general, has been reached through this project, and opens a promising pathway for future clinically relevant toleragenic transplant protocols, long term patient care strategies, and potentially therapeutics that make use of the understanding of immunologic systems organization found here. This project has contributed to the advancement of the state of the art in transplant tolerance and has helped clinicians understand the risks of indiscriminant immunosuppression, and how to reduce the risks of over-immunosuppression. The models and methods developed in this project has provided collaborating clinicians with tools and information to better understand the likely immunologic mechanisms of transplant tolerance, and illuminated a new approach that may lead to less toxic immunosuppression regimens, and potentially new classes of drugs that act on the immune signaling process itself.

A significant amount of new information about how the immune system organizes its response to several specific types of controlled insult have revealed that there the immune system is capable of rapidly disseminated large amounts of information throughout its constituent components, and that these components themselves are able to rapidly act in ways more sophisticated than previously believed. In short, we see the immune system responding locally but coordinating systemically, and that in addition to pattern recognition at the level of antigen presentation cells, the overall degree of immune activation, the location of immune activation, and the specific cell types involved in immune signaling can all have a significant affect on the subsequent immunologic response.
The quantitative models generated in this project were built specifically to allow the toleragenic paradigm to be personalized and adapted to each individual patient, based on the makeup of their individual immunologic characteristics. This can allow better allograft management and the treatment of rejection prior to the accumulation of damage that affects

organ function. With better proactive management, a larger pool of organs that could be used in transplant can be made available, helping to alleviate the worldwide shortage of kidney, liver, pancreas and other organs.

The specific goal of this project was to elucidate immunological patterns and mechanisms in pre-clinical composite tissue and solid organ transplants under various conditions of rejection, inflammation and tolerance. To achieve this, machine learning, computational linguistics, agent based modeling, and differential equation based methods of analysis were applied to proteomic and genetic data. The raw data was gathered through collaboration with Johns Hopkins Medical Institute, the University of Pittsburgh Departments of Immunology, Plastic and Reconstructive Surgery, The McGowan Institute for Regenerative Medicine, The Pittsburgh Nuclear Magnetic Resonance Center, the University of Innsbruck School of Medicine, and the Cleveland Clinic.

We capture these key immunologic patterns with models that have predictive power for rejection state, when given proteomic information about the immune signaling milieu present in a local area, annotated with post-operative time point, and graft type. As part of the work for this project, models were built for specific tissues including skin, muscle, heart, lung, liver, kidney, pancreas, lymph node, and serum. Protein data was provided in mass values (pg specific protein/mg total protein) as read by a Luminex 100-IS machine at the University of Pittsburgh Department of Immunology.

Genetic analysis was possible with gene array data provided by the composite tissue transplant program at the University of Pittsburgh Department of Plastic and Reconstructive Surgery as well as the University of Innsbruck School of Medicine. With this data we hoped to develop models with predictive power for rejection state, identification of rejection specific gene expression and enumeration of these targets for verification, and corroboration of gene-protein expression under various rejection and inflammation conditions. Through exploratory data analysis, we were able to identify many promising correlations between elevated levels of specific proteins and elevated gene expression levels, as well as differentially elevated genes that expressed only during allograft rejection or only during isograft engraftment. The majority of these differentially enumerated genes remain uncharacterized in the literature and present an exciting array of targets for further investigation. We hope in future projects to collect additional data and find more effective ways to explore and make use of this potentially very valuable perspective into the workings of the immune response.

Early Detection and Classification of Rejection

This project has developed computation models that are capable of accurately identifying rejecting samples in the early stages of rejection (grade three or below) utilizing protein quantification data to identify predictive cytokine network profiles in both solid organ and composite tissue transplant contexts. A listing of all groups and sampling time points is made in Table 1.

Surgical Model	*Tissues Sampled*	*Time Points Sampled*
Hind Limb VCA	Skin, Muscle, Serum	POD 3,5,7,9,11
Heart and Lung Transplant	native heart, graft heart, native lung, graft lung, lymph node, spleen, kidney, liver and serum	5 hours, POD 1,3,5,7,9
Naïve Wound	Skin, Muscle	N/A
Full-Thickness Skin Transplant	Skin, Muscle, Serum	Hours 6, 12, 24 POD 3, 21, 45
SMB Transplant	Skin, Muscle, Serum	Hours 6, 12, 24 POD 3, 21, 45

Table 1: Rat surgical models and tissue types analyzed

Data for solid organs came from rat cardiac (heart and lung) transplant surgeries conducted at the Nuclear Magnetic Resonance (NMR) center of Carnegie Mellon University, using a Brown-Norway (BN) to Agouti (DA) heterotopic heart and lung transplant technique described in Ho 2001 [Macrophage Accumulation Associated With Rat Cardiac Allograft Rejection Detected by Magnetic Resonance Imaging With Ultrasmall Superparamagnetic Iron Oxide Particles. Kanno et al. Circulation. 2001; 104: 934-938 doi: 10.1161/hc3401.093148]; allogeneic [n=13] and syngeneic [n=4] cardiac transplants were studied, as well as naïve [n=2] rats. The heterotopic transplant technique allows the native heart and lungs to remain functional in the animal, allowing samples to be drawn from both native organs as well as allografts. At every time point native heart, graft heart, native lung, graft lung, lymph node, spleen, kidney, liver and serum samples were taken.

Data for cytokine dynamics in composite tissue transplants was collected from rat hind limb transplant surgeries conducted at the Department of Plastic and Reconstructive Surgery - Composite Tissue Allotransplantation Group, at both the Johns Hopkins Medical Institute and the University of Pittsburgh using a Brown-Norway (BN) to Lewis (LEW) rat hind-limb allotransplant technique; syngeneic [n=10], allogeneic transplants without immunosuppression [n=10], allogeneic grafts [n=10] treated with tacrolimus and naïve rats [n=10] were studied. 180 skin and muscle biopsies were taken between postoperative days 0 and 11. Protein levels of 14 cytokines known to be relevant in cellular inflammatory responses were assessed by Luminex™.

All tissue samples are 4mm squares preserved in RNALater and frozen at -25C for 24 hours, then moved to -80C. Blood samples are processed into Serum, then frozen at -25C for 24 hours, then moved to -80C.

To quantify the amount of relevant protein present, samples were homogenized and protein levels of 14 cytokines known to be relevant in cellular inflammatory responses (GMCSF, Il-1α, Il-1β, Il-5, Il-4, Il-6, Il-8, Il-10, Il-18 TNF-α, IFN-γ, Il-12 p70, GRO/KC and MCP-1) were assessed by Luminex® 100 IS using the Rat Cytokine/Chemokine Luminex® Bead immunoassay Kit, LINCOplex, 14 Plex. Each sample was bisected and run in two separate

arrays of wells. The final protein concentration values were calculated as the mean of the values from the two arrays.

All computational analyses utilized Matlab, the Weka machine learning toolkit (v 3.6), and Microsoft Excel (v 2008). Luminex® protein readings were provided in a comma separated value (CSV) format that were then converted to attribute-relation file format (ARFF) for further analysis in Weka Explorer. Normalization of data was performed by manually programmed min-max normalization formula in MS Excel, with the form

$$X^* = (X - \min(X))/\mathrm{range}(X).$$

Linear classification analysis utilized the Naive Bayes classifier tool within Weka, while decision tree analysis utilized the J48 algorithm. Feature extraction was aided by the information gain attribute evaluator (ranker search algorithm), as well as the correlation-based feature subset evaluator (best first algorithm). Decision trees, entropy values and classification results were generated with Weka visualizer and explorer, while graphs were generated with MS Excel.

In the solid organ context, Allograft/Isograft groups and Native tissue can be clearly distinguished from each other at all time points. Allograft and Isograft groups are also linearly separable from each other as early as Post-Operative Day (POD) 1, but with a slightly increased chance of false positive.

The average cytokine concentration levels present in the skin of a composite tissue allograft (Figure 1) are distinct from those present in the muscle (Figure 2). It is immediately clear that the specific cytokines that vary between isograft and allograft groups are also distinct between in skin versus muscle.

In skin, IL-18 is one of the single most important differentiator between Native tissue and Isograft/Allograft tissue (Figures 1, 3). Allograft IL-18 levels at POD 1 are higher than Native tissue by a factor of approximately 7, and Isograft IL-18 levels at POD 1 are higher than Native tissue by a factor of approximately 5. The J48 decision tree algorithm (Weka) was able to create a classification rule that distinguished Native heart tissue from Allograft heart tissue at all time points in this preliminary data set. The J48 decision tree algorithm was also able to derive a rule that distinguished allograft from native lung tissue based on elevated IL-1a levels in the allograft.

In the context of composite tissue allotransplant, protein levels for postoperative day 3 are standard normalized (mean 0, standard deviation 1) as a preprocessing step for each of the cytokines. Feature selection was then applied with the least absolute selection and shrinkage operator (LASSO). This method applies an L1 constraint while minimizing residual sum of squares driving some of the predictors exactly to 0, eliminating them from the model. Picking the following cytokines with the highest weight thus reduced the number of relevant features to three: TNF-a, Il12 p70 and Il-5.

Using mass-normalized protein levels for the chosen three cytokines, multiclass logistic regression was applied to distinguish between Allograft, Isograft and tacrolimus treated samples. Logistic regression is a linear regression technique that applies logit function to compute posterior probabilities of classes given the data. Under this model, probability of a class j with the exception of the last class is for sample X_i is:

$$P(C_j \mid X_i) = \exp(\beta_j^T X_i) / \Sigma_k \exp(\beta_k^T X_i) \quad k = \text{number of classes}$$

where parameters β_i are estimated using an efficient algorithm called iterative reweighted least squares. We applied leave one out cross validation (LOOCV) to test the classifier.

When multiclass logistic regression is applied without feature selection, 4 out of 14 samples are misclassified leading to 71.4% accuracy, indicating that multiple cellular or physiological processes generate the cytokine concentrations observed. The final concentration value of the cytokines/chemokines likely represents overlapping distributions from these multiple processes. However, when we reduce the feature space to be more tractable through LASSO feature selection (yielding TNF-a, Il12 p70 and Il-5), and then apply a logistic regression classifier (utilizing 10-fold cross validation), all samples were successfully classified to their correct groups at post operative day 3, obtaining 100% accuracy. Lasso reduced noise from the model, revealed cytokines that are discriminative for rejection, inflammation and increased the learning performance. Differentiation between Allograft and Isograft groups is possible at post-operative day (POD) 1, and becomes progressively more pronounced at PODs 3, 5, 7 and 9.

These results are promising and suggest an important link exists between cytokine network dynamics and the state (as well as progression) of rejection in solid and composite tissue transplantation. The cytokine dynamics indicative of rejection are separable at early time points and apparently form an effective set of bases for the prediction of acute rejection. Further, the cytokine network profiles observed suggest that there is significant attribute interaction in the underlying immunological processes that drive rejection, and that interrupting these processes, either at a cytokine or cellular level, may be likely to alter the course of rejection or rejection pathology.

In both the composite tissue allotransplantation and cardiac transplantation models additional connections to underlying mechanisms through corroborative findings in the literature or direct observation has helped direct the clinical application of these findings.

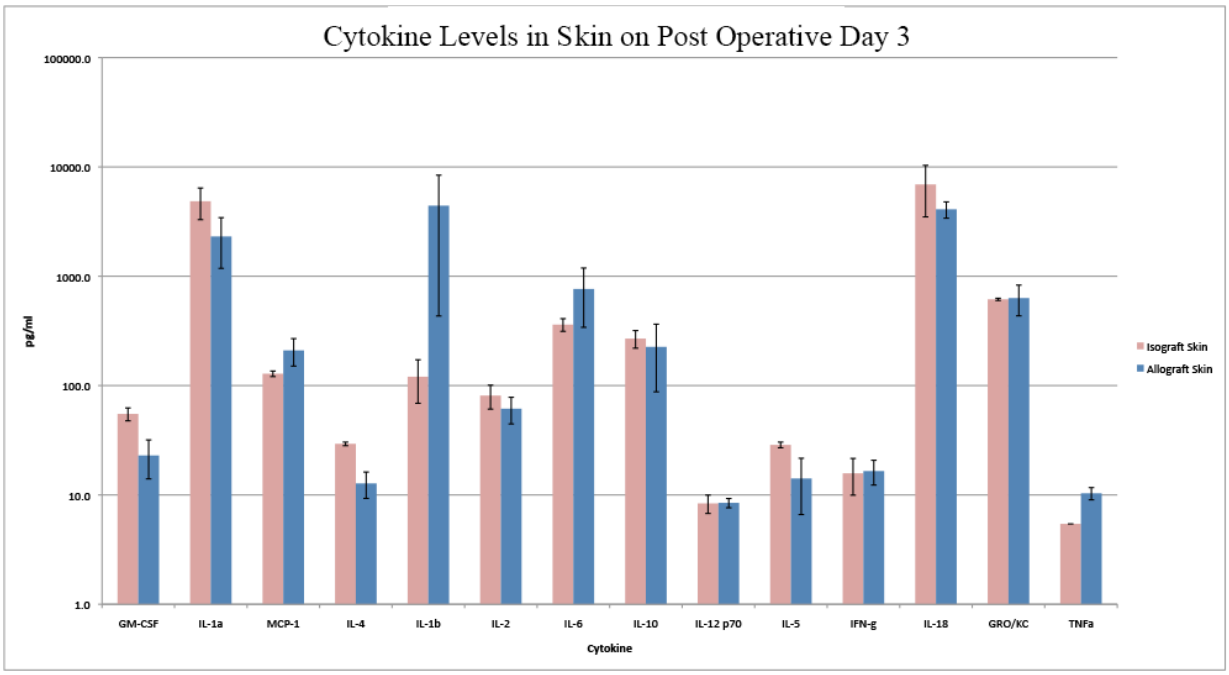

Figure 1: Cytokine levels in skin of composite tissue allografts on postoperative day 3.

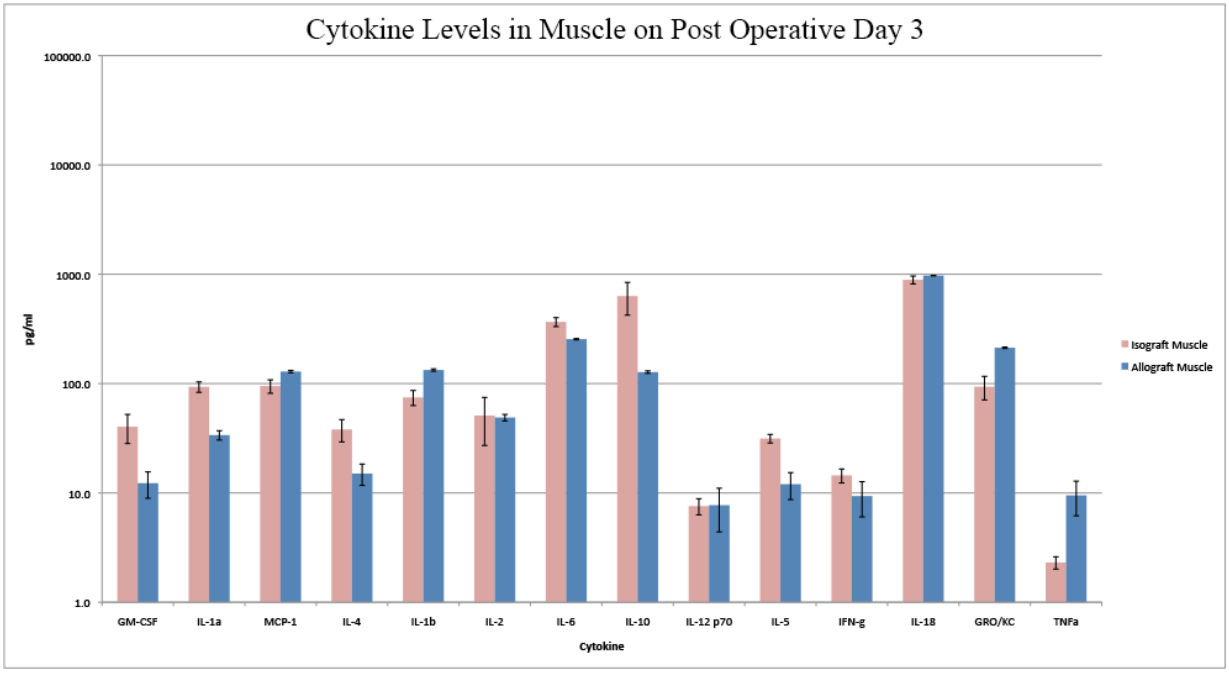

Figure 2: Cytokine levels in muscle of composite tissue allografts on postoperative day 3.

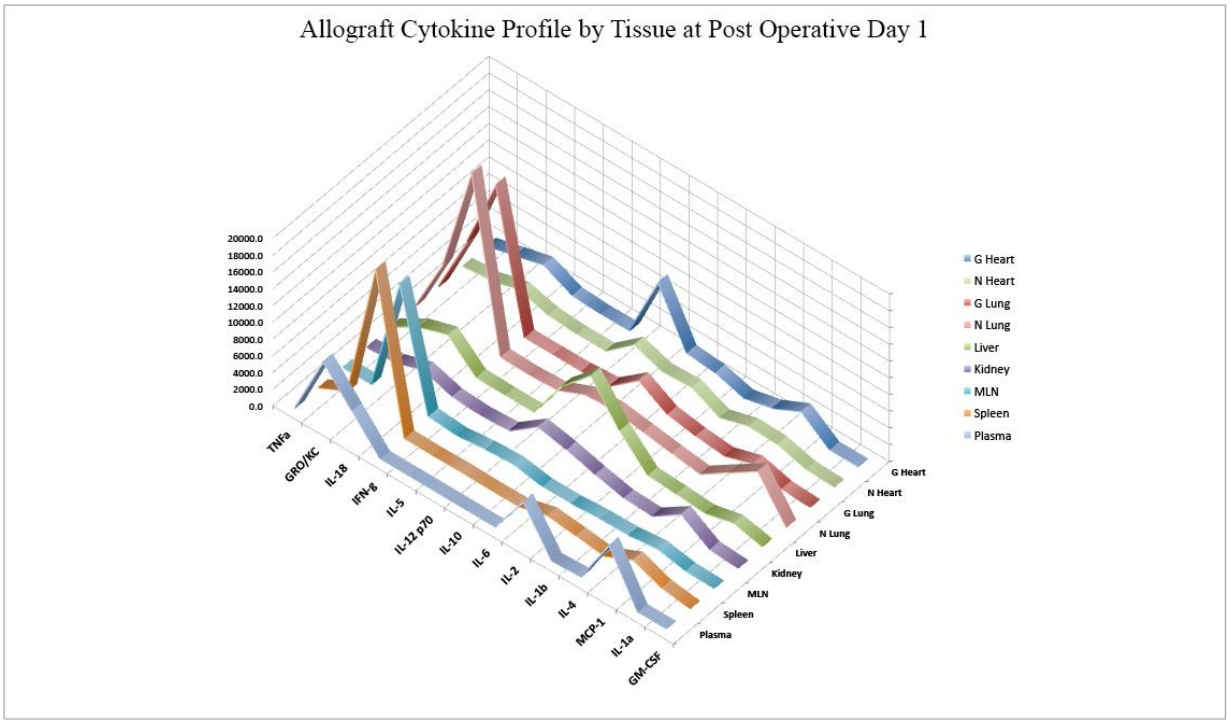

Figure 3: Cytokine levels in all sampled tissues of allograft cardiac transplant on postoperative day 1.

As is already well understood, the primary platforms for immunological signaling in the immune system are cytokines and chemokines. These immune signaling proteins affect cell differentiation, migration, degranulation, and a host of other essential immune activities. Understanding the patterns of immune signaling in greater depth allowed us to build upon these promising results by extending the analysis into more informative discriminative models and further explore the underlying factors that drive the immunologic response.

Differential Gene Expression

Samples were measured for gene expression by microarray, providing data on 22,226 individual genes. The output of this is a data set that represents expression intensity values by gene. This provides a probabilistic measure of whether the gene is expressed, relative to other readings on the microarray. Thus the readings of microarrays do not indicate definitively which genes are or are not expressed, but provide a probabilistic view of the activation matrix, and require additional assessment to determine their potential relevance.

To complete this additional assessment, we established a method for inferring with a very high level of confidence whether the gene was in fact expressed, based entirely on the intensity values in the microarray dataset. Inherently, there are two classes present in gene expression; on and off. Using this fact, we used simple K-Means clustering algorithm to determine a threshold for high-confidence discretization of the data into these on/off values. This revealed that 298 genes of the original 22,226 had a differentially high expression in either isograft or allograft samples. This shows that approximately 1% of the data is of interest. A great simplification, however 298 is still a large pool of potential genes to investigate.

We further differentiated what genes were most interesting by plotting the two classes against each other, with isograft (ISO) plotted along the X axis and allograft (ATC) along the Y axis (Figure 4).

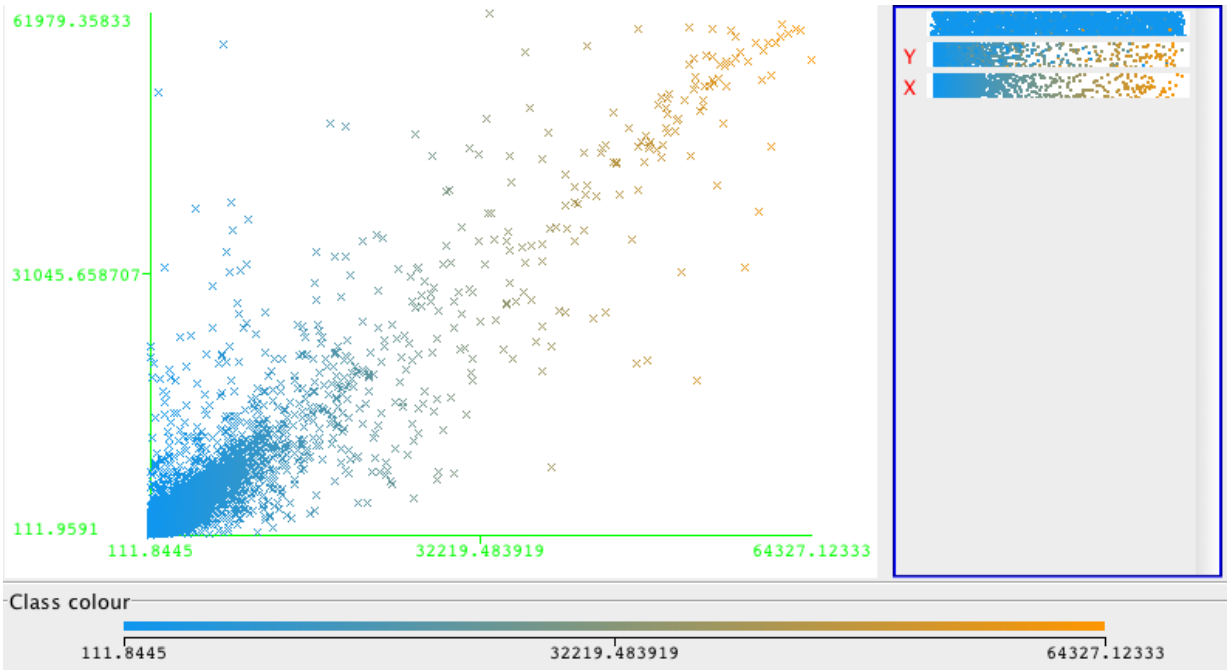

Figure 4: Expression levels of allograft and isograft genes as measured by micro-array.

This figure shows that there is a handful of outliers that are expressed highly in allograft samples but not isograft samples, and vice versa. At first glance it appears that simply dividing the graph into quadrants will provide sufficient filtering, but there are still a significant number of genes that express near the center of the graph. We therefore divided the graph into eight parts, to achieve more specific rules for the classification criterion:

Allograft criterion) Expression in ATC > 32,000 and expression in ISO < 16,000
Isograft criterion) Expression in ISO > 32,000 and expression in ATC < 16,000

Application of these filtering rules to the data set yields several specific genes of interest, enumerated in Table 2 and highlighted in Figures 5 and 6.

ATC Target ID (Gene)	**ATC Value (Signal)**	**ISO Target ID (Gene)**	**ISO Value (Signal)**
ILMN_49461	38789.52167	ILMN_55504	39161.48583
ILMN_54193	32110.55333		
ILMN_69223	52496.52917		
ILMN_55502	33611.325		
ILMN_61448	39438.44417		
ILMN_60657	58145.53583		
ILMN_47707	36127.24667		
ILMN_53575	37381.755		

Table 2: Genes of that differentially express in either allograft or isografts,

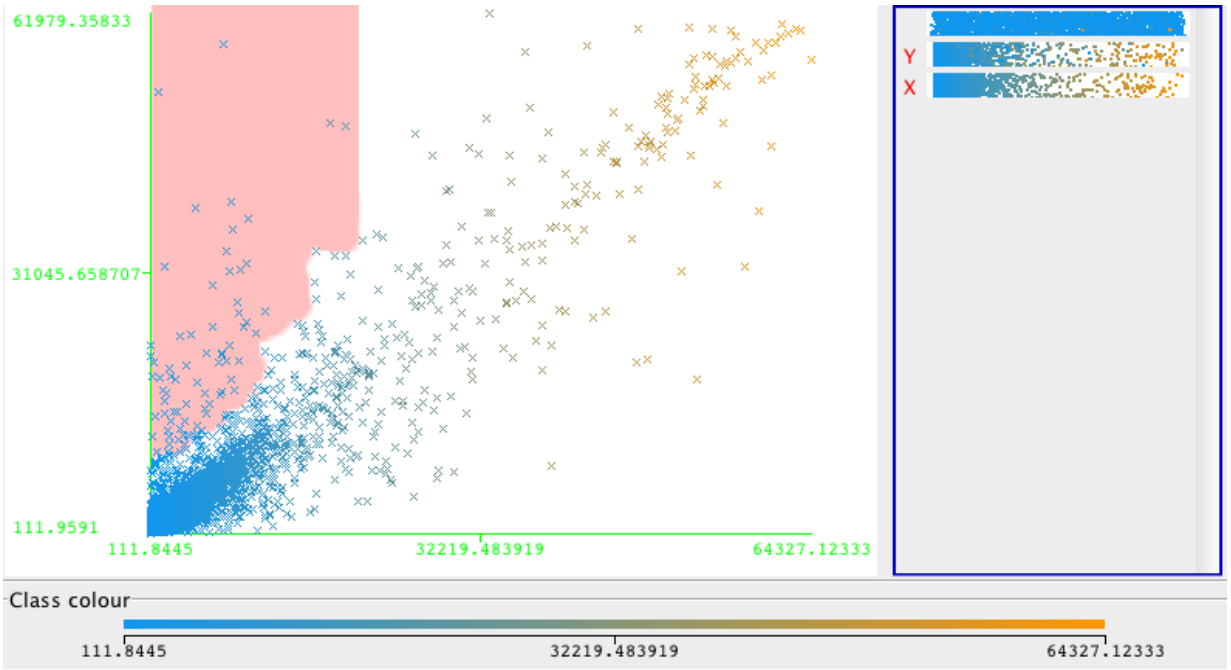

Figure 5: Highlighted genes expressed are specific to allograft samples

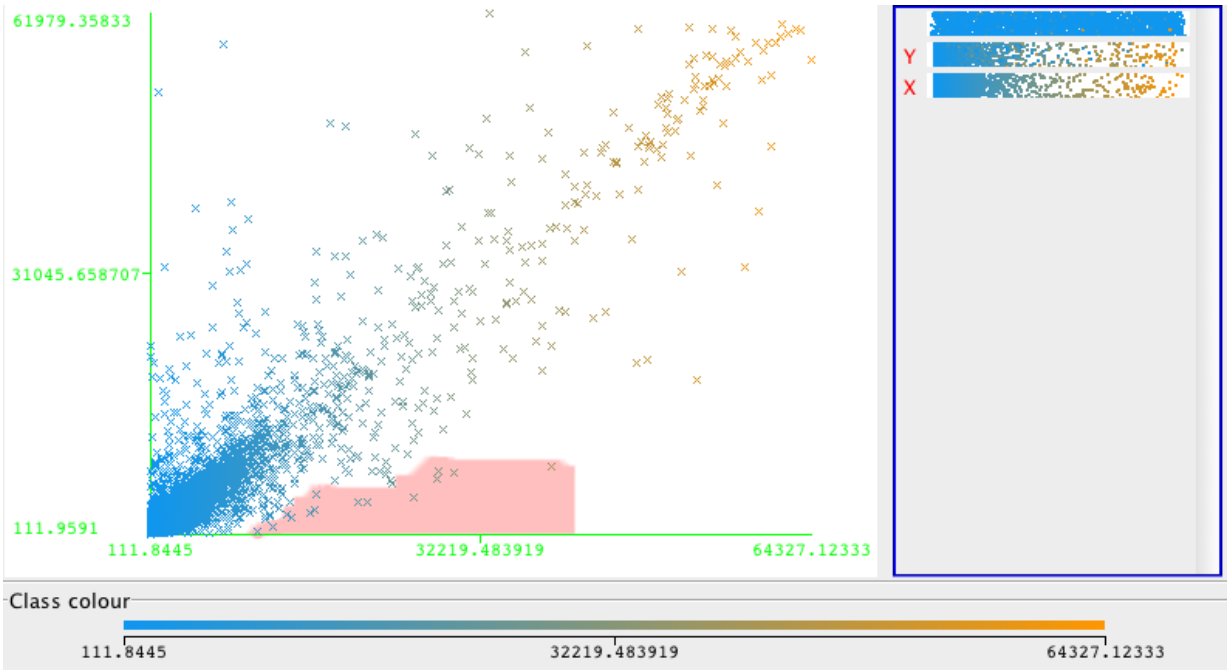

Figure 6: Highlighted genes expressed are specific to isograft samples.

Towards Development of an Extensible General Model for Inferring the Relationships between Proteins and Cell Populations

To help provide additional information and data about relevant contexts, additional samples were collected from a Naive Wound group to clarify inflammation profiles surrounding

surgical trauma and reperfusion injury, and a full-thickness skin transplant group to provide information about the inflammation profile associated with skin-only transplant. This model showed some of the most vigorous rejection response, as the skin is highly antigenic and there are no "protective" tissues such as muscle or bone that may ameliorate the acute immune response. Finally Skin-Muscle-Bone (SMB) transplant samples were also collected to provide information about the inflammation profile associated with transplants that incorporate a skin, muscle, and bone component. The vigor of rejection in this model is not as vigorous as Full-Thickness Skin (most vigorous), but more vigorous than Vascularized Composite Allografts (VCA). These groups were all sampled within 24 hours of transplant, and showed a significant amount of predictive immune activity occurred within the first 24 hours post transplant.

This work developed methods by which the functional cellular population can be inferred from the cytokine mélange as measured by Luminex. This allowed us to produce a new type of graph termed the "cellulograph," which is a snapshot of the functional cell population as it is most likely to exist in order to have produced the cytokine profile observed. The output of this method applied to the data from heart allograft samples can be seen in Figure 7.

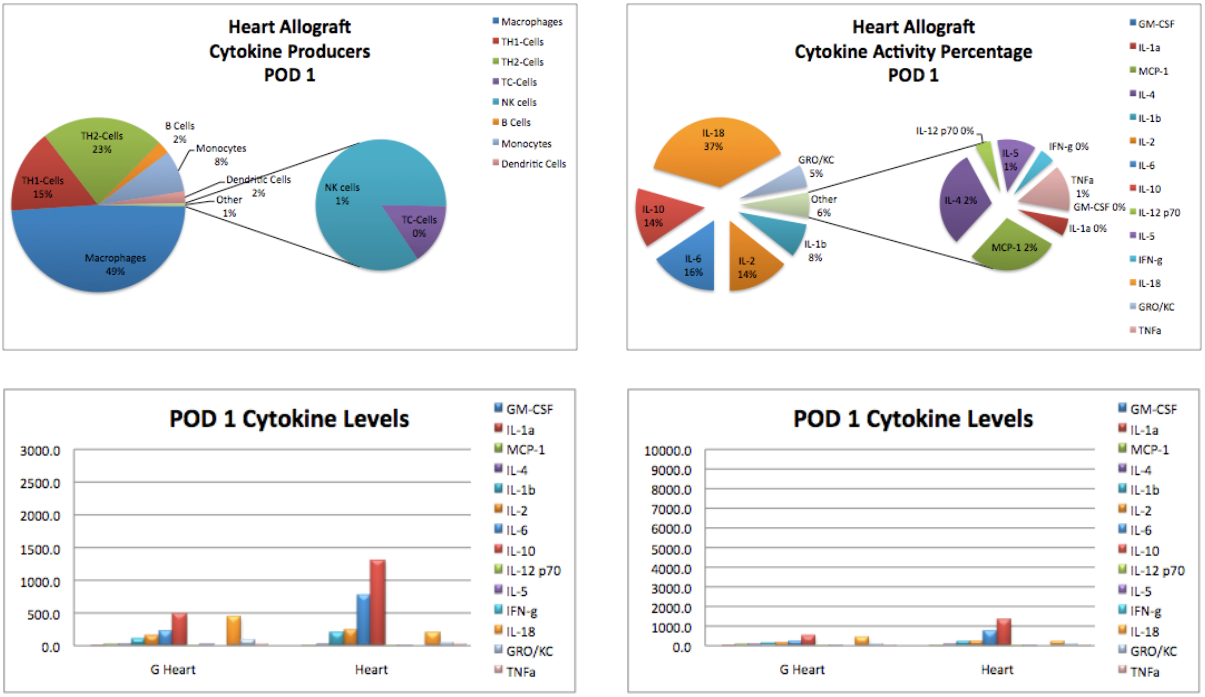

Figure 7: The cellulograph method of representing the cell population inferred by the cytokine milieu present in the sample.

To create this snapshot, the individual cytokine activity percentages are calculated and each of the 14 measured cytokines is associated with the specific cell type(s) that are known to emit them. For cases where more than one emitting cell type is known, we assume independent identical distribution (IID) of emission, except in cases where rate of emission actually is known. This method produces the cellulograph panel.

We have identified three apparent stages of rejection, labeled here as innate response, adaptive response, and end-stage response (Figure 8). These stages appear in both VCA and solid organ contexts. Importantly, the earliest phases of immune response appear to be driven by cytokines that are produced predominantly by macrophages or other monocytes, indicating that the innate response has a key role in the downstream adaptive immune response, although the exact nature of this interaction is not fully understood yet.

It is also evident from the data that there are tissue-specific immune signaling profiles at each stage of rejection. In Figure 9, data from the cardiac transplant model is shown to elucidate this point, because it is the model with tissue samples from the widest range of anatomical locations. However, the finding that each tissue has distinct cytokine profiles holds for all tissue analyzed, in all models. Additionally, patterns of cytokine expression can be found within each tissue type.

To further elucidate the underlying drivers of the observed patterns in immune signaling, and to enhance discriminative classifier performance, feature transformation methods such as Principle Components Analysis (PCA) were applied to the data and analyzed. As a result, key immune signaling molecules were identified in each tissue and model that has a role in governing the local immune response (Figures 10, 11). Each model and tissue reveals a unique set of governing cytokines.

PCA allows the cytokines that contribute the most observed variance to the immunologic process to be identified. Further, we can understand what immune signaling components are exerting influence on other components, and how strong this influence is through PCA. By combining this information with a more detailed analysis of the underlying biological mechanisms of each of these cytokines, and the cells that produce them, we can derive a narrative that explains why the local cytokine milieu takes on the particular form observed in each tissue and immune response. We call this process biological function analysis and use it as a method of disambiguating cytokine function in context. It also plays a role in identifying parameters to incorporate into future model development.

This approach elucidated new information about the distinct immune signaling patterns present in different tissues at different time points. We developed a method to unify this information under a single analytic framework that allows development of accurate tissue specific descriptive and discriminative models.

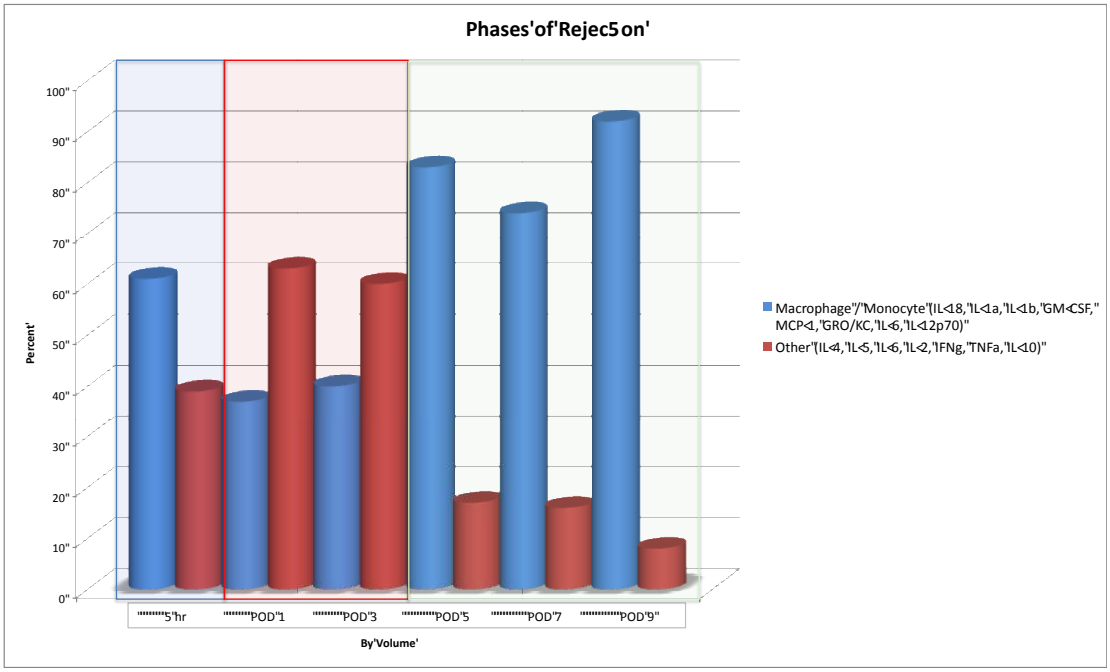

Figure 8: Three phases are evident in the course of cardiac transplant rejection, as well as in other measured transplant models.

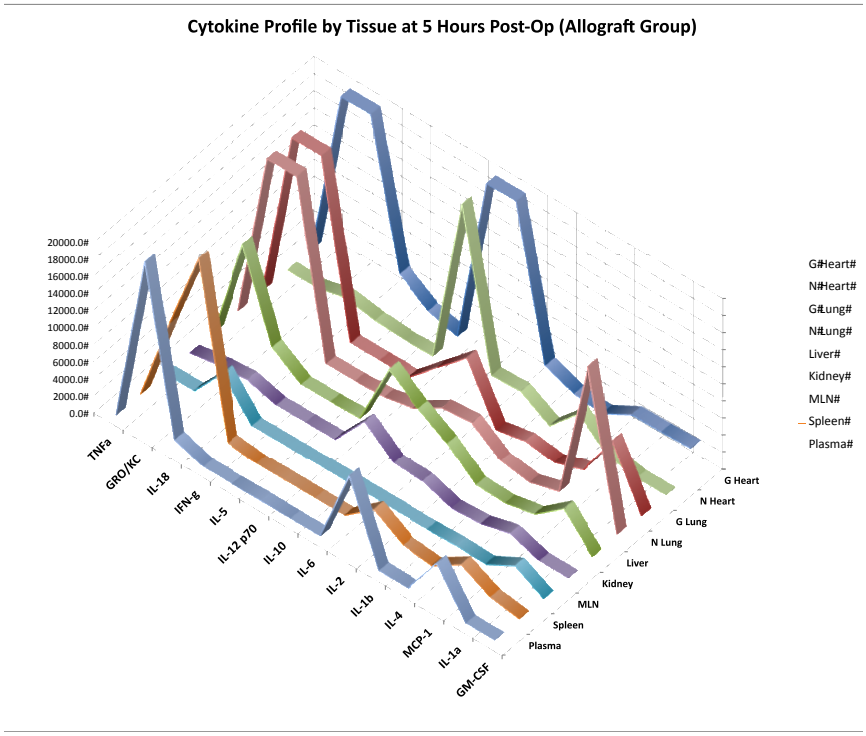

Figure 9: Distinct cytokine network profiles are evident for each tissue in this graph of the cardiac transplant model at 5 hours post op.

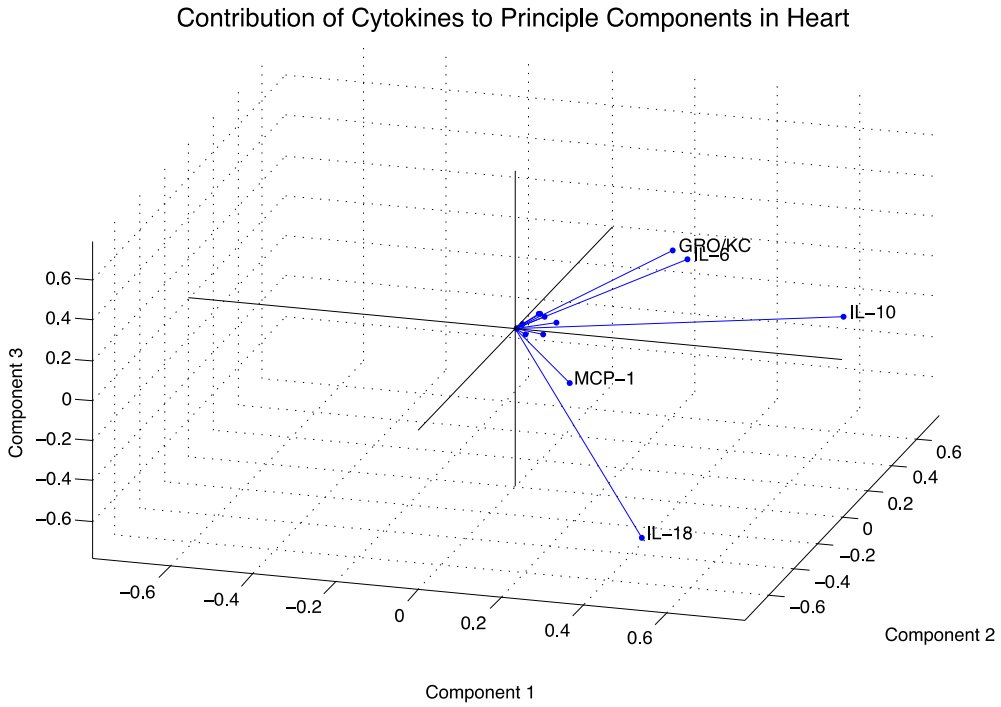

Figure 10: Contribution of cytokines to each of the three principle components in the graft heart from the cardiac transplant model. Three principle components explain >95% of variance. Length and direction of vectors indicates contribution to principle components.

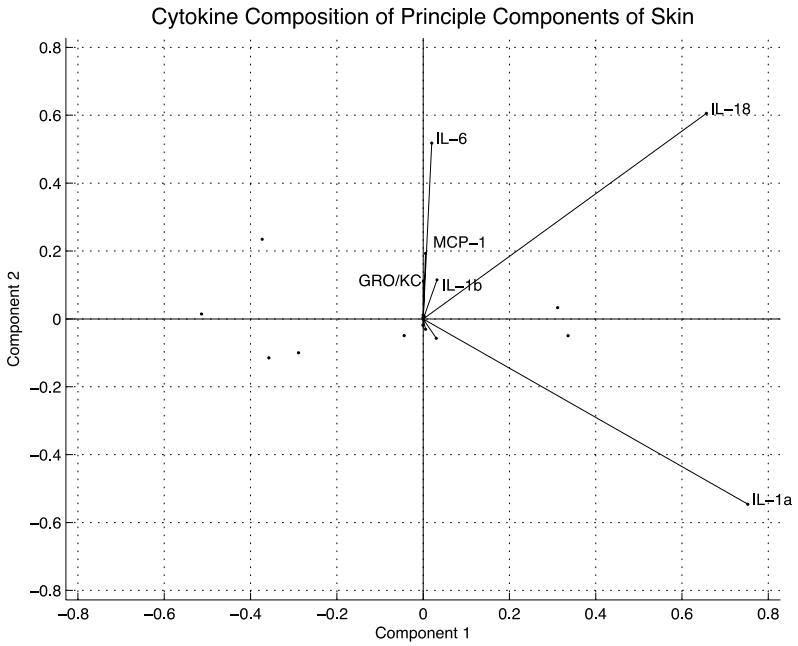

Figure 11: Contribution of cytokines to each of the two principle components in the graft skin from the skin wound model. Two principle components explain >95% of variance. Length and direction of vectors indicates contribution to principle components.

Detection of Early Rejection in the Presence of Immunosuppression

Focusing on further developing the model for differentiation of rejecting tissue from non-rejecting tissue, a robust computational model was created that is able to correctly identify rejecting tissue from non-rejecting tissue in the skin and muscle of rat hind limb vascularized composite allografts (VCA), both with and without immunosuppressive treatment. This model, and the methodology used to construct it, provide a potential platform for generalization of the findings of this project to other inflammatory conditions as well as a key step in further establishing clinical relevance.

Group	Procedure	N=
Naïve controls	None	9 LEW, 9 BN
Isograft	LEW-LEW Hind-Limb Transplantation	24 LEW Recipients, 12 LEW Donors
Allograft	BN-LEW Hind-Limb Transplantation	24 LEW Recipients, 12 BN Donors

Table 3: Groups in the non-invasive sampling investigation cohort

This cohort was designed to provide a large amount of cytokine signaling data about the distinguishing signaling characteristics of allogeneic, syngeneic, naive, and immunosuppressed hind limb transplants. In untreated skin the adaptive immune response

takes approximately two to three days to develop and the course of rejection runs over approximately 11 days, so samples were taken at post-operative days 3,5,7,9, and 11 in the allogeneic and syngeneic groups. At each time point, biopsy samples were taken from both skin and muscle. Surgery and sampling in collaboration with surgeons from the University of Innsbruck School of Medicine, Johns Hopkins School of Medicine Department of Plastic Surgery, and University of Pittsburgh Medical Center Department of Plastic and Reconstructive Surgery.

In skin, high concentrations of IL-1a and IL-18 were seen in all groups (naïve, isograft, and allograft). Allografts display particularly elevated levels of IL-6, IL-1b, and GRO/KC. Larger variance and more outliers are also characteristics of the allograft group.

A different set of cytokines appears to be activated in the muscle than in skin, in all groups. There is little overlap with the cytokine profiles of skin, and there are clearly distinct cytokine profiles between each group. IL-18 or IL-1b also seem to play an important role in muscle. Similar to what is seen in skin, higher levels of variance in cytokine levels are also seen in the allograft muscle than in the naïve or isograft muscle.

It seems clear from this data that raw cytokine levels hint at important differences between the groups, but does not directly illustrate a naïve or linear separation amongst them. In order to achieve our objective, we turned to more advanced methods of computational analysis and statistical inference modeling.

The matrix of the original 14-cytokine features was extended in both skin and muscle data sets by adding feature interaction variables. The selection of these features was driven by the intuition that feature interactions which are present in only one group provides additional information that will help separation of groups and consequently improvement in classifier performance. For each tissue, the matrix of r-values for each group was subtracted from each other. For example, the matrix of r-values for the group “rejecting” was subtracted from the matrix of r-values for “not rejecting.” This yielded a new 14 x 14 square matrix that represents pairwise correlation between features present in one group but not both. Feature pairs with higher values (either positive or negative) are more desirable. This process was performed for both muscle and skin data. In skin the feature interactions selected by r-value comparison were GM-CSF*TNFa and IL-2*TNFa, while in muscle they were MCP-1*IL-18 and GRO/KC*IL-18. A random forest classifier was then applied to this extended dataset.

The random forest is an ensemble classifier that aggregates many individual decision trees and makes a classification that is the mode of the classes output by the ensemble of decision trees. Each tree in the ensemble is grown to make a decision on a subset of the data set features, and a training set for the tree is selected by taking a bootstrap sample. The remaining observations are used to estimate the error of the tree. For each subsequent node in the tree, randomly choose another subset of features on which to base the next split. This is continued until the tree is fully grown. This method combines bootstrap aggregation (bagging), and random feature selection. It is capable of very high classification accuracy in non-linear problem spaces.

The random forest method is able to classify rejecting (allograft) versus non-rejecting (isograft) tissue with 96.15% accuracy in skin ($p<0.01$, Table 4 and 95.16% accuracy in muscle ($p<0.05$, Table 5) at all time points. For time points at or prior to POD 5 accuracy in skin is 93.41% ($p<0.05$), and in muscle is 98.68% ($p<0.01$). Performance was calculated utilizing 10-fold cross-validation, with statistical significance calculated by Welch's t-test.

When performance was measured by measuring the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUROC), the random forest classifier achieved an AUROC score of >0.994 in skin (Figure 12) and an AUROC score of >0.989 in muscle (Figure 13).

VCA Allograft vs Isograft in Skin Random Forest (all time points, hybrid features, 50 trees)

Accuracy	**Confusion Matrix**			
96.15%	*Classified As*			*True Class*
p-value	Allograft	Isograft		
.0002	74	3	Allograft	
	3	76	Isograft	

Table 4 Random forest classifier performance in skin

VCA Allograft vs Isograft in Muscle Random Forest (all time points, hybrid features, 50 trees)

Accuracy	**Confusion Matrix**			
95.16%	*Classified As*			*True Class*
p-value	Allograft	Isograft		
.0352	48	5	Allograft	
	2	69	Isograft	

Table 5 Random forest classifier performance in muscle

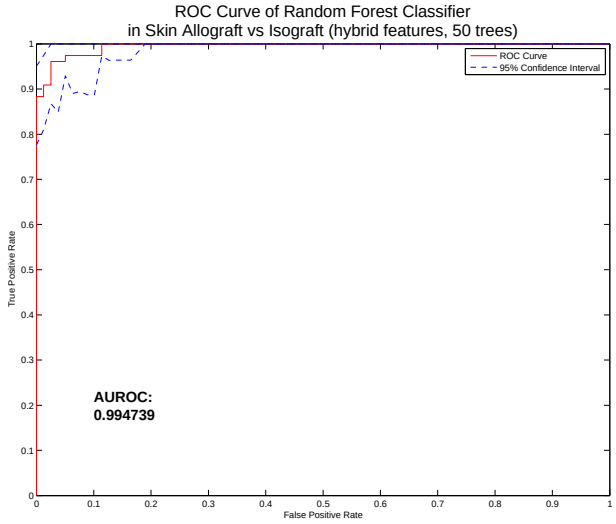

Figure 12: Skin Allograft vs. Isograft

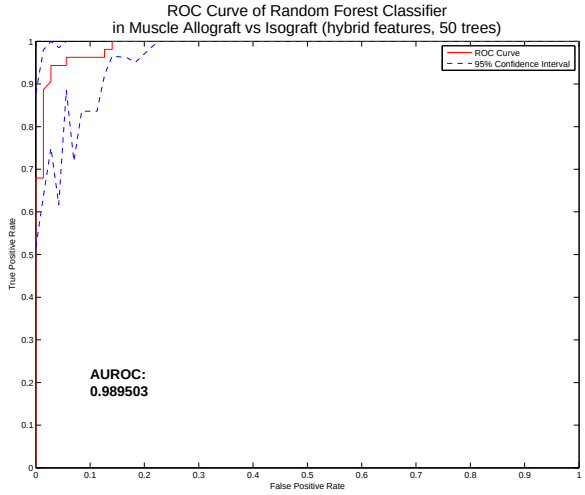

Figure 13: Muscle Allograft vs. Isograft

Towards a Non-Invasive Method of Sampling

While the ability to accurately classify rejection state by cytokine profiles (or immune signaling) is a significant advance, one of the primary hurdles to the adoption of cytokine-based immune monitoring in VCA transplantation is the current requirement to sample skin tissue by biopsy. The cutting associated with a taking a biopsy causes inflammation in the allograft, and inherently carries risks of infection and rejection. Therefore, although cytokine-based immune monitoring is a significant advance in rejection early warning, biopsy-dependent application of the technology would be limited to analysis of routine program biopsies. This project also demonstrated the accuracy of cytokine-based immune monitoring in VCA with a non-invasive tape stripping method. This method does not require biopsies to be taken for cytokine level analysis of the skin, and would circumvent much of the risk associated with frequent biopsy sampling of allografts.

Collaborating surgeons at Johns Hopkins University supplied raw samples from hind-limb transplant surgeries to collaborating researchers at the University of Pittsburgh, where cytokine levels were read by Luminex 100 IS machine. The raw quantification data was then sent to this project's investigators, where it was analyzed for patterns consistent or correlating with previously reported findings. Animal surgeries and cytokine quantification for this aspect of the project was made possible by our research collaborators (not funded by this project) who saw the value in demonstrating the non-invasive immune monitoring concept and had other funds available to conduct the work.

There were eight Isograft (ISO) and eight Allograft (ALLO) complete hind-limb transplants conducted with Brown Norway (BN) rat recipients and Lewis (LEW) rat donors. These animals were selected because they are a complete Major Histocompatibility Complex (MHC) mismatch, ensuring rejection. In each sampling, a biopsy was taken as a control, a sebutape sample was taken (non-invasive), and a D-Squame sample was taken (non-invasive). To accommodate sampling with both sebutape and d-squame at the same time point, we sampled from the lateral and the interior of the thigh, to give maximum sampling surface area. D-squame and sebutape sampling were alternated at each time point (e.g. lateral - interior - lateral).

A serial sampling method was used to eliminate excessive dead skin or detritus on the outermost layers of skin. With D-Squame 15 consecutive sampling patches were applied and collected as per manufacturer protocol. With Sebutape, 5 consecutive sampling patches were applied by gloved hand and collected as per manufacturer protocol.

After patches were collected and prior to diluting proteins off the surface, protein levels were measured with the CUDerm 850A scanning instrument. Patches with high protein loads were then diluted into solution using PBS buffer wash. This solution was then processed with BCA assay, and samples with a high protein load were processed by Luminex to quantify individual cytokine levels.

The cytokine patterns associated with allogeneic transplants can be differentiated from those associated with syngeneic transplants across all time points with high specificity and sensitivity. Additional pattern-recognition analyses were able to further elucidate distinctive patterns among cytokine profiles originating in skin biopsy samples versus those obtained from allogeneic or syngeneic transplants. The analysis is done by transforming the data using the robust MANOVA methodology and then utilizing the transformed data as the subject of analysis for the random forest classifier algorithm (growing the forest to a size of 50 trees). Performance of the classifier was then evaluated through leave-on-out cross validation (jack-knifing), and comparison of the confusion matrix across all runs. This preliminary analysis demonstrates the accuracy of this classifier to be 85.7143% with a p value of 0.002303 for differentiating allograft from isograft across all the time points using the samples collected by non-invasive method. The Receiver Operating Characteristics (ROC) of this method (Figure 14) shows a desirable relationship between specificity and sensitivity. In a ROC curve the true positive rate (sensitivity) is plotted in function of the false positive rate (specificity) for different cut-off points of a diagnostic test. The area under the ROC curve (AUROC) is 0.874, with 95% confidence interval

extending to 0.791 and 0.937, delivering evidence that this test does have an ability to distinguish between the allogeneic and syngeneic groups with a clinically desirable relationship between specificity and sensitivity.

The work from this project has demonstrated that cytokine signaling patterns can foretell rejection in both VCA and solid organ contexts with biopsy sampling, can provide insight into the underlying factors that drive the immune response, and has demonstrated the viability of non-invasive immune-monitoring by tape stripping.

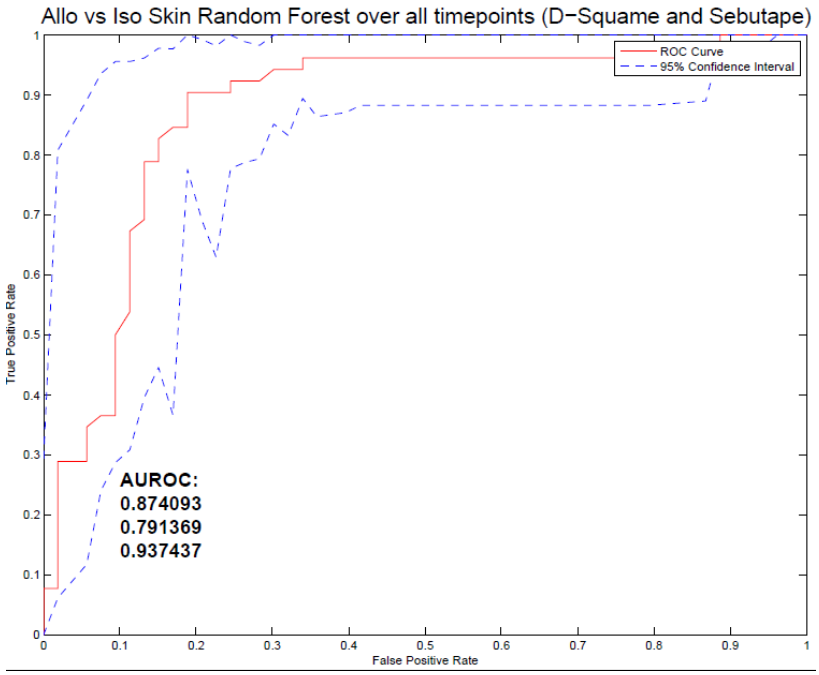

Figure 14: ROC of the classifier

Conferences and invited talks:

1. Wolfram D, Starzl R, Barclay D, Vodovotz Y, Wang Y, Brandacher G, Lee WPA, Schneeberger S. *Profiling gene expression and cytokine networks mediating acute rejection after composite tissue allotransplantation.* Plastic Surgery Research Council 55th annual meeting. May 2010. Plenary Talk.
2. Starzl R, Brandacher G, Wolfram D, Barclay D, Vodovotz Y, Wang Y, Gorantla V, Pratschke J, Lee WPA, Schneeberger S. *Analysis of genetic and proteomic factors of the immune response mediating cellular infiltration and trafficking in skin rejection of composite tissue allografts*. XXIII International Congress of The Transplantation Society. August 2010. Poster Presentation.
3. Starzl R, Wolfram D, Barclay D, Vodovotz Y, Wang Y, Brandacher G, Pratschke J, Lee WPA, Schneeberger S. *Allogeneic and syngeneic cytokine networks in skin muscle and composite tissue allografts*. XXIII International Congress of The Transplantation Society. August 2010. Poster Presentation.
4. Starzl R, Zhang W, Wang Y, Brandacher G, Schneeberger S, Gorantla V, Carbonell J, Barclay D, Vodovotz Y, Zheng XX. *Inflammatory Biomarker Patterns Identify Specific Types of Rat Skin Rejection*. 3rd International Conference on Transplantomics and Biomarkers in Organ Transplantation. March 2012. Poster Presentation.
5. Starzl R, Zhang W, Wang Y, Brandacher G, Schneeberger S, Gorantla V, Carbonell J, Barclay D, Vodovotz Y, Zheng XX. *Distinctive Immune Signaling Patterns and Factors in Rat Skin Rejection.* American Transplant Congress. June 2012. Poster Presentation.
6. Starzl R, Wolfram D, Zamora R, Jefferson B, Barclay D, Ho C, Brandacher G, Schneeberger S, Lee WPA, Carbonell J, Vodovotz Y. *Tissue-Specific Patterns of Caspase-1 and Cytokines in Excisional Wounds are Altered by Shock in Rat Skin and Muscle.* 11th International Conference on Complexity in Acute Illness. September 2012. Poster Presentation.

18. Extent of Clinical Activities Initiated and Completed. Items 18(A) and 18(B) should be completed for all research projects. If the project was restricted to secondary analysis of clinical data or data analysis of clinical research, then responses to 18(A) and 18(B) should be "No."

18(A) Did you initiate a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___x___No

18(B) Did you complete a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___x___No

If "Yes" to either 18(A) or 18(B), items 18(C) – (F) must also be completed. (Do NOT complete 18(C-F) if 18(A) and 18(B) are both "No.")

18(C) How many hospital and health care professionals were involved in the research project?

_____Number of hospital and health care professionals involved in the research project

18(D) How many subjects were included in the study compared to targeted goals?

_____Number of subjects originally targeted to be included in the study
_____Number of subjects enrolled in the study

Note: Studies that fall dramatically short on recruitment are encouraged to provide the details of their recruitment efforts in Item 17, Progress in Achieving Research Goals, Objectives and Aims. For example, the number of eligible subjects approached, the number that refused to participate and the reasons for refusal. Without this information it is difficult to discern whether eligibility criteria were too restrictive or the study simply did not appeal to subjects.

18(E) How many subjects were enrolled in the study by gender, ethnicity and race?

Gender:
_____Males
_____Females
_____Unknown

Ethnicity:
_____Latinos or Hispanics
_____Not Latinos or Hispanics
_____Unknown

Race:
_____American Indian or Alaska Native
_____Asian
_____Blacks or African American
_____Native Hawaiian or Other Pacific Islander
_____White
_____Other, specify:_____
_____Unknown

18(F) Where was the research study conducted? (List the county where the research study was conducted. If the treatment, prevention and diagnostic tests were offered in more than one county, list all of the counties where the research study was conducted.)

19. Human Embryonic Stem Cell Research. Item 19(A) should be completed for all research projects. If the research project involved human embryonic stem cells, items 19(B) and 19(C) must also be completed.

19(A) Did this project involve, in any capacity, human embryonic stem cells?
_____Yes
___x___ No

19(B) Were these stem cell lines NIH-approved lines that were derived outside of Pennsylvania?
_____Yes
_____ No

19(C) Please describe how this project involved human embryonic stem cells:

20. Articles Submitted to Peer-Reviewed Publications.

20(A) Identify all publications that resulted from the research performed during the funding period and that have been submitted to peer-reviewed publications. Do not list journal abstracts or presentations at professional meetings; abstract and meeting presentations should be listed at the end of item 17. **Include only those publications that acknowledge the Pennsylvania Department of Health as a funding source** (as required in the grant agreement). List the title of the journal article, the authors, the name of the peer-reviewed publication, the month and year when it was submitted, and the status of publication (submitted for publication, accepted for publication or published.). Submit an electronic copy of each publication or paper submitted for publication, listed in the table, in a PDF version 5.0.5 (or greater) format, 1,200 dpi. Filenames for each publication should include the number of the research project, the last name of the PI, and an abbreviated title of the publication. For example, if you submit two publications for Smith (PI for Project 01), one publication for Zhang (PI for Project 03), and one publication for Bates (PI for Project 04), the filenames would be:

Project 01 – Smith – Three cases of isolated
Project 01 – Smith – Investigation of NEB1 deletions
Project 03 – Zhang – Molecular profiling of aromatase
Project 04 – Bates – Neonatal intensive care

If the publication is not available electronically, provide 5 paper copies of the publication.

Note: The grant agreement requires that recipients acknowledge the Pennsylvania Department of Health funding in all publications. Please ensure that all publications listed acknowledge the Department of Health funding. If a publication does not acknowledge the funding from the Commonwealth, do not list the publication.

Title of Journal Article:	Authors:	Name of Peer-reviewed Publication:	Month and Year Submitted:	Publication Status (check appropriate box

				below):
1.				☐Submitted ☐Accepted ☐Published
2.				☐Submitted ☐Accepted ☐Published
3.				☐Submitted ☐Accepted ☐Published

20(B) Based on this project, are you planning to submit articles to peer-reviewed publications in the future?

Yes___x_____ No_____

If yes, please describe your plans:

We plan on submissions to immunology journals.

21. Changes in Outcome, Impact and Effectiveness Attributable to the Research Project. Describe the outcome, impact, and effectiveness of the research project by summarizing its impact on the incidence of disease, death from disease, stage of disease at time of diagnosis, or other relevant measures of outcome, impact or effectiveness of the research project. If there were no changes, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None

22. Major Discoveries, New Drugs, and New Approaches for Prevention Diagnosis and Treatment. Describe major discoveries, new drugs, and new approaches for prevention, diagnosis and treatment that are attributable to the completed research project. If there were no major discoveries, drugs or approaches, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

None

23. Inventions, Patents and Commercial Development Opportunities.

23(A) Were any inventions, which may be patentable or otherwise protectable under Title 35 of the United States Code, conceived or first actually reduced to practice in the performance of work under this health research grant? Yes_____ No___x___

If "Yes" to 23(A), complete items a – g below for each invention. (Do NOT complete items a - g if 23(A) is "No.")

a. Title of Invention:

b. Name of Inventor(s):

c. Technical Description of Invention (describe nature, purpose, operation and physical, chemical, biological or electrical characteristics of the invention):

d. Was a patent filed for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____

If yes, indicate date patent was filed:

e. Was a patent issued for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes_____ No____
If yes, indicate number of patent, title and date issued:
Patent number:
Title of patent:
Date issued:

f. Were any licenses granted for the patent obtained as a result of work performed under this health research grant? Yes_____ No____

If yes, how many licenses were granted?_____

g. Were any commercial development activities taken to develop the invention into a commercial product or service for manufacture or sale? Yes___ No____

If yes, describe the commercial development activities:

23(B) Based on the results of this project, are you planning to file for any licenses or patents, or undertake any commercial development opportunities in the future?

Yes_____ No_____x_____

If yes, please describe your plans:

24. Key Investigator Qualifications. Briefly describe the education, research interests and experience and professional commitments of the Principal Investigator and all other key investigators. In place of narrative you may insert the NIH biosketch form here; however, please limit each biosketch to 1-2 pages. *For Nonformula grants only – include*

information for only those key investigators whose biosketches were not included in the original grant application.

BIOGRAPHICAL SKETCH

NAME Jaime G. Carbonell	POSITION TITLE Professor & Director of the Language Technologies Institute
eRA COMMONS USER NAME JCARBONELL	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Massachusetts Institute of Technology	S.B.	1975	Mathematics
Massachusetts Institute of Technology	S.B.	1975	Physics
Yale University	M.S.	1976	Computer Science
Yale University	Ph.D.	1979	Computer Science

A. Personal Statement

Jaime Carbonell is an expert in machine learning (co-founded the field circa 1980, and his recent focus is on active and proactive learning), in language technologies, and in computational proteomics. He has worked on computational predictive models for tertiary and quaternary protein structure on with collaborators at CMU and U Pitt, inducing within-species and cross-species protein-protein interactomes

B. Positions and Honors

Positions and Employment

1996 - Present Director, Language Technologies Institute, Carnegie Mellon University
1995 - Present Allen Newell Professor of Computer Science, Carnegie Mellon University
2006 - Present Adjunct Faculty, Dept of Computational Biology, U Pittsburgh Medical School.
2001 - Present Co-Founder, Board chairman, Carnegie Speech Incorporated
1995 - 2003 Co-Founder, Board chairman, Wisdom Technologies Corporation
1986 - 1996 Director, Center for Machine Translation, Carnegie Mellon University
1987 - 1995 Professor of Computer Science, Carnegie Mellon University
1983 - 1987 Associate Professor of Computer Science, Carnegie Mellon University
1983 - 1996 Co-Founder, Director, Scientific advisor, Carnegie Group Inc.
1979 - 1983 Assistant Professor of Computer Science, Carnegie Mellon University
1975 - 1978 Research Fellow in Computer Science, Yale University
1974 - 1975 Research Assistant, Center for Space Research, MIT
1971 - 1975 Research Programmer, Artificial Intelligence, BBN, Cambridge, MA

Professional Organizations and Activities

Member: ACM (elected Chair of SIGART 1983-85), ACL, AAAI (AAAI Fellow 1988-present, AAAI executive committee 1990-92), Cognitive Science Society, Sigma Xi, AMTA.

Government Committees: NIH Human Genome Scientific Advisory Committee, aka "Watson Committee" (1988-1992). DFKI (National German AI Lab) Scientific Advisory Board (1988-2003). National Institute of Standards and Technology (NIST/IAD), Scientific Advisory Committee (1997-2001), DOE Oakridge National Laboratories Scientific Advisory Committee (1985-1987), GMD/IPSI Information Sciences (Germany) Scientific Advisory board (1990-2001), Citigroup Technology Advisory committee (1987-1995).

Book Series co-Editor: *Lecture Notes in Artificial Intelligence* North American editor, Springer 1995-2008.

Editorial Boards: Machine Learning Journal (1984-2000, editor-in-chief 1988-1993), Machine Translation Journal (1980's), Artificial Intelligence Journal (1984-2007).

Inventor: Interlingua Machine Translation, Maximal-Marginal-Relevance algorithm, Seeded Version Space Learning algorithm, Radial-Density-Differential method for novelty detection, Proactive Machine Learning, Linked-Segmented-CRF method for protein structural induction (co-inventor).

Honors

Named "University Professor" at Carnegie Mellon University as of 2012

Over 300 invited talks and key-note addresses between 1987 and 2012
IJCAI Best Paper Award in 1997 for Translingual Information Retrieval
Elected AAAI fellow in 1988 for scientific accomplishments
Received the Herbert A Simon Computer Science teaching award in 1987
Sperry fellowship for research excellence in 1986
"Recognition of service" award in 1985 from the ACM for the SIGART presidency
Received multiple unsolicited industrial gifts to support research (Hughes, Hitachi, IBM, Alcoa, …)

C. Selected recent peer-reviewed publications (from more than 300 publications)

1. Yang, L., Hanneke, S., Carbonell, J. "The Identifiability of Priors from Bounded Sample Sizes with Applications to Transfer Learning", *Proc of the 24th Annual Conference on Learning Theory* (COLT), 2011.
2. Balakrishnan S., Tastan O., Carbonell J., and Klein-Seetharaman J., "Alternative Paths in HIV-1 Targeted Human Signal Transduction Pathways." BMC Genomics, 2010
3. He, J. and Carbonell, J. "Prior-Free Rare Category Detection" Proceedings of the SIAM International Conference on Data Mining (SDM09), 2009.
4. Tastan O, Qi Y, Carbonell, JG, and Klein-Seetharaman, J. "Prediction of Interactions between HIV-1 and Human Proteins by Information Integration." In proceedings of the Pacific Symposium on Biocomputing 14:516-527, 2009.
5. Donmez, P., Carbonell, J. "Proactive Learning: Cost-Sensitive Active Learning with Multiple Imperfect Oracle", iIn Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM), Napa Valley 2008.
6. Liu, Y. Carbonell, J., and Gopalakrishnan, V. "Protein Quaternary Fold Recognition Using Conditional Graphical Models." In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07).* Hyderabad, India. 2007.
7. Carbonell, J., Klein, S., Miller, Steinbaum, M., D., Grassiany, T., and Frey, J., "Context-Based Machine Translation" In *Proceedings of the Association for Machine Translation of the Americas (AMTA-2006)*, Boston, 2006.
8. Liu Y, Carbonell J, Weigele P, Gopalakrishnan. V, ''Segmentation Conditional Random Fields (SCRFs): A New Approach for Protein Fold Recognition,'' *Journal of Computational Biology*, Vol. 13, No. 2, 2006. [Longer version of RECOMB-05 paper.]

D. Research Support

Ongoing Research Support

MURI - US Army Carbonell (PI) 09/30/2010 to 09/29/2015
The Linguistic Core Approach to Structured Translation and Analysis of Low-Resource Languages
We research structurally-driven machine translation and text analysis for typologically-diverse low-resource languages. Our approach is to design and construct a linguistic core for each identified low-resource language, design and implement novel core computational methods and algorithms, and design and implement prototype MT and TA systems based on the new linguistic cores and algorithmic advances.
Role: PI

DARPA Hovy (PI)
09/27/2012 - 03/27/2017
Structured Distributed Semantics: Analysis and Filtering of Text
This research involves discovering anomalous events based on analysis of semantic relations, as reported in temporal streams.
Role: Co-PI

NSF Carbonell (PI)
09/01/2011 to 08/31/2014
RI: Medium: Interactive Transfer Learning in Dynamic Environments
The aim of this project is to bring richer interaction into the realm of machine learning by developing frameworks as well as machine learning methods that take advantage of fuller mixed-initiative communication.
Role: PI