

Center for Translational Molecular Medicine

iReport: HTML Reporting in Galaxy

Saskia Hiltemann (1,2), Youri Hoogstrate (1,2), Hailiang Mei (3), Peter van der Spek (2), Guido Jenster (1), Andrew Stubbs (2)

1)Dept. of Urology, Josephine Nefkens Institute, Erasmus MC, Rotterdam, The Netherlands; 2) Dept. of Bioinformatics, Erasmus MC, Rotterdam, The Netherlands 3) Leiden University Medical Center (LUMC), Leiden, The Netherlands

IFUSE calc stats Var unts on data 4, data 5, and others - TSV file

I: IFUSE calc stats TIT HetHom on data 4, da a 65, and others - TSV file

74: Chromosomes mo ntage on data 1, data 6 5, and others 73: Whole genome plot 💿 🖋

iReport

TraIT Galaxy server: galaxy.ctmm-trait.nl Example iReport (published history): tinyurl.com/llrzz9w

The CTMM TralT Project

TraIT will develop a long-lasting IT infrastructure for translational medicine that will facilitate the collection, storage, analysis, archiving, sharing and securing of the data generated in the CTMM operational translational research projects.



Basic structure



Report (version 1 Name of Report

Link to cover image

Cover page User-defined title and cover image (url)

Tab-based report

Add any number of tabs, then add any number of *content items* to each tab.

ional. A default image will be used if not spe Content items can be one of: Tab 1 - Text Enter tab name: Content-Items - Image Content-Item - Table Select Item Tvp Please choose item type - PDF Remove Content-Iter Add new Content-Item - Link Remove Tab 1 Add new Tab

Content Items

Text

Specify a text file from history, or enter text directly into a text field.

Content-Items

Content-Item

Select Item Type

Text Field

Text to display

<h3> Report: HTML reporting in Galaxy </h3> Report is a Galaxy tool for the easy creation of HTML reports from workflow outputs. An iReport consists of a cover page, and a tab-based report page

Users specify a title for their iReport, and optionally a link to a cover image. If no cover image is specified, a default image is used. An example (the cover page for this report) is shown below. Clicking on the image or on the link above it, will take the user to the main report page

Input is sanitized and printed verbatim, with the exception of a few html tags (h1-h6, strong, em)

\n and \t interpreted as newline and tab respectively to allow for some

The project builds on existing expertise to create an IT infrastructure that will help to accelerate the translational research in the Dutch Life Sciences and Health sector.

An example: iFUSE2

CGtag Galaxy workflow for downstream analysis of Complete Genomics Data (Hiltemann et al, *GigaScience*, 2013) generates many outputs.

iFUSE (Hiltemann et al, *Bioinformatics*, 2012) is a web-based tool for exploration, visualisation and prioritization of candidate fusion genes, and has since been ported into Galaxy.

iFUSE2 combines these two pipelines with newly incorporated copy number variation methods (OncoFUSE) and presents the results in an iReport

Overview page containing Circos plots and tables with summary statistics



can explicitly a	add whitespace adding \r	in your text for a newline	
------------------	--------------------------	----------------------------	--

control over whitespace

Insert break after item?

Image

Specify image file from history, and an image width (0 for original size). If image is scaled down, a jQuery *zoom-on-mouseover* effect is added to the image



Table

Specify tab delimited file from history

jQuery library datatables (http://datatables.net) used to create paginated, searchable, and sortable tables (if desired)

Show 10 • entries	how 10 🔻 entries														
Cars	mpg	mpg cyl d		hp	drat	wt	qsec	vs	am	gear	carb				
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2				
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4				
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4				
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4				
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2				
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6				
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1				
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4				
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1				
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1				
Showing 1 to 10 of 31 ent	ries									Pre	vious Next				

	TILET OTTEE OT	nrX chrY		<u>06 cnr07</u>	chr08 ch	nr09 ch	r10 chr	11 chr1	2 chr13 chr	14 chr15	chr16	CHIT		<u>.8</u>
Metric	Whole Genome	Exonic		All Variants	snps	ins	del	sub	Gene	All Variants	snps	ins	del	sub
structural	2.415	999	Whole Genome	279,465	131,248	75,703	48,978	23,536	PTPRD	287	139	73	63	12
/ariants	-,		exonic;splicing	1	1	0	0	0	CSMD1	236	146	40	32	18
small variants	279,466	2,250	intergenic	159,127	78,580	39,360	26,835	14,352	EYS	221	118	54	26	23
SNP	131,248	1,663	downstream	1,741	691	568	347	135	CNTNAP2	207	103	50	41	13
NS	75,703	277	intronic	102,604	43,510	32,048	19,310	7,736	DLG2	199	98	48	33	20
DEL	48,978	149	UTR5;UTR3	1	1	0	0	0	LRP1B	181	88	42	28	23
SUB	23,536	161	ncRNA_splicing	1	1	0	0	0	RBFOX1	178	74	53	39	12
fi/Tv ratio	1.8048	1.76246	ncRNA_intronic	8,294	3,982	2,119	1,398	795	NRG1	178	104	41	25	8
let/Hom ratio	1.21596	1.26153	UTR3	2,060	887	654	393	126	MACROD2	175	67	64	37	7
			upstream	2,342	1,267	488	412	175	CTNNA3	170	90	46	24	10
			ncRNA_exonic	469	299	91	51	28	MAGI2	154	69	44	29	12
			UTR5	874	592	135	99	48	2					
			ncRNA_UTR3	33	9	15	8	1						
			upstream;downstream	90	47	22	17	4						
			ncRNA_UTR5	10	5	1	3	1						
			splicing	38	13	16	7	2						
			exonic	1,780	1.363	186	98	133						

Tabs for Structural Variants (fusion gene candidates), Small Variants, Copy Number Variations (e.g. OncoFUSE), SNP array data, and more

Genome Overview	Structural Variants	Iral Variants B-allele Frequency Coverage Virtual Normal																								
OncoSNPSeq OncoSNP-SEQ is an analytical tool Yau at Department of Mathematics,	for characterising copy number altera Imperial College London.	tions and loss-of-heteroz	ygosity <mark>(</mark> LOH) e	vents in cancer san	ples from (Co	nplete <mark>Genom</mark> i	cs) whole ge	nome seque	ncing data. It was	огі																
 OncoSNP-SEQ contains the following novel features: Loci classified into one of a number of copy number aberration states, each state describes a Ability to model both normal DNA contamination and intra-tumour heterogeneity jointly. Estimates baseline haploid read level for polyploid samples. Inference using Bayesian methods allowing uncertainty in estimates to be characterised. 		Genome	Overview	Structural \	ariants	B-allele F	Frequency	Covera	ge Vii	tual Norma	al															
		Fusion G	Gene Candidate	s:		1	Genom	Conome Quention Structure Variante R allela Fraguency Coverage Virtual Narmal																		
O Deno: 1 of 25 - + Automatic Zoom :			Show :	LO 🔻 entries			- P	Genom																		
Overview		Junction Related Associated Shared Gene					Fusion Gene Candidates:																			
		244	rj119	NA	sg160 yes	ir	Г																			
		761	rj1734	rj1734 NA gj120 yes i													1.7579km									
			837	rj1915	NA	g264 yes	yes i LPPR2													1.7579bp chr1.9:2869474(+) 10313bp						
			899	rj1304	NA	ig1437 yes	âr	chri	ZNF 57/LPPR2	. • •	-	· .	*	<u> </u>	•	- 1 - 1			•				11.326bp	3/3/4(+)		
			1517	rj38	NA	sg39 yes	ir	Gene	Name		ZNF57 (N	Left M 17348	0)				Gene	Name		LPPR2 (N	Right IM 022737)					
			1565	rj39	NA	sg40 yes	ir	Codir Trans Junc	ng Sequence script tion Site	>	chr19:285 chr19:285 chr19:285	2043-286 1895-286 5662-285	9287(+) 9474(+) 6002(+)				Codin Trans Junct	g Seque cript ion Site	nce	chr19:113 chr19:113 chr19:113	29349-11336 27061-11337 30155-11330	055(+) 374(+) 430(+)				
1			1567	rj41	NA	sg2422 yes	ir	DNA	tion Position		chr19:285	6002				Event	Sequen	ion Posit ces	ion	chr19:113	30155					
		1570	rj44	NA	sg2422 yes	ir	<pre>>ZNF57_+ chr19:2851895,28560024LPPR2_+_chr19:11330155,1133 7374_DNA(+) TrGGAGGCCCCGAGGGGGGGGGGGGGGGGGCCTGCCTGCCGCGCGCG</pre>									56002&LPPI de) TGTTTTTGT	IGTTIGTTIGTTTTTTGAG									
		1585	rj41		ig2422 yes	11										560024LPPR2_+_chr19:11330155,1133										
a.2			Showing 1 to 10 of 48 entries					Predicted Protein >ZNF57 + chr19:2851895,28560024LPPR2 + chr19:11330155,11337374 mRNA(+)_to_protein MSVLGIVILLAYRLEFTDTFPVHT0GFFCYDSTYARPYPGPEAASRVPPALVYALVTAGPTLTILLGELARAFFPAPPSAVPVIGESTIVSGACCRFSPPVRRLVRFLGVYSFGLFTTT														TT				
				All breakpoints:						51895-24	352046(+)	4 Exor	IS PORPGPD	REVIDQ	GACAGS	PSLVAA	1	chr19	:11327061	-11327248(10 Exons	CPSLCLAI	LCPAFLV	SVVRVAET	PCN	
U 8 4								3 4	2 chr19:2866519-2866646(+) 2 chr19:11328477-11328652(+) 3 chr19:2867075-2867247(+) 3 chr19:1329335-11329415(+) 4 chr19:2867921-2869474(+) 4 chr19:11331207-11331986(+)																	
			Show 10 v entries						6 chr19:11332692-11331632(*) 6 chr19:11332692-1133164(*) 7 chr19:1133488-11334365(*) 8 chr19:11335488-11335495(+)																	
	·····		GG.II	n Related Junctions	Associated Junctions	Shared Ge Genes Misn	ne atch		-								9 10	chr19 chr19	:11335836 :11336022	-11335888(-11337374((+) (+)					
		· · · · ·	70 103	rj1674 rj12	NA NA	g1870 NA g13 NA	ir ir	Show	10 - ontrior														Soarch:		^	
°			104	rj13 rj14	NA NA	g14 NA g15 NA	ir ir	Junct	ion Related	Associated	Shared	Gene	Single Event	Fusion	Left	Right	Left	Right	Gene	Gene	Gene	Gene	Gene	Gene	Gene	
			109 195	rj15 rj92	NA NA	g16 no g97 NA	ir	CG. 70	ID Junctions	Junctions NA	Genes sg1870	Mismatch NA	inversion	Gene	in CDS NA	in CDS NA	in Exon	in Exon NA	Left.name2	Right.name2	Left.name NA	Left.chrom NA	Left.strand NA	Left.txStart NA	NA Left.txEnd	
12			240	rj145	NA	sg110 no	in	103	rj12	NA	sg13	NA	inversion	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
			241 242	rj151	NA	g150 NA g159 no	in	104	rj13	NA	sg14 sg15	NA	inversion	NA	NA	yes	NA	no	<u>NA</u>	OPHN1	NA	NA	NA	NA	NA	
			244	rj119	NA	g160 yes	in	109	rj15	NA	sg16	no	inversion	opposing	yes	yes	no	no	IL1RAPL1	IL1RAPL1	NM_014271	chrX	+	28515601	2988393	
			Showin	g 1 to 10 of 2,433	entries		1	240	rj92 rj145	NA	sg97 sg110	no	inversion	opposing	yes	no	no	no	TMPRSS2	TMPRSS2	NA_005656	chr21	- NA	41758347	4180195	
			•					241	rj151	NA	sg156	NA	inversion	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
							-	242	rj154	NA	sg159	no	inversion	opposing same	yes	yes	no	no	APP	APP	NM_201413	chr21	-	26174731	2646500	
								Showi	ng 1 to 10 of 2 4	33 entries	sg160	yes	interchromosomal	orientation	yes	yes	no	yes	UNK	LSM12	NM_001080419	cnr17	*	Previous	7133348 Next -	
					4	0 10 01 2)1																Þ				

Can also generate weblinks from column values by specifying url prefix and suffix (e.g. from column with gene names, create links to gene cards by prepending fixed url to all cells)

ilter and Sort

dmin iPER tes

PDF

Specify PDF file from history.

Depending on browser settings this will be shown in page itself or a link will appear to download or view file in new window



Links

Links to web addresses, files from history, or all files in an archive in the history can also be created



chr05 chr06 chr07 <u>chr19 chr20 chr21 chr22 chrX chrY</u>

Tools used: CGATools, Circos, GNUplot, R, OncoFUSE, OncoSNPSeq

TraIT Galaxy HPC CLOUD Architecture

The rapid evolution of NGS technologies together with decreasing cost are creating a challenge to store and analyze the vast amount of sequencing data that are generated by experimental biologists. Configuring suitable data analysis software and having access to readily available computation and storage are the two major bottlenecks faced by many research groups

