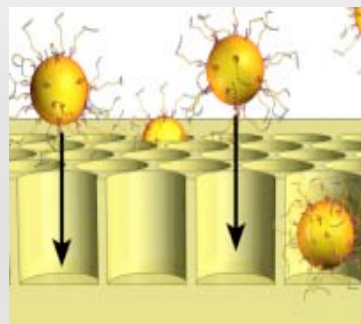


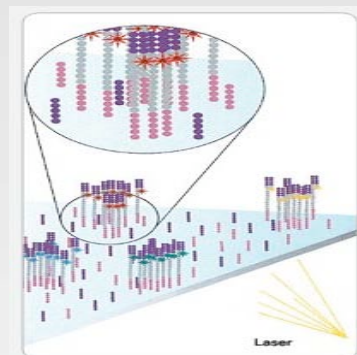
New Technology, New Ideas, New People

Dawei Lin, Ph.D.
Bioinformatics Core
UC Davis Genome Center
(Twitter: @iGenomics)

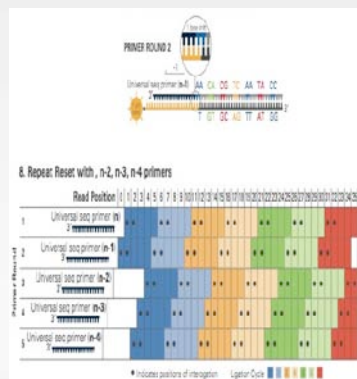
What is after ultra scale sequencing



454/Roche

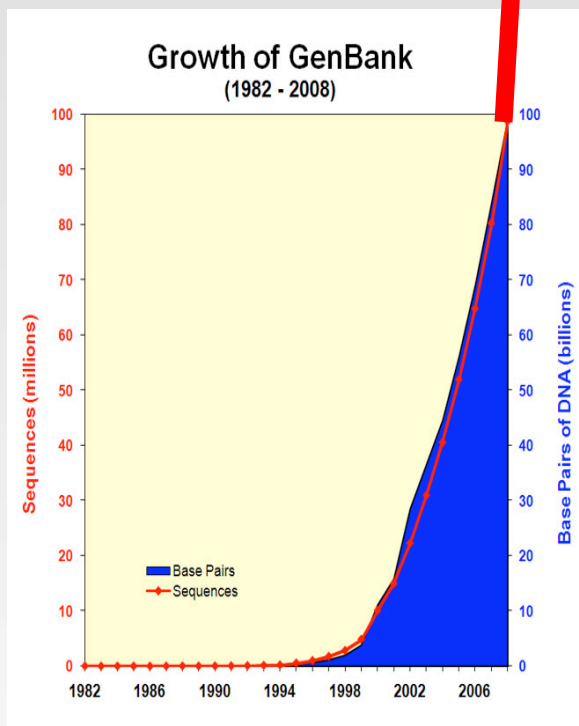


Solexa/Illumina



SOLiD/ABI

Human Genome/5 minutes



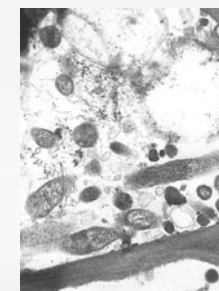
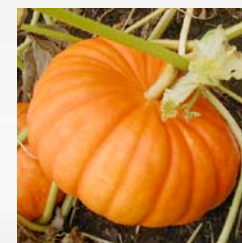
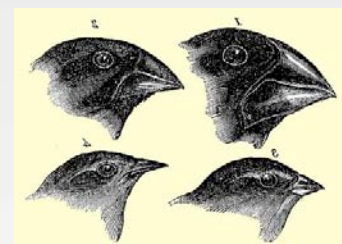
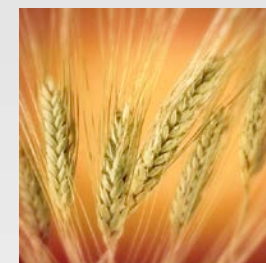
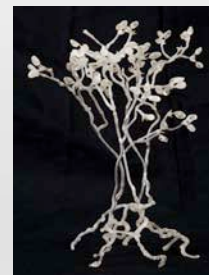
98M sequences (99Gbp), 2008

- ◆ *De Novo Sequencing****
- ◆ SNP discovery***
- ◆ Resequencing
- ◆ Methylation Sites Detection
- ◆ ChIP-Seq*
- ◆ Nucleosome Positioning
- ◆ Genome Structural Variation
- ◆ Chromosomal Rearrangement
- ◆ Transcriptome Profiling**
- ◆ Capture Array Sequencing
- ◆ microRNA Profiling
- ◆ Metagenomics*
- ◆ Novel Use of New Capabilities*

Next next

New Ideas, New People

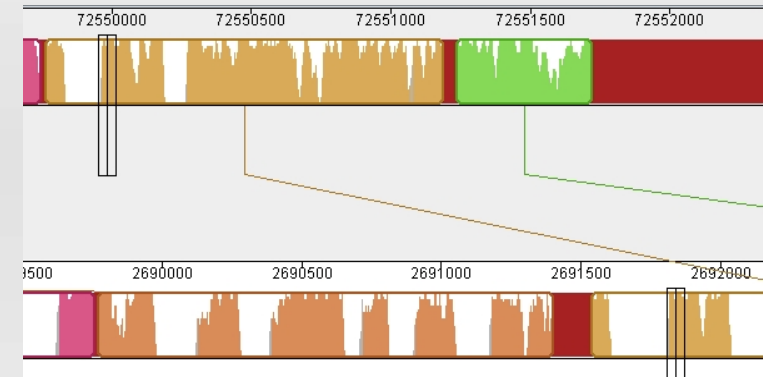
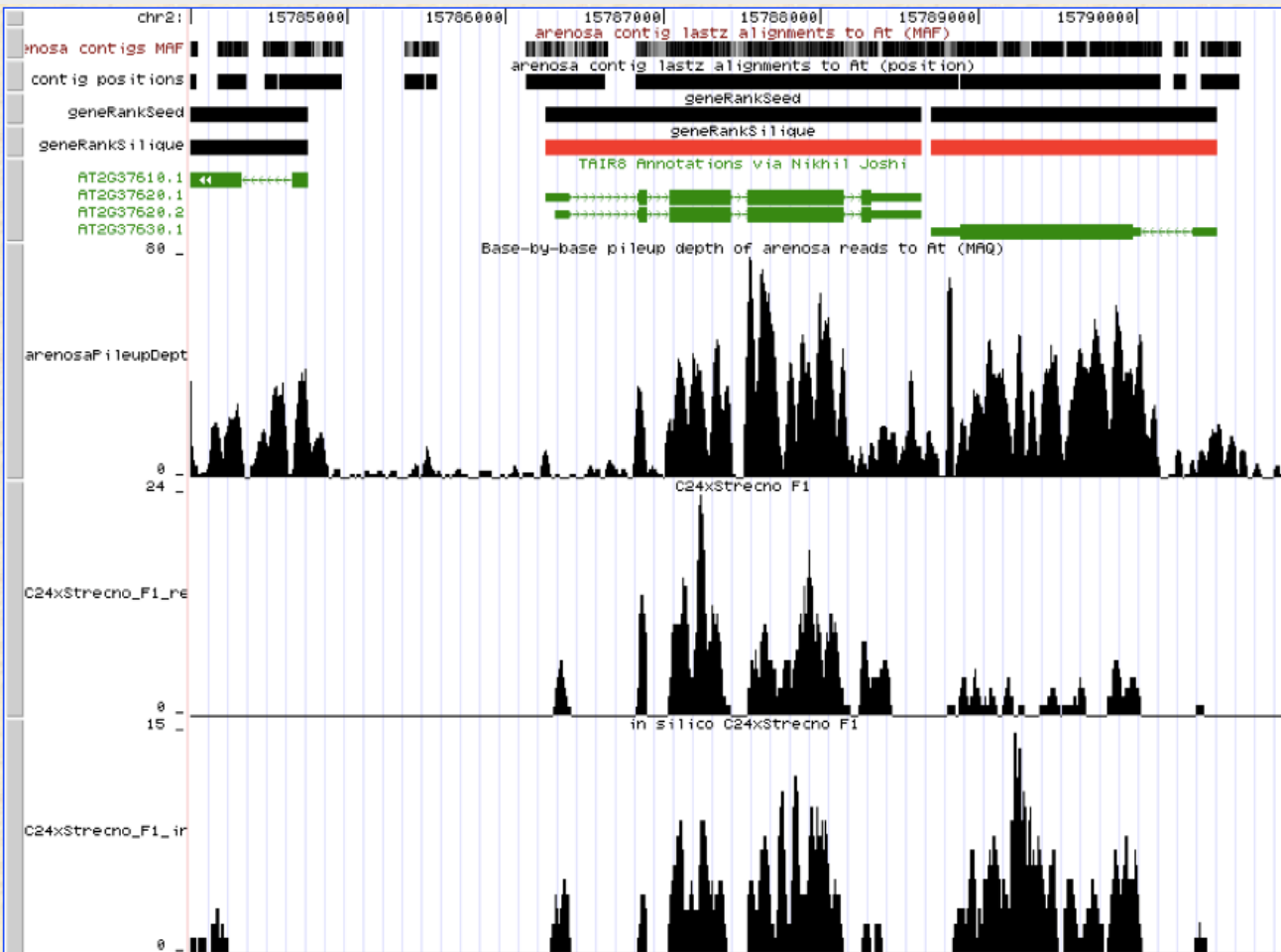
- ◆ **De Novo Sequencing**
Focused sequencing
Large number of small genomes
- ◆ **SNP Discovery**
No reference genome sequences
- ◆ **Transcriptome Profiling**
Unknown transcriptomes
- ◆ **Metagenomics**
Special interests
- ◆ **Novel Use of New Capabilities**
High-throughput mapping



Frequent Asked Questions

- We know nothing about sequencing!
- It is an unknown genome, can you do it?
- Which sequencing technology should I use?
- How many sequences (coverage) are needed?
- We can only do one run, is it enough?
- How many samples do I need?

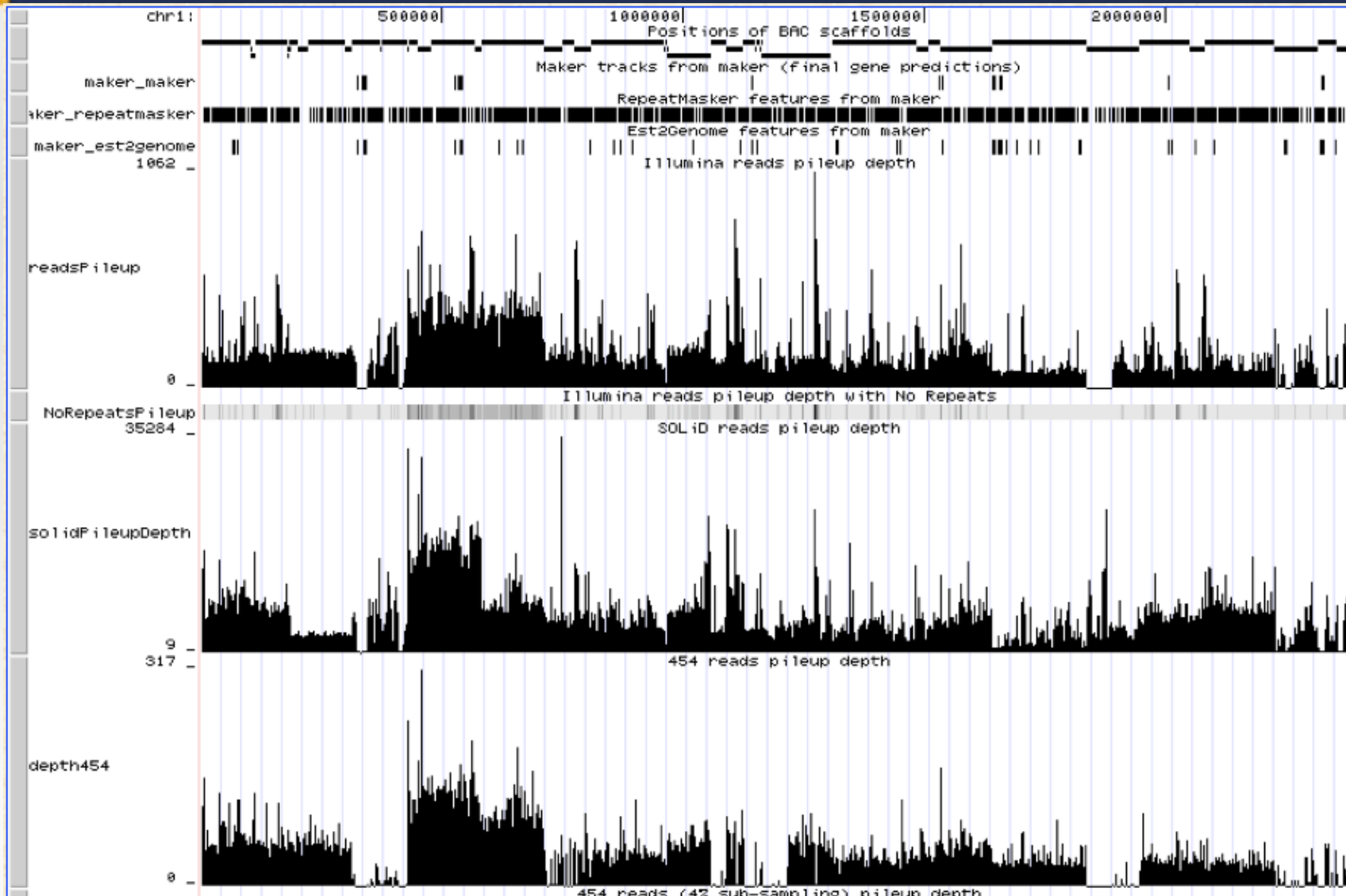
Post Assembly Analysis for *De novo* Sequencing



- Map to Close Relative Genomes
Blast, LastZ, Mauve, Syntenic mapping
- Gene Prediction and Annotation
Maker, EST mapping

It is an unknown genome, can you do it?

Limitations of Next Gen Sequencing



Known Sequence

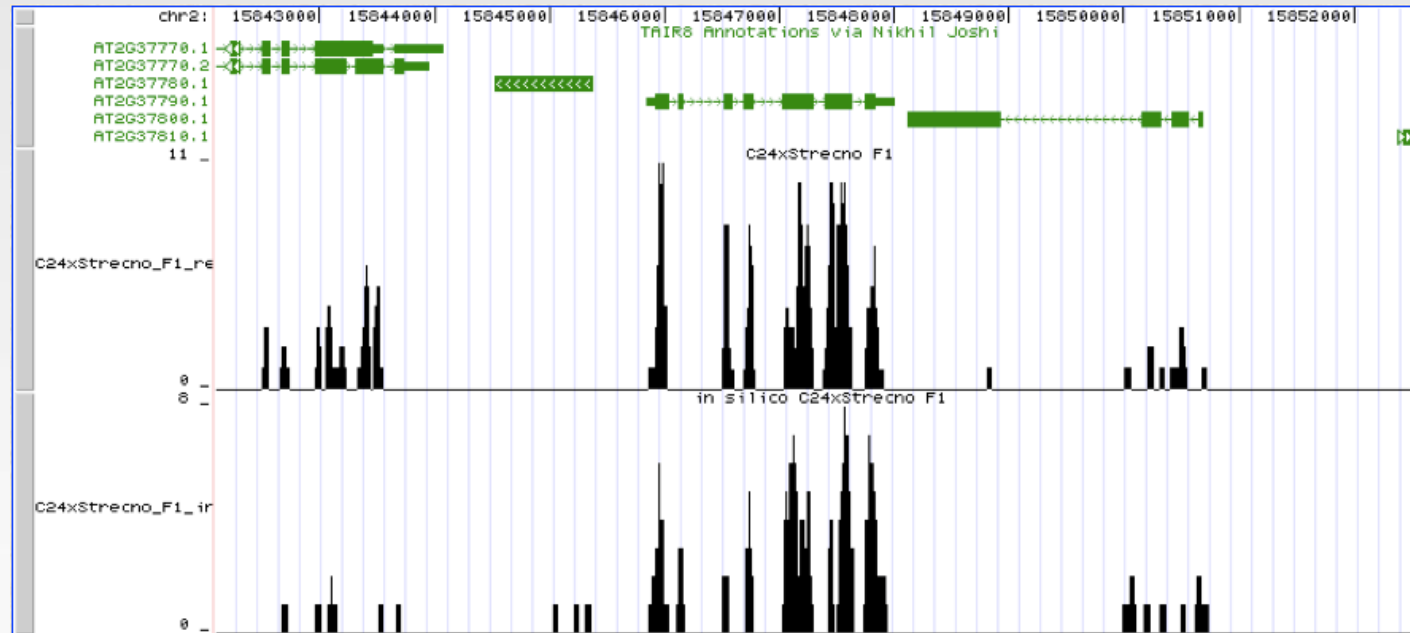
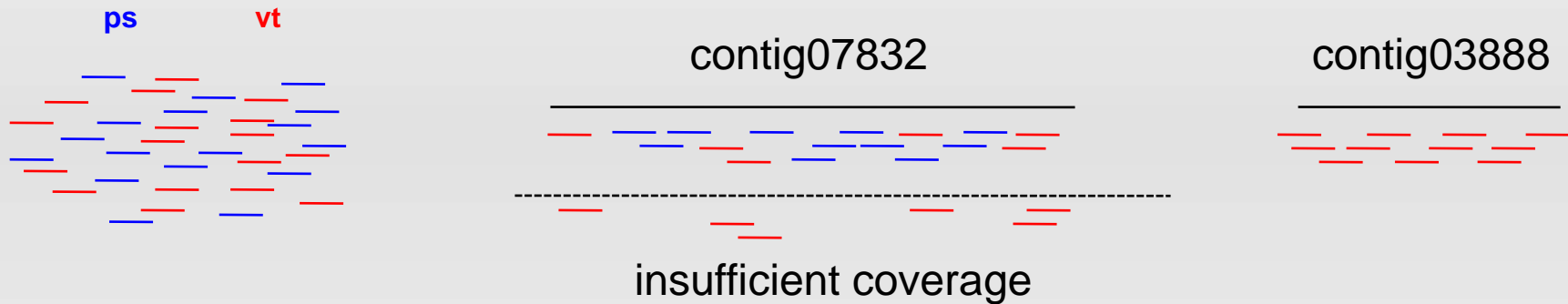
Solexa

SOLiD

454

Which sequencing technology should I use?

Post Assembly Analysis of Transcriptome Profiling



How many sequences (coverage) are needed?

Important Infrastructure

Solexa

Manage | Solexa Tools | Administrative | Help | Logout

View Solexa Run

General Data

Machine SOLEXA1
Cycles 45
Run Directory 080522_SOLEXA1_20813AAXX
Run Date 2008-05-22
Submitted 2008-05-28 11:24:26
Description Combination Care genomic, outside user, and Marta libraries

Notes

Lane Group	Organism	Concentration	Description	Summary	Files
1	Richard Michelmore Arabidopsis thaliana	3.5	Na-	Summary	View Files
2	Richard Michelmore Arabidopsis thaliana	3.5	Na+	Summary	View Files
3	Richard Michelmore Arabidopsis thaliana	3.5	TMAC-	Summary	View Files
4	Richard Michelmore Arabidopsis thaliana	3.5	TMAC+	Summary	View Files
5	Nipper Unknown	3	unknown	Summary	View Files
6	Luca Comai Arabidopsis var.	4	Care	Summary	View Files
7	Luca Comai Arabidopsis var.	4	Care	Summary	View Files
8	Charlie phix174	4	Phix	Summary	View Files

Storage in a Cloud

BIOSHARE

Manage | Bioshare Tools | Administrative | Help | Logout

View Sub Projects

Submitted	Type	Description	Files	Upload	Edit	Delete
2008-07-07	Presentation	ISMB Panel discussion on "best practice on building and managing a bioinformatics Core"	View Files	Upload Files	Edit	Delete
2008-07-09	proteomics	Add proteomics data	View Files	Upload Files	Edit	Delete
2008-07-10	Database	test database	View Files	Upload Files	Edit	Delete
2008-07-10	BLAST		View Files	Upload Files	Edit	Delete
2008-07-28	Presentation	Canada trip	View Files	Upload Files	Edit	Delete
2008-08-19	Presentation	Presentations from collaborators	View Files	Upload Files	Edit	Delete
2008-08-19	Presentation	Presentation of mine	View Files	Upload Files	Edit	Delete
2008-09-10	Presentation	Bioinformatics Core presentation templates	View Files	Upload Files	Edit	Delete

Project in a Cloud

Overview

Details

Gene G1uma01g40650

Phaseolus_vulgaris_McClean g1307

Marker g1307

1. See Detail about g1307
2. Search Phaseolus Genes Toolbox for marker
3. Search CMap for marker

Tracks

- Analysis All on All off
- Restriction Sites
- General All on All off
- CDS
- Clone from soybase
- Genes All on All off
- Gene

Five Prime UTR Phaseolus_vulgaris_McClean
 Three Prime UTR

Genome Browser



Biorhythm – Time Tracking for Projects

View entries:

By date:

from: Jun 08 to: Jun 17

By employee:

 amschaal
 ellisocj
 hansvg
 jboveda
 jfass
 lcomai
 lhslin
 najoshi
 nik
 zwluxx

Search

Timesheets by project

Bioinformatics Website

total time: 31 hours

DATE	HOURS	EMPLOYEE	DESCRIPTION	CONTROLS
JUN 17	4.00	jboveda	Debugging PHP issues	
	1.00		Programming authors page	
	0.50		Research into wiki-style links	
	0.50		Programming wiki-style links into site	
JUN 15	4.00		Debugging PHP handling issue	
JUN 11	3.00		Programming bioinfo template search.php	
	2.00		Programming - loading info on site	
JUN 10	7.00		Programming bioinfo admin integration to homepage for logged in user	
JUN 09	1.50		Debugging is_post issue in 2.7.1	
JUN 08	1.00		Debugging the_author_meta issue	
	0.50		Installation/Configuration of wordpress 2.8 RC	
	3.00		Programming visual updates	

Timesheets by employee

jboveda

total time: 46 hours

DATE	HOURS	PROJECT	DESCRIPTION	CONTROLS
JUN 17	1.50	Comai: Tilling	Meetings with Kathie & Helen - outlined additional parameters	
	4.00	Bioinformatics Website	Debugging PHP issues	
	1.00		Programming authors page	
	0.50		Research into wiki-style links	
	0.50		Programming wiki-style links into site	
JUN 16	8.00	Internal Affairs	Sick	
JUN 15	4.00	Bioinformatics Website	Debugging PHP handling issue	
	1.00	Drug Rediscovery	Meetings with Adam on rediscovery	
JUN 11	3.00	Bioinformatics Website	Programming bioinfo template search.php	
	0.50	Drug Rediscovery	Research into codebase	

Grants, Hourly Based Projects

Acknowledgement

Thank You!