



THE CENTER FOR BIOMOLECULAR SCIENCE & ENGINEERING  
ENGINEERING BUILDING 2, ROOM 501  
SANTA CRUZ, CALIFORNIA 95064

(831) 459-1477  
(831) 459-1809 FAX

June 9, 2014

Dr. Simona Volpi, Program Director  
Genotype-Tissue Expression Project (GTEx)  
Division of Extramural Research  
5635 Fishers Ln, Room 4107  
Rockville, MD 20892

Dear Dr. Volpi,

In response to our recent communication regarding integration of GTEx data into the UCSC Genome Browser and the development of visualization for this, we provide here a proposal for a three year supplement to our Browser award (grant # 5U41HG002371-14).

We have also provided a detailed budget and justification, and a checklist with details on indirect costs (the university rate for this project is 53%, 53.5% and 54% successively in years 1-3 of the supplement).

We are requesting direct costs of \$710,655 and total costs of \$1,090,942.

Please let us know if you have any questions or need anything further for this proposal. We are extremely excited to be able to expand our browser to encompass the GTEx data.

Sincerely,

A handwritten signature in black ink, appearing to read "W. James Kent".

W. James Kent  
Director, UCSC Genome Browser  
Center for Biomolecular Science and Engineering  
UC Santa Cruz

Department of Health and Human Services Public Health Services  <b>Grant Application</b>  <i>Do not exceed character length restrictions indicated.</i>	<b>LEAVE BLANK—FOR PHS USE ONLY.</b>									
	<table border="1" style="width:100%; border-collapse: collapse;"> <tr> <td style="width:33%;">Type</td> <td style="width:33%;">Activity</td> <td style="width:34%;">Number</td> </tr> <tr> <td>Review Group</td> <td></td> <td>Formerly</td> </tr> <tr> <td>Council/Board (Month, Year)</td> <td></td> <td>Date Received</td> </tr> </table>	Type	Activity	Number	Review Group		Formerly	Council/Board (Month, Year)		Date Received
Type	Activity	Number								
Review Group		Formerly								
Council/Board (Month, Year)		Date Received								

1. TITLE OF PROJECT (Do not exceed 81 characters, including spaces and punctuation.)  
**The UCSC Genome Browser (GTEEx Supplement)**

2. RESPONSE TO SPECIFIC REQUEST FOR APPLICATIONS OR PROGRAM ANNOUNCEMENT OR SOLICITATION  NO  YES  
 (If "Yes," state number and title)  
 Number: PAR-11-095 Title: Genomic Resource Grants for Community Resource Projects (U41)

**3. PROGRAM DIRECTOR/PRINCIPAL INVESTIGATOR**

3a. NAME (Last, first, middle) Kent, W. James	3b. DEGREE(S) BA MA PhD	3h. eRA Commons User Name JKENT123
3c. POSITION TITLE Research Scientist, Director	3d. MAILING ADDRESS (Street, city, state, zip code) MS: CBSE/ITI 1156 High Street Santa Cruz, CA 95064	
3e. DEPARTMENT, SERVICE, LABORATORY, OR EQUIVALENT Center for Biomolecular Science & Engineering		
3f. MAJOR SUBDIVISION Baskin School of Engineering		
3g. TELEPHONE AND FAX (Area code, number and extension) TEL: 831-459-1401 FAX: 831-459-1809		E-MAIL ADDRESS: kent@soe.ucsc.edu

4. HUMAN SUBJECTS RESEARCH <input type="checkbox"/> No <input checked="" type="checkbox"/> Yes	4a. Research Exempt <input type="checkbox"/> No <input checked="" type="checkbox"/> Yes	If "Yes," Exemption No. 4
---	--	------------------------------

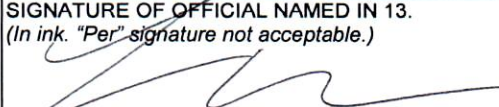
4b. Federal-Wide Assurance No.	4c. Clinical Trial <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes	4d. NIH-defined Phase III Clinical Trial <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes
--------------------------------	---	---

5. VERTEBRATE ANIMALS <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes	5a. Animal Welfare Assurance No.
---	----------------------------------

6. DATES OF PROPOSED PERIOD OF SUPPORT (month, day, year—MM/DD/YY) From 7/1/14 Through 6/30/17	7. COSTS REQUESTED FOR INITIAL BUDGET PERIOD 7a. Direct Costs (\$) \$228,659	8. COSTS REQUESTED FOR PROPOSED PERIOD OF SUPPORT 7b. Total Costs (\$) \$349,848 8a. Direct Costs (\$) \$710,655 8b. Total Costs (\$) \$1,090,942	
---	---	--	--

9. APPLICANT ORGANIZATION Name Regents of the University of California Address 1156 High Street Santa Cruz, CA 95064	10. TYPE OF ORGANIZATION Public: <input checked="" type="checkbox"/> Federal <input type="checkbox"/> State <input type="checkbox"/> Local Private: <input type="checkbox"/> Private Nonprofit For-profit: <input type="checkbox"/> General <input type="checkbox"/> Small Business <input type="checkbox"/> Woman-owned <input type="checkbox"/> Socially and Economically Disadvantaged
11. ENTITY IDENTIFICATION NUMBER 194-1539563A1 DUNS NO. 125084723 Cong. District CA-017	

12. ADMINISTRATIVE OFFICIAL TO BE NOTIFIED IF AWARD IS MADE Name Deirdre Beach Title Contract and Grant Officer Address 1156 High Street Office of Sponsored Projects Santa Cruz, CA 95064-1077 Tel: (831) 459-2778 FAX: E-Mail: dbeach@ucsc.edu	13. OFFICIAL SIGNING FOR APPLICANT ORGANIZATION Name Lynn Crowder Title Proposal Analyst Address 1156 High Street Office of Sponsored Projects Santa Cruz, CA 95064-1077 Tel: (831) 459-1378 FAX: E-Mail: lcrowder@ucsc.edu
---	--

14. APPLICANT ORGANIZATION CERTIFICATION AND ACCEPTANCE: I certify that the statements herein are true, complete and accurate to the best of my knowledge, and accept the obligation to comply with Public Health Services terms and conditions if a grant is awarded as a result of this application. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties.	SIGNATURE OF OFFICIAL NAMED IN 13. (In ink. "Per" signature not acceptable.) 	DATE 6/9/14
--	---	----------------

## Project Summary

In this supplement we propose one new specific aim – integrating the GTEx data into the UCSC Genome Browser. The NIH Common Fund's Genotype-Tissue Expression (GTEx) program measures genotype and gene expression in multiple tissues for many human donors. It has already released a substantial pilot data set at <http://www.gtexportal.org>, and is working on a larger production data set. We propose to integrate the data into the UCSC Genome Browser starting with the pilot phase data in the first year of this supplement, and extending to production data sets in later years. The integration will include at least two new tracks in the Genome Browser, one that shows the tissue-by-tissue expression data associated with each gene, and another that shows significant allele-specific differences in expression. We will also add new columns for the GTEx data to the Gene Sorter, and new panels to the existing gene details display. As the project progresses we anticipate adding a “phenotype associated significant gene expression differences” track as well.

## Relevance

Gene RNA expression levels provide a detailed view of the inner state of a cell. Both genetically caused and other diseases are often associated with gene expression changes. RNA expression levels are helpful in understanding the normal as well as disease physiology of an organism. This work will thus help scientists develop better treatments for a broad range of diseases.

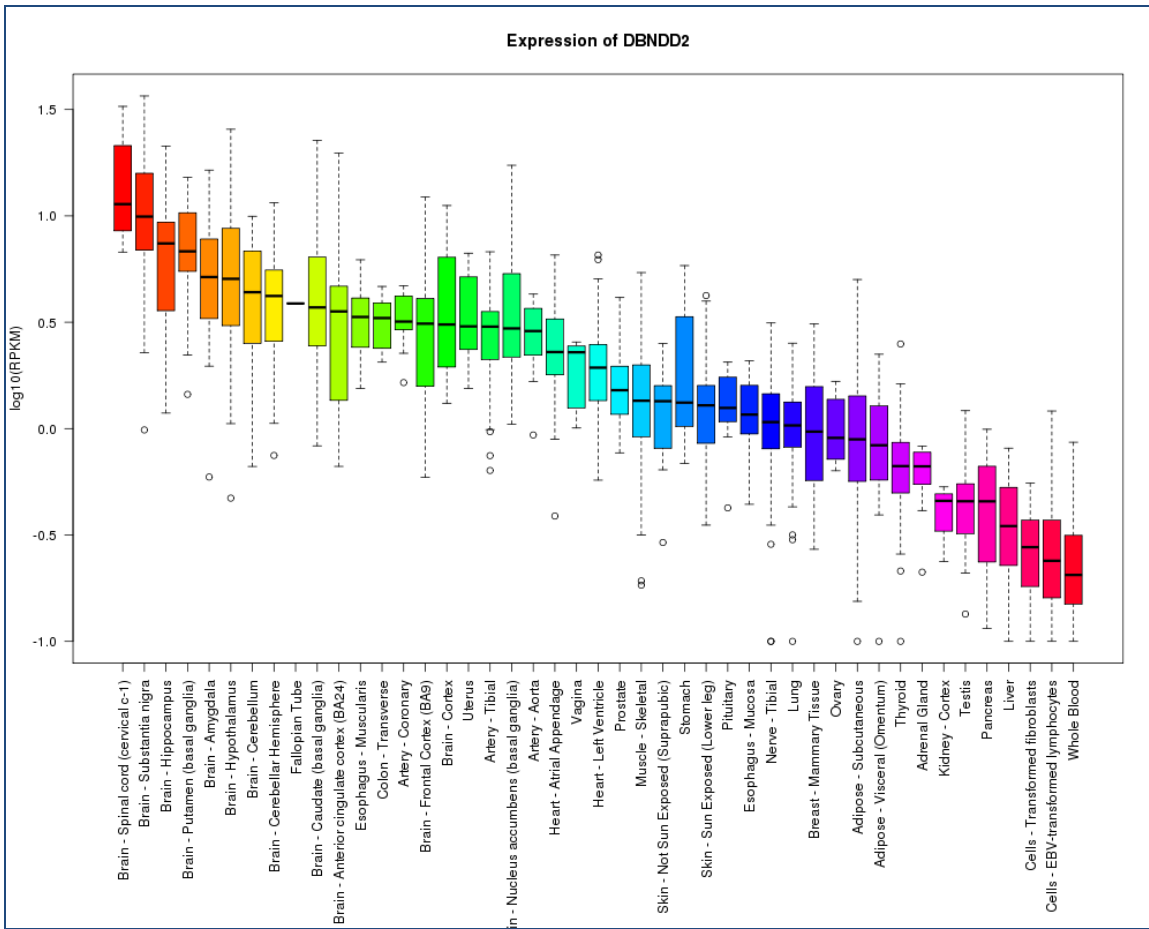
## Supplement to 5U41HG002371-14, UCSC Genome Browser: Integrate GTEx data into the UCSC Genome Browser

In this supplement we propose to integrate the GTEx data into the UCSC Genome Browser for visualization and analysis. This will make the carefully collected, genotyped, and analyzed human RNA expression data from GTEx more accessible by the research community. It will also result in new display capabilities in the Genome Browser that may be useful for other data sets as well. We will use these capabilities to display allele-specific expression by SNPs that show significant eQTLs, as well as overall gene expression levels in a variety of tissues. The GTEx data will be displayed as annotation tracks aligned to the genome assembly, as part of the gene details page, and as columns in the Gene Sorter. We will update the GTEx data in the Genome Browser at least once yearly, and will synchronize new releases to match major data releases from GTEx when possible. UCSC will collaborate with the GTEx analysis team, integrating new analyses as well as new data, and advising on analysis methods when useful.

The pilot phase GTEx data is available now and includes RNA levels measured by a variety of methods on 45 tissues collected from 185 phenotypically normal donors, who were genotyped as well. Most of the tissues have samples from at least 20 different donors. This is enough to obtain very robust gene expression data compared to earlier projects, and in some cases to tie variations in gene expression to specific SNPs in a significant way as eQTLs. The GTEx project plans to extend these data, and to add data from donors with various phenotypes as well.

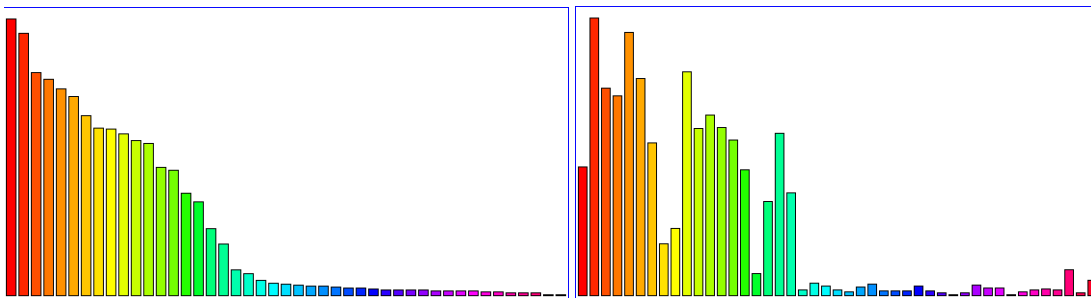
The first step in integrating the GTEx data into the browser will be to parse it, after which it will be formatted and indexed for fast display. Because the data are available and documented in reasonably simple formats at the GTEx portal, we anticipate that this step will take about 3 months. At the end of this step the pilot data will be made available in the UCSC Table Browser and on the Genome Browser download server.

Next we will develop a display for the expression information for a gene, similar to that shown in Figure 1. We will add this as a new section on the genes details page that displays when a user clicks on a gene in the Genome Browser. This step should take about 2 months.



**Figure 1.** Gene expression Tukey boxplots for a single gene from the GTEx Portal. The order shown in these plots is always from highest expressed to lowest expressed tissue.

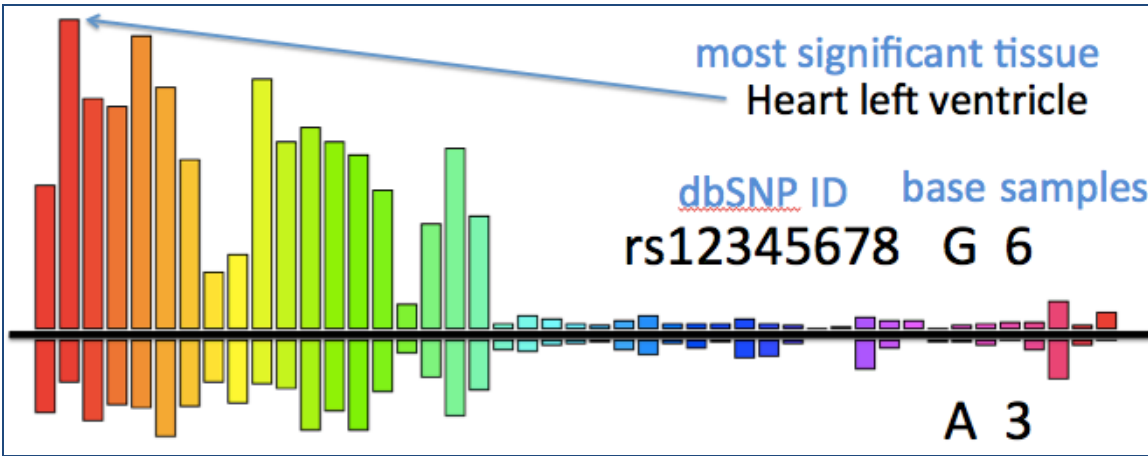
With this groundwork in place we will develop tracks for the Genome Browser. Since the browser displays data at many scales, from base level to complete chromosomes, the track graphics must work at many levels as well. For this reason the predominant view at the gene level will be a bar graph (Figure 2a), which is readable at smaller scales than the Tukey boxplots in Figure 1. Frequently users look at multiple genes simultaneously. Because of this we will keep a uniform, logarithmic-based scale for all genes, and we will order the tissues identically for all genes. The uniform ordering will be arranged such that related tissues are in close proximity (Figure 2b).



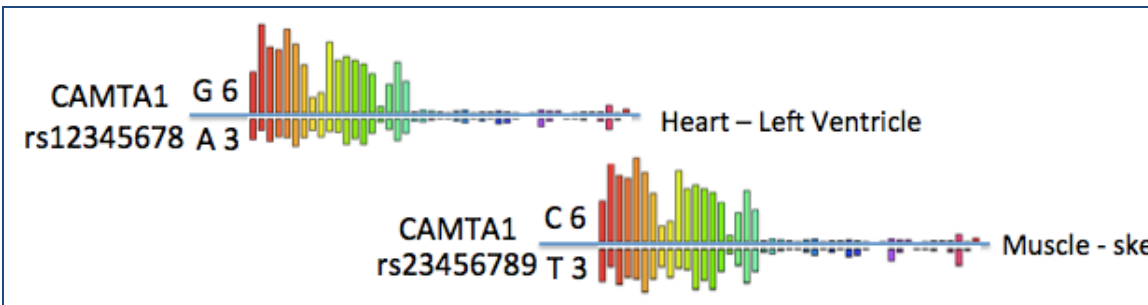
**Figures 2a & 2b.** Bar graphs sorted by gene expression (left (2a)) and by a canonical ordering of tissue types (right (2b)). Though some detail is lost, the bar graph is legible at a greater range of scales than the Tukey plot in Figure 1. The canonical ordering is preferred when expression patterns for multiple genes or multiple alleles of the same gene are shown together, since the two figures can be compared visually.

We will develop two Genome Browser tracks to represent the pilot data. One track will show an item for each gene, and will display a bar graph with a height of ~100 pixels at each gene for which there is GTEx data. The

second track will show an item for each SNP associated with an eQTL. It will display expression graphs for both alleles, one extending up and one extending down from a common axis (Figures 3 and 4). These new tracks will require 3-5 months to develop.



**Figure 3.** Mock-up of imagery and labeling for an eQTL SNP track item. Here the G allele would be associated with higher expression in general, and particularly higher on the biosample second from left.



**Figure 4.** Mock-up showing how two nearby eQTL SNPs might appear in a track. Each item would include the gene name, SNP name, tissue where most significant difference is observed, allele frequencies at that SNP for that tissue, and median expression graph for each allele. It is possible that we could include an overall p-val for the QTL in the display at this level. As with all our data tracks, more detail about an item would be available by clicking on it. This would include links back to the GTEx portal where appropriate.

We will finish the first year of the supplement and the integration of the pilot phase GTEx data by adding the data to the UCSC Gene Sorter. The Gene Sorter display is useful for visualizing expression levels on multiple genes at once, and for selecting gene sets by expression, GO annotation, and many other criteria.

In subsequent years of this project we will update the UCSC databases with new data from GTEx, and as necessary, develop new tracks and other new displays as novel types of data become available. We will answer user questions about the GTEx data, and when possible enhance the GTEx displays and analysis according to user suggestions, and according to feedback from the GTEx scientists.

<b>DETAILED BUDGET FOR INITIAL BUDGET PERIOD DIRECT COSTS ONLY</b>	FROM 7/1/14	THROUGH 6/30/17
--	----------------	--------------------

List PERSONNEL (*Applicant organization only*)  
 Use Cal, Acad, or Summer to Enter Months Devoted to Project  
 Enter Dollar Amounts Requested (*omit cents*) for Salary Requested and Fringe Benefits

NAME	ROLE ON PROJECT	Cal. Mnths	Acad. Mnths	Summer Mnths	INST.BASE SALARY	SALARY REQUESTED	FRINGE BENEFITS	TOTAL
Jim Kent	PD/PI	1			██████████	██████████	██████████	██████████
Kate Rosenbloom Pohl	Tech. Proj. Mngr	12			██████████	██████████	██████████	██████████
TBD	Quality Assurance	4			██████████	██████████	██████████	██████████
Ann Zweig	Eng & Proj Mngr	1			██████████	██████████	██████████	██████████
Donna Karolchik	Genome Browser Mngr	1			██████████	██████████	██████████	██████████
TBD	Grant Administratio	1			██████████	██████████	██████████	██████████
<b>SUBTOTALS</b> →						<b>154,784</b>	<b>64,375</b>	<b>219,159</b>

CONSULTANT COSTS	
EQUIPMENT ( <i>Itemize</i> )	
SUPPLIES ( <i>Itemize by category</i> ) Minor equipment (e.g. laptop workstation) - \$2,500; project-specific supplies, software - \$1,000	3,500
TRAVEL Travel to agency meetings and other project-related travel (4 trips at \$1,500 per trip)	6,000
INPATIENT CARE COSTS	
OUTPATIENT CARE COSTS	
ALTERATIONS AND RENOVATIONS ( <i>Itemize by category</i> )	
OTHER EXPENSES ( <i>Itemize by category</i> )	

CONSORTIUM/CONTRACTUAL COSTS	DIRECT COSTS	
<b>SUBTOTAL DIRECT COSTS FOR INITIAL BUDGET PERIOD</b> ( <i>Item 7a, Face Page</i> )		<b>\$ 228,659</b>
CONSORTIUM/CONTRACTUAL COSTS	FACILITIES AND ADMINISTRATIVE COSTS	
<b>TOTAL DIRECT COSTS FOR INITIAL BUDGET PERIOD</b>		<b>\$ 228,659</b>

**BUDGET FOR ENTIRE PROPOSED PROJECT PERIOD  
DIRECT COSTS ONLY**

BUDGET CATEGORY TOTALS	INITIAL BUDGET PERIOD <i>(from Form Page 4)</i>	2nd ADDITIONAL YEAR OF SUPPORT REQUESTED	3rd ADDITIONAL YEAR OF SUPPORT REQUESTED	4th ADDITIONAL YEAR OF SUPPORT REQUESTED	5th ADDITIONAL YEAR OF SUPPORT REQUESTED
PERSONNEL: <i>Salary and fringe benefits. Applicant organization only.</i>	219,159	228,923	239,073		
CONSULTANT COSTS					
EQUIPMENT					
SUPPLIES	3,500	1,000	1,000		
TRAVEL	6,000	6,000	6,000		
INPATIENT CARE COSTS					
OUTPATIENT CARE COSTS					
ALTERATIONS AND RENOVATIONS					
OTHER EXPENSES					
DIRECT CONSORTIUM/ CONTRACTUAL COSTS					
<b>SUBTOTAL DIRECT COSTS</b> <i>(Sum = Item 8a, Face Page)</i>					
F&A CONSORTIUM/ CONTRACTUAL COSTS					
<b>TOTAL DIRECT COSTS</b>	228,659	235,923	246,073		
<b>TOTAL DIRECT COSTS FOR ENTIRE PROPOSED PROJECT PERIOD</b>					<b>\$ 710,655</b>

JUSTIFICATION. Follow the budget justification instructions exactly. Use continuation pages as needed.

See continuation page

## Personnel

Salaries are projected to increase 3.5% annually.

PI W. James Kent (0.083 FTE; 1 month/year) – Design and prototype of software tools. Scientific oversight. Communicate with GTEx project and with funding agency.

Technical Project Manager Kate Rosenbloom Pohl (1.0 FTE; 12 months/year) – Refine design and implement production software associated with tracks and other features. Communicate with GTEx project.

QA TBD (0.33FTE; 4 months/year) – Test software, provide usability feedback, and help move software from development to test to production servers.

Engineering & Project Manager Ann Zweig (0.083 FTE; 1 month/year) – Project manager for genome browser. Tracks and coordinates work and release on new features and new data sets.

Genome Browser Manager Donna Karolchik (0.083 FTE; 1 month/year) – Provides personnel management for programming staff, helping communication, managing performance, and assembling yearly performance reviews.

Grant Administration TBD (0.083 FTE; 1 month/year) – Provide oversight and coordination of proposals, reporting, financial management, resources and personnel.

## Supplies

Funds are requested for a laptop workstation or other minor equipment in year 1 (\$2,500). An additional \$1,000 per year is requested each year for project-specific supplies and software.

## Travel

Funds are requested for 1-2 travelers to GTEx meetings and workshops in the Washington DC area (4 trips at \$1,500 each).



**CHECKLIST**

**TYPE OF APPLICATION** (Check all that apply.)

- NEW application. (This application is being submitted to the PHS for the first time.)
- RESUBMISSION of application number: \_\_\_\_\_  
(This application replaces a prior unfunded version of a new, renewal, or revision application.)
- RENEWAL of grant number: \_\_\_\_\_  
(This application is to extend a funded grant beyond its current project period.)
- REVISION to grant number: **5U41HG002371-14**  
(This application is for additional funds to supplement a currently funded grant.)
- CHANGE of program director/principal investigator.  
Name of former program director/principal investigator: \_\_\_\_\_
- CHANGE of Grantee Institution. Name of former institution: \_\_\_\_\_
- FOREIGN application     Domestic Grant with foreign involvement    List Country(ies) Involved: \_\_\_\_\_

INVENTIONS AND PATENTS (Renewal appl. only)     No     Yes  
 If "Yes,"     Previously reported     Not previously reported

**1. PROGRAM INCOME (See instructions.)**

All applications must indicate whether program income is anticipated during the period(s) for which grant support is request. If program income is anticipated, use the format below to reflect the amount and source(s).

Budget Period	Anticipated Amount	Source(s)

**2. ASSURANCES/CERTIFICATIONS (See instructions.)**

In signing the application Face Page, the authorized organizational representative agrees to comply with the policies, assurances and/or certifications listed in the application instructions when applicable. Descriptions of individual assurances/certifications are provided in Part III and listed in Part I, 4.1 under Item 14. If unable to certify compliance, where applicable, provide an explanation and place it after this page.

**3. FACILITIES AND ADMINSTRATIVE COSTS (F&A)/ INDIRECT COSTS.** See specific instructions.

- DHHS Agreement dated: **7/12/12**     No Facilities And Administrative Costs Requested.
- DHHS Agreement being negotiated with \_\_\_\_\_ Regional Office.
- No DHHS Agreement, but rate established with \_\_\_\_\_ Date \_\_\_\_\_

CALCULATION\* (The entire grant application, including the Checklist, will be reproduced and provided to peer reviewers as confidential information.)

a. Initial budget period:	Amount of base \$	<u>228,659</u>	x Rate applied	<u>53.00</u>	% = F&A costs	\$	<u>121,189</u>
b. 02 year	Amount of base \$	<u>235,923</u>	x Rate applied	<u>53.50</u>	% = F&A costs	\$	<u>126,219</u>
c. 03 year	Amount of base \$	<u>246,073</u>	x Rate applied	<u>54.00</u>	% = F&A costs	\$	<u>132,879</u>
d. 04 year	Amount of base \$	_____	x Rate applied	_____	% = F&A costs	\$	_____
e. 05 year	Amount of base \$	_____	x Rate applied	_____	% = F&A costs	\$	_____
						TOTAL F&A Costs	\$ <span style="border: 2px solid black; display: inline-block; width: 100px; height: 20px; vertical-align: middle;"></span>

\*Check appropriate box(es):

- Salary and wages base     Modified total direct cost base     Other base (Explain)
- Off-site, other special rate, or more than one rate involved (Explain)

Explanation (Attach separate sheet, if necessary.):

**4. DISCLOSURE PERMISSION STATEMENT:** If this application does not result in an award, is the Government permitted to disclose the title of your proposed project, and the name, address, telephone number and e-mail address of the official signing for the applicant organization, to organizations that may be interested in contacting you for further information (e.g., possible collaborations, investment)?     Yes     No