Content       Content       Destruction       Destruction       Destruction         Content       Content<		f	
In the contract of the cont	🔶 📿 🔕 🏠 🏹 http://main.g2.bx.psu.edu/		* Galaxy Genomics Penn State
And And And And And And And And And And	Galaxy	Info: report bugs   wiki   screencasts	blog Logged in as aun1@psu.edu: manage   logou
Cutotic         Cutotic <t< th=""><th>ls</th><th>Join</th><th>History (options)</th></t<>	ls	Join	History (options)
Extension (Constraints)         Extension (Constraints)         Extension (Constraints)           Text and States (Constraints)         Extension (Constraints)         Extension (Constraints)         Extension (Constraints)           Text and States (Constraints)         Extension (Constraints)         Extension (Constraints)         Extension (Constraints)           Extension (Constraints)         Extension (Constraints)         Extension (Constraints)         Extension (Constraints)           Extension (Constraints) <t< td=""><td>Data</td><td></td><td>refresh   collapse all</td></t<>	Data		refresh   collapse all
International Sector       International Sector         Distributional Sector       Sector         Distrint Intervent Sector       Sector	ENCODE Data	1: RelSeq coding exons	2: Human/Chimp/Dog alignments for   ●   0
International constraints       One         0.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1	t Manipulation	First query	Exons from Step 6 957 serviences formati fasta database ho18
On. National Solutions         Caretae         Caretae         Caretae           Constructions         Exercise         Exercise         Exercise         Exercise           Constructions         Exercise         Exercise         Exercise         Exercise           Exercise         Exercise         Exercise         Exercise	er and Sort	with:	Info: 319 regions were extracted successfully.
Answer         Answer           Exercision Exercisio	n, subtract and Group	2. SNP5	save
According Exploring Statistics School Scho	ivert Formats		>hg18.chr8(+):18301793-18302666
Endemontant         Endemontant           Endemontant	ch Sequences	with min overlap:	ATGGACATTGAAGCATATTTTGAAAGAATTGGCTATAAGAACTC >pantro2
Generations       Contract in the intervale of two queries       Contract in the intervale of two queries <ul> <li> </li> </li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></ul> <li> </li> </li> <li> </li> <li> </li> <li> <li> <li> <li> <li> <li> <li> <li> </li> <li> <li> <li> <li> </li> <li> </li> <li< td=""><td>ch Alianments</td><td></td><td>ATGGACATTGAAGCATATTTTGAAAGAATTGGCTATAAGAACTC</td></li<></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li></li>	ch Alianments		ATGGACATTGAAGCATATTTTGAAAGAATTGGCTATAAGAACTC
Return Instance Interest the network of two queries         Return Interval Int	Genomic Scores		>canFam2
<ul> <li>interesting the networks of two queries</li> <li>interesting the networks of two queries</li> <li>interesting the networks of a query the networks of a query three networks of the titter networks of a query three networks of the titter q</li></ul>	state on Genomic Intervals	Return:	
<ul> <li>Execute the reveals of two contexis</li> <li>Execute the every pay intervals of a cuery.</li> <li>Consistents two queries into one query more and intervals of a lintervals of a cuery.</li> <li>Execute the every of a query more and a subject contexponent, the "citie attributer" to set chromosome, start, a reveal of a query.</li> <li>Execute the every of a query more and a subject contexponent of the more query and an interval function.</li> <li>Execute the every of a query more and a subject contexponent of the more query and an interval function.</li> <li>Execute the every of a query of the more query and more and a subject contexponent.</li> <li>Execute the every of a query of the more query and more and a subject contexponent.</li> <li>Execute the every of a query of the more query and more and a guery are filted in with a period. (Note that this ana query are filted in with a period. (Note that this ana query are filted in with a general (Note that this and prevaled of the first query and and interval file, since a period. (Si a query and and interval file, since a period. (Si a query and and interval file, since a period. (Si a query and and interval file, since a period. (Si a query and and and interval file, since a period. (Si a query and and and the provid. Si a query and and and the provid. Si a query and and and interval file, since a period. (Si a query and and and the provid. Si a query and and and interval file, since a period. (Si a query and and and prevaled.) Interval file, since a period. (Si a query and and and and and and and and and and</li></ul>	ntersect the intervals of two queries	Only records that are joined (INNEK JC	
<ul> <li>Known or eventpoint intervols of a curve</li> <li>Known or eventpoint intervols of a curve</li> <li>Concretence of a littervols</li> <li>Concretence of a set of intervols on second set of a set of intervols of a curve vision of a curve vis</li></ul>	ubtract the intervals of two queries	Execute	6: Coordinates of Exons with at least 5  V SNPs
<ul> <li>Constants two queries into one query activation one query activation on queries into one queries of a set of intervals of a set of a set of intervals of i</li></ul>	erge the overlapping intervals of a query		
<ul> <li>and contents</li> <li>and c</li></ul>	oncatenate two queries into one query	0 11P: If your query does not appear in the pulldown menu -> it is not in interval format. Use "edit attributes" to set chromosome, start, end, and	
• Concesso of a set of intervals on second set of intervals of a cuery       • Complement intervals of a cuery         • Complement intervals of a cuery       • Complement intervals of a cuery         • Data intervals of a cuery       • Laster the intervals of a cuery         • Data intervals of a cuery       • Calaxy intervals of a cuery         • Laster the intervals of a cuery       • Calaxy intervals of a cuery         • Laster the intervals of a cuery       • Calaxy intervals of the intervals of the intervals of the first query, and any intervals that cuery any cuery any any intervals that cuery and a	ase Coverage of all intervals		4: Number of SNPs per Exon
<ul> <li>Constant intervals of a cuery</li> <li>Constant intervals of a cuery</li> <li>Linster the intervals of a cuery</li> <li>Linster the intervals of a cuery</li> <li>Linster intervals of two cuerenes side-by-sale</li> <li>Linstervals of two cuerenes side-by-sale</li> <li>Linstervals of two cuerenes side-by-sale</li> <li>Linstervals of the first cuery that allows them to be joined.</li> <li>Linstervals of the first cuery that allows them to be joined.</li> <li>Linstervals of the first cuery that allows them to be joined.</li> <li>Linstervals of the first cuery and intervals that allows them to be joined.</li> <li>Linstervals of the first cuery, and any intervals that allow them allows them to be joined.</li> <li>Linstervals of the second query, and any intervals that do not join an intervals of the first cuery and interval fils, since a period(1) is an exampter to a transport of the second query. and any intervals that do not join an interval of the second query.</li> <li>Linstervals that any produce an invalid interval fils, since a period(1) is not a valid chrom, start, end or starend or start, end or start, end or start, end or start, end or</li></ul>	overage of a set of intervals on second set of tervals	Screencasts!	3: Join of Exons and SNPs
<ul> <li>cluster the intervals of a query</li> <li>cluster the intervals of a query</li> <li>bill the intervals of two queries side-by-side</li> <li>bill the intervals of two queries side-by-side</li> <li>certianis returns flanking region/s for every</li> <li>certianis return and records of first query (fill null with "")" returns all intervals of the first query, and any intervals that do not join an introvision the second query. This is and every intervals that do not join an introvision there are invalid interval fille, since a period(.) This is may produce an invalid interval file, since a period(.) Sin</li></ul>	omplement intervals of a query	See Galaxy Interval Operation Screencasts (right dick to open this link in another window).	2: SNPs
<ul> <li>Juin the intervals of two queries side-byside</li> <li>Sortax</li> <li>Buin the intervals of two queries side-byside</li> <li>Certifiands returns flanking region/s for every gene Return and intervals of the first query that join to a record query are flight in with "".") returns all intervals of the first query, and any intervals that do not join an intervals that intervals of the first query that join to a record query are flight in with a sendogue to a LETT JOIN.</li> <li>Return and intervals of first query (fill null with "") returns all intervals of the first query, and any intervals that do not join an interval for miter strate are priorided. The second query of fill null with "") returns all intervals of the ond join.</li> <li>Return and in the second query fill null with "") returns all intervals of the ond point.</li> <li>Return and intervals of both quere (fill null with "") returns all intervals of the ond point.</li> <li>Return and intervals of both quere (fill null with "") returns all intervals of the ond point.</li> <li>Return and interval file, since a period(.) The second query fill null with "") returns all intervals of the ond point.</li> <li>Return and interval file, since a period(.) The second query fill null with "") returns all intervals of the ond point.</li> <li>Return and interval file, since a period(.) The since a period(.) The</li></ul>	luster the intervals of a query		
Get flanks       Cet flanks <td>in the intervals of two queries side-by-side</td> <td>Syntax</td> <td>259,080 regions, format: bed, database: hg18</td>	in the intervals of two queries side-by-side	Syntax	259,080 regions, format: bed, database: hg18
gene statististic     event on NURE ADIN.     event on NURE ADIN.       Carabilization for the second query are first query (fill null with "") returns all intervals (that do not join an intervals the second query are filed in with a sendogus to a LET JOIN.     Intervals that do not join an intervals the second query are filed in with a sendogus to a LET JOIN.       Evolution: HVPhy returns of interval file in with a sendogus to a LET JOIN.     Intervals of the first query (fill null with "") returns all intervals of the acond query are valid chorn, start, end in with a sendogus to a LET JOIN.       Return all records of the query (fill null with "") returns all intervals of the acond query are valid chorn, start, end in with a produce an invalid interval file, since a period(.) Is produce an invalid interval file, since a period(.) Is not a valid chronn, start, end or strand.       Return all records of both queries (fill null with "") returns all records of the query wild chorn start, end or strand.       Return all records of both queries (fill null with "") returns all records of the query wild chorn, start, end or strand.       Return all records of both queries (fill null with "") returns all records of the query wild chorn, start, end or strand.       Return all records of both queries (fill null with "") returns all records of the query wild chorn, start, end or strand.       Return all records of both queries (fill null with "") return all records of the query wild chorn, start, end or strand.       Return all records of both queries (fill null with "") return all records of the query wild chorn start.       Return all records of both queries (fill null with "") return all records the ruph or return with records of the strand.	et flanks returns flanking region/s for every	Where overlap specifies the minimum overlap between intervals that allows them to be joined.	Info: Uploaded from UCSC
Statistication         Return all records of the first, query (fill null with "") returns all intervals of the first, query in an into two the first, query (fill null with "") returns all intervals of the first, query into the price of the first, query (fill null with "") returns all intervals of the first, query into the price of the tits analyzed on the price of the tits analyzed on the first, query (fill null with "") returns all intervals that do not join.           Return all records for on the first, query are filled in with a period(). Note that this analyzed on the first, query are filled in with a period(). Note that this anay produce an invalid interval file, since a period() is on a valid work of the view of the tit period.           Return all records of both query are filled in with ") returns all intervals that on other that interval file, since a period() is not a valid or form, start, end or strand.           Return all records of both query are filled interval file, since a period() is not a valid or file with ").           Return all records of both query are filled interval file, since a period() is not a valid or file with periods.           Return all records of both query are filled interval file, since a period() is not a valid chrom, start, end or strand.           Return all records of both query are filled interval file, since a period() is not a valid chrom, start, end or return.           Return all records of both query are filled interval file, since a period() is not a valid chrom, start, end or return.           Return all records fille, since a period() is not a valid chrom, start, end or return.           Return all records fille and the return all records fille and and interval file, since a period() is not a valid	ene	<ul> <li>Return only records that are joined returns only the records of the first query that join to a record in the second query. This is analogous to an INNER JOIN.</li> </ul>	<u>save</u> I display at UCSC <u>main</u>
Intervent     The second query area freed in with a period(), has a reargadous on any intervals that do not join interval file, since a period() is interval file, since a period() is vitery and any intervals that do not join interval file, since a period() is vitery and any intervals that do not join interval file, since a period() is vitery and any intervals that do not join interval file, since a period() is vitery and any interval file, since a period() is vitery and any intervals that do not join interval file, since a period() is vitery and any interval file, since a period() is vitery and any interval file, since a period() is vitery and any interval file, since a period() is vitery and any interval file, since a period vitery and any interval file and any interval file and any interval file and any interval file any preany and any	tistics ab / Diculary Data	• Return all records of first query (fill null with), returns all intervals of the first query, and any intervals that do not join an interval from	1 2 3 4
EMBODS       Interval from the first owner are filled in with a period(.). Note that this may produce an invalid interval file, since a period(.) is not a valid driven, start, end or strand.         • Return all records of both quories (fill nulls with "")" returns all records from both quories, and fills on either the right or left with period(.) is not a valid driven, gart, end or strand.         • Return all records of both quories (fill nulls with "")" returns all records from both quories, and fills on either the right or left with period(.) is not a valid driven, gart, end or strand.         • Return 100 100 densrp11.2       Effect         • Main 100 100 densrp11.2       Effect query is:         • Return 100 100 densrp11.2       Effect query is:         • Return 200 0 densrp1.2       Effect densrp1.2         • Return 200 0 densrp1.2       Effect densrp1.2         • Return on those will generate:       Effect densrp1.2         • Return onther exclude sthat are joined:       • Return onther exclude sthat are poined:	lution: HvPhv	the second query are lined in whith a periodic). This is an allogous to a LEFI JULN • Return all records of second query (fill null with "") returns all intervals of the second query, and any intervals that do not join an	chr1 67052400 67052451 NM_024763_cds_0_0_chr
Return current, sark end or stratus.     Return all records from both queries, and fills on either the right or left wit periods. Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or left wit in a log on early in a log on e	3055	Interval from the first query are filled in with a period(.). Note that this may produce an invalid interval file, since a period(.) is not a	Chrl 5/06051 5/060188 NM_024/53_CdS_1_0_Chr chrl 57055090 57055317 NM 024753 cds 2 0 chr
Periods. Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.         Example       If First query is:         chail 10 100 duery1.1       chail 10 100 duery1.1         chail 200 luony1.2       chail 200 luony1.2         chail 200 luony1.2       chail 200 luony1.2         chail 200 luony1.2       chail 200 luony1.2         chail 200 luony1.2       chail 200 luony2.2         chail 200 luony2.2       c		valid curroin, setty can or a stratiu. • Return all records of both queries (fill nulls with "") returns all records from both queries, and fills on either the right or left with	chr1 67066082 67066181 NM 024763 cds 3 0 chr
Example           First query is:           chi 1 00 100 0uery1.1           chi 1 100 1230 0uery1.2           chi 1 200 200 0uery1.2           chi 1 200 300 0uery2.2           chi 1 200 300 0uery2.2           chi 1 200 100 0uery2.2		periods. Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.	chr1 67071855 67071977 NM 024763 cds 4 0 chr
If First query is:           chr1 10 100 Query1.1           chr1 10 100 Query1.2           chr1 100 1290 Query1.2           chr1 120 20 Query1.2           chr1 200 204 Query1.2           chr1 200 204 Query2.2           chr1 200 204 Query2.2           chr1 200 300 Query2.2           chr1 200 724 Query2.2           chr1 200 724 Query2.2           chr1 200 Yun Cuery2.2		Example	
chr1       10       100       Query1.1         chr1       100       Query1.2         chr1       100       Query1.2         chr1       200       Query1.2         chr1       200       Query2.2         chr1       200       Query2.2         chr1       200       Query2.2         The four return options will generate:          • Return only records that are joined:		If First query is:	
chr1 100 1200 Query1.12       chr1 110 1200 Query1.13       and Second query is:       chr1 20 80 Query2.13       chr1 20 90 Query2.14       chr1 20 90 Que	_	chr1 10 100 Query1.1	
and Second query is:         aheri 200 % 0 warp2:1         aheri 200 % 0 warp2:2		chr1 500 1000 Query1.2 chr1 1100 1250 Query1.3	
att: 20 80 Ouer92.1 att: 260 200 Ouer92.2 att: 2500 200 Quergr.2 The four return options will generate: • Return only records that are joined:	_	and Second query is:	
chr1 2500 3000 Guer22.3 The four return options will generate: • Return only records that are goined:		chri 20 80 Query2.1 chri 2000 2204 Query2.2	
The four return options will generate: • Return only records that are goined:		chri 2500 3000 Query2.3	
Return only records that are joined:		The four return options will generate:	
		• Return only records that are joined:	
Done			

## Tools (left)

Tools (currently ov 100) are organized i categories. Clicking on category expands an shows available tool For example, here th Operate on Genom Intervals category expanded.

# Interfaces (cente

If you click a tool, its interface will appear he For example, clicking on *Join* (left pane) w bring up its interface (shown in the cen pane). It this case we are joining RefSeq exc with SNPs. The names of datasets in the t dropdowns ("1:RefSeq coding exons" a "2:SNPs") correspond to items listed in ( right (History) pane. The tool interface a provides help information.

# **History (right)**

This is where all your data and analyses results go. The output of every tool creates another history item, so that your original data is never changed. For example, we started with two datasets imported from UCSC: "1:RefSeq coding exons" and "2:SNPs". Joining the two datasets created a new History Item called "3:Join on Exons and SNPs". And so on... Thus the History serves as a complete record of every analyses. Clicking "options" (just above the right pane) allows saving and sharing (*l*) of histories.

### DON'T DO GENOMICS ALONE!

scion fulvieolis

Introducing GALAXY, a revolutionary tool for collaborative genomic research! http://g2.bx.psu.edu

(since you cannot click the paper - flip it!)

### See what **GALAXY** can do for you:

### **Biologists**

Use our public site (http://g2.bx.psu.edu) to access popular sources of data like the UCSC Table Browser. Run analyses right on the spot using a variety of integrated tools. Your results are always available and can be easily shared with others.

### **Software Developers**

Galaxy is an easy-to-use, open-source, scalable framework for tool and data integration. Stop wasting time writing interfaces and get your tools used by biologists! GALAXY includes everything you need to get started, so download and start integrating!



You are an experimental biologist. You keep watching databases fill with more and more data. You keep thinking: "even if I knew how to use Excel as a pro, it would probably not load 12,435,654 SNPs". So how do you perform analyses without calling somebody on the Computer Science side of campus? Suppose you want to find human exons with the highest SNP density. There is no straightforward way of doing it without learning programming first. And this is why:

#### Databases are not analyses tools

Databases are where you get the data. Browsers are where you visualize the results. For a bench biologist there is not much in between besides spreadsheets or Perl scripting. GALAXY fills this gap by providing a tool-rich analysis medium.

#### No tools for essential datatypes

Some datatypes generated by high throughput genomics are so new that there are no tools to analyze them. For example, how do you extract seq-uences of coding exons from the latest 28-way alignments of vertebrate genomes or analyze genome-wide disease association data? With GALAXY!

### Is genomics reproducible?

The Methods section of too many papers sound like *"the data were analyzed using a collection of in-house scripts"*. How do you repeat such an analysis? GALAXY saves every step of your analysis and allows you to share these workflows with others.

#### **Too many tools**

*Bioinformatics* has published 166 Application Notes since January 2007. How does one know which tool to use? GALAXY integrates a multitude of different tools by giving them the same "look and feel" and linking them to data warehouses.

### How?

So is it all hot air or can you actually do something with GALAXY that you cannot do otherwise? Let's try this:

find all human exons that contains at least 5 nonsynonymous (amino acid changing) SNPs and extract their alignments with chimp and dog

This analysis requires 7 steps outlined below but can be completed easily and quickly with GALAXY (in fact, 90% of that time is spent downloading SNP data from the UCSC Table Browser). These steps correspond exactly to the contents of GALAXY's left pane on the opposite side of this brochure. You can watch a QuickTime screencast detailing this analysis at http://g2.bx.psu.edu.

#### **Download exon and SNP coordinates**

Here we just grab coordinates of all coding exons and SNPs from the UCSC Table Browser (the efforts of the UCSC Team were crucial for this integration). This creates History items #1 and #2 (right pane of Galaxy's interface on the opposite side)

### Join exons with SNPs

This will find all exons that contain SNPs. This operation is performed using the *Join* function and creates History item #3. *Join* is a part of our large suite of tools for manipulation of genomic intervals that include intersection, coverage, and other functions.

### **Compute number of SNP per Exon**

Here we first compute the number of SNPs per exon with the *Count* function (History item #4) and then use the *Filter* tool to find exons with at least 5 SNPs (History item #5).

### **Extract human/chimp/dog alignments**

Finally, we extract human/chimp/dog alignments corresponding to the exons from the previous step. These alignments are trimmed per exon boundaries. If multiple alignment blocks overlap an exon - they are stitched together (History item #6).

### See it!

Still not convinced? Go to http://g2.bx.psu.edu. There you will see a link to screencasts (you will need a free QuickTime player available from Apple website to view them). Screencasts highlight all aspects of GALAXY's functionality and also show examples of very complex analyses.

### **1** Share!

Best of all, GALAXY's history system provides a complete analyses record that can be shared. Every history is an analysis workflow, which can be used to reproduce the entire analyses.

### History is an analysis record

Look at GALAXY's interface on the opposite side. Every step of the analyses is recorded. You can have any number of histories saved. This way you can go back to your analyses anytime.

### Share your analyses

Alice works at Penn State, while Bob suffers from the terrible San Diego climate. Alice wants Bob to see her analyses. Alice clicks the "*share*" link and enters Bob's e-mail address. Now Alice's history is visible to Bob (see "*Sharing history*" screencast).

### Now your results are reproducible!

When publishing results, replace "the data were analyzed using a collection of in-house scripts" with a URL pointing to Galaxy's history. Your reviewers will have no further questions. That's reproducible genomics!