

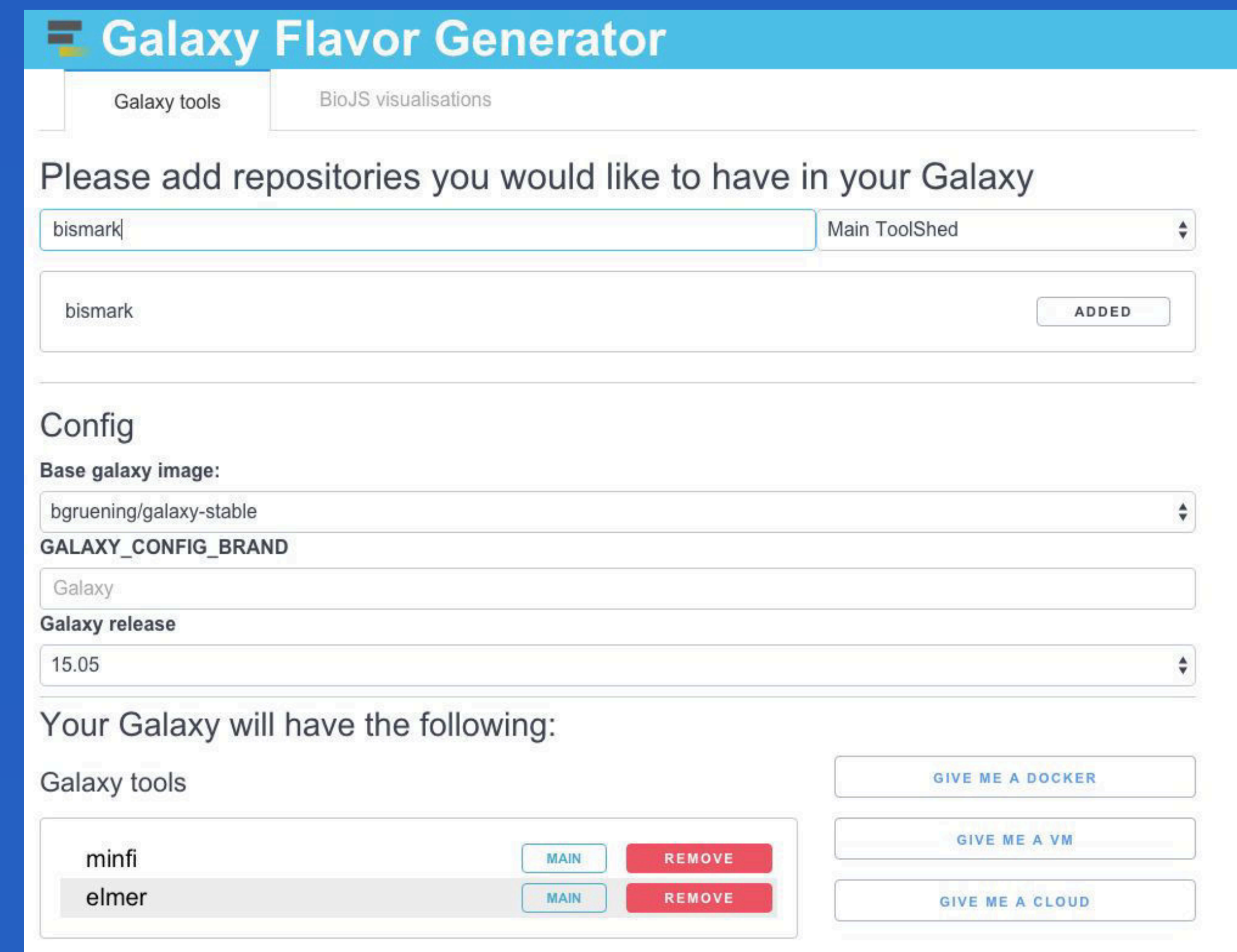
Galaxy Methylation Toolkit as a Galaxy Flavor

Nitesh Turaga¹, Enis Afgan¹, Benjamin Berman^{2,3}, James Taylor¹ and The Galaxy Team

1 Johns Hopkins University, Department of Biology. 2 University of Southern California, Keck School of Medicine
3 Cedars-Sinai Medical Center, Department of Biomedical Sciences

What is a Galaxy Flavor?

- Galaxy Flavor: an instance of Galaxy with a toolkit that has been tailored for a specific purpose
- Each Flavor's toolset is defined in a plain text file, which is then used to build a corresponding Galaxy instance, whether it is a local VM, a Docker based, or a cloud-based instance
- You can generate your own galaxy flavor at: bit.ly/galaxy-flavor-generator



```
3 tools:
4 - name: 'column_maker'
5   owner: 'devteam'
6   tool_panel_section_id: 'textutil'
7   revisions:
8 - '08a01b2ce4cd' # v1.1.0
9 - name: 'tabular_to_fasta'
10  owner: 'devteam'
11  tool_panel_section_id: 'convert'
12  revisions:
13 - '0b4e36026794' # v1.1.0
14 - name: 'fasta_to_tabular'
15  owner: 'devteam'
16  tool_panel_section_id: 'convert'
17  revisions:
18 - '9d99908f2ad' # v1.1.0
19 - name: 'fastqtofasta'
20  owner: 'devteam'
21  tool_panel_section_id: 'convert'
22  revisions:
23 - '3571553aeb20' # v1.0.0
24 - name: 'dna_filtering'
25  owner: 'devteam'
26  tool_panel_section_id: 'filter'
27  revisions:
28 - 'a6f0d355b05f' # v1.0.3
29 - name: 'subtract_query'
30  owner: 'devteam'
31  tool_panel_section_id: 'group'
32  revisions:
33 - '5f6ebef89722' # v0.1
```

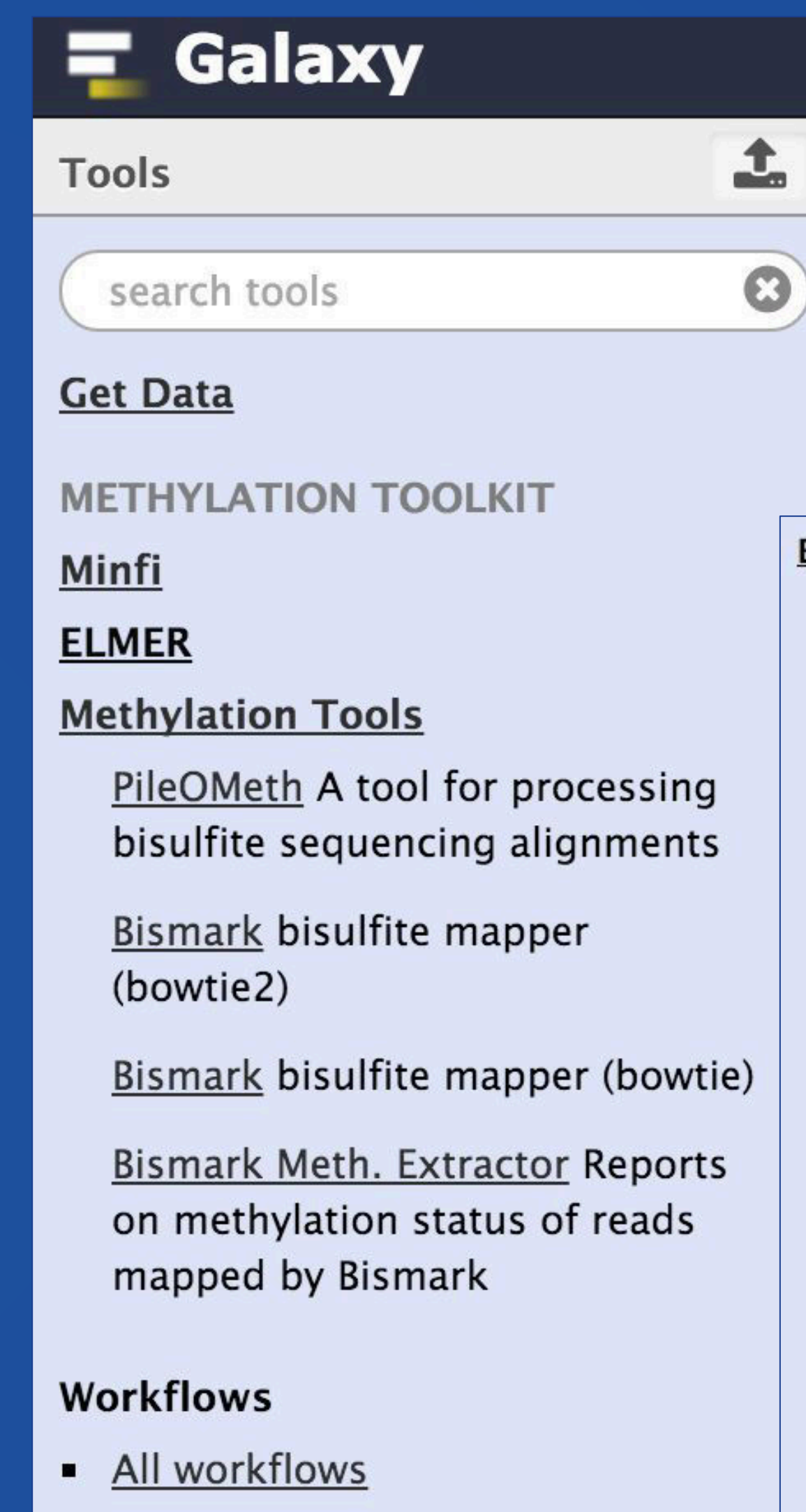


New to Galaxy: Methylation toolkit

- **Minfi**: Tools for analyzing and visualizing Illumina's 450k array data
- **ELMER**: Inferring Regulatory Element Landscapes and Transcription Factor Networks using Cancer Methylomes
- **PileOMeth**: A (mostly) universal methylation extractor for BS-seq experiments
- **Bismark**: map bisulfite converted sequence reads and determine cytosine methylation states

New tools allow access to TCGA data from within Galaxy:

- Use Minfi & ELMER for analyzing methylomes of cancer datasets available on TCGA
- Both, sequencing and array-based tools are available on Galaxy



ELMER

ELMER-TCGA

- Download TCGA data DNA Methylation, Expression and Clinical data for all samples of certain cancer type from TCGA
- Complete ELMER analysis pipeline for a TCGA data
- Download TCGA 450K Methylation data for a cancer type
- Download TCGA Clinical data for a cancer type
- Download TCGA RNAseq data for a cancer type

ELMER-CUSTOM DATA

- ELMER pipeline distal enhancer DNA methylation analysis and predict putative target genes, motif analysis and identify regulatory transcription factors

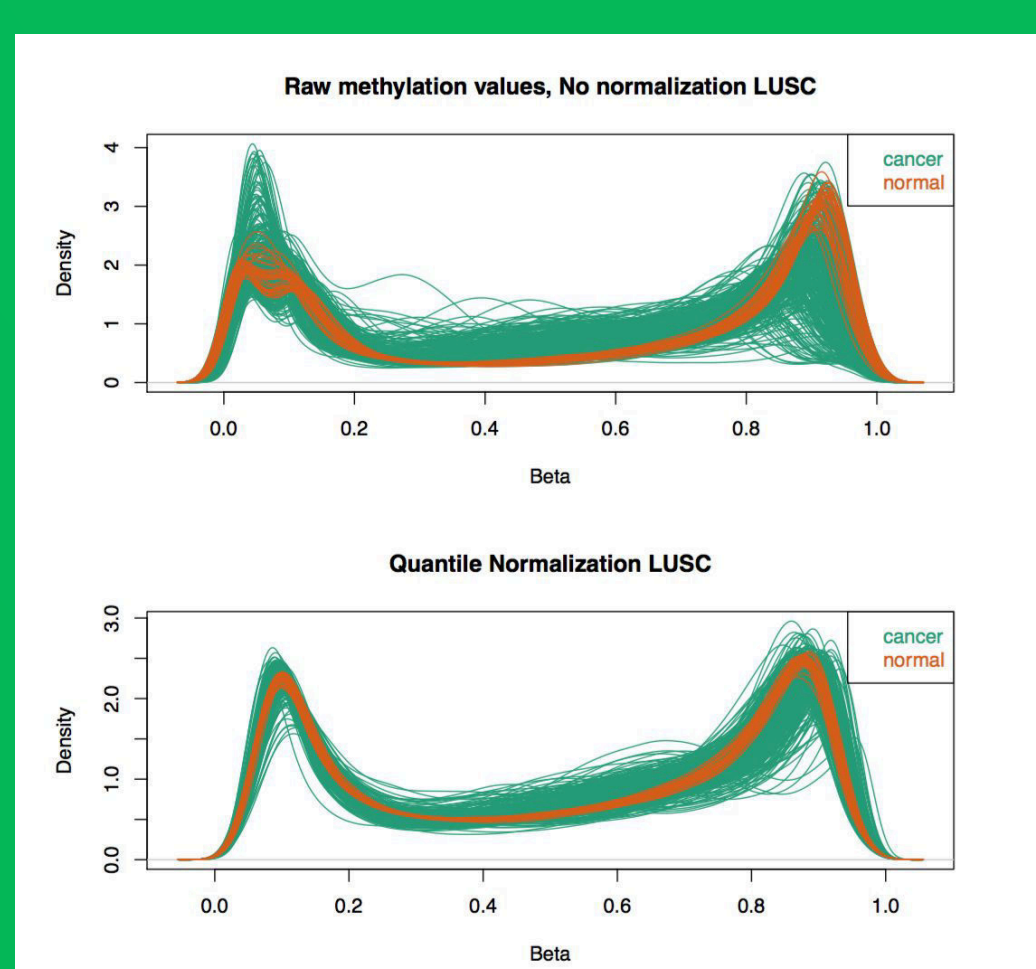
METHYLATION TOOLKIT

Minfi

- Unarchive files
- Read TCGA Illumina 450k data in IDAT format from GDAC-Broadinstitute
- Read Illumina 450k IDAT format to get Beta Values
- Density Plot of 450k data beta values
- QC report of Illumina Human Methylation 450k arrays
- Cell Proportion Estimation across all samples and cell types for 450k data
- Multi dimensional scaling(MDS) plots to an overview of similarities and differences between samples
- Differentially Methylated Position Finder to identify CpGs
- Bumphunter to estimate differentially methylated regions

Analysis of TCGA LUSC Illumina Methylation 450k Array and Expression Dataset

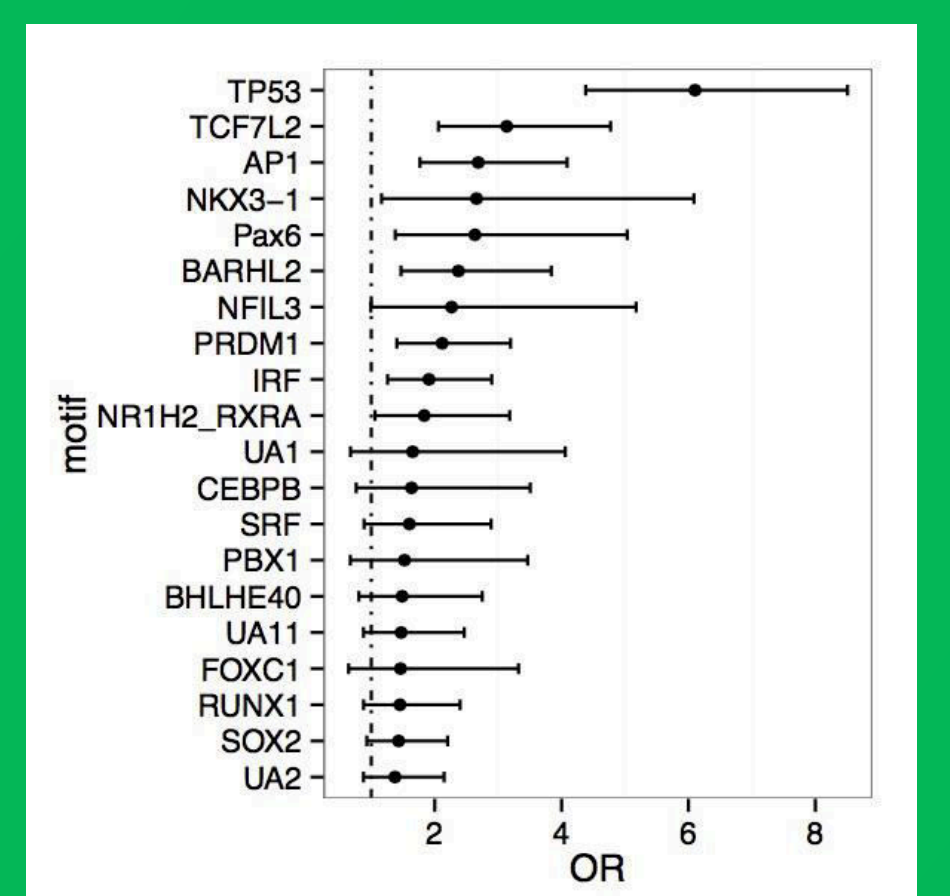
Step 1 - Quality Control:
Density plot of methylation 450k arrays using Minfi



Step 2 - Differential Methylation analysis and Annotation

chr	start	end	cluster	value	strand	area	clusterL	p.value	gene_name
chr2	63279495	63297208	107365	1.670508404	+	66.73664495	41	0	OTX1
chr12	115134291	115136308	92018	1.76243499	+	60.64461177	74	0	TBR3
chr7	27140787	27145159	175203	1.474682829	+	57.50972093	39	0	HOKA2
chr7	27183401	27185282	175222	1.301186802	-	55.95103247	53	0	HOKA6
chr2	176865174	176899586	114073	1.849534093	+	51.78975462	28	0	HOKD9
chr7	27203430	27209807	175231	1.651364759	-	45.39522974	27	0	HOKA9
chr13	100640914	100644957	96561	1.878575343	+	43.64295891	26	0	ZIC2
chr6	29520752	29521803	163140	1.144302945	-	43.48351189	40	0	URB
chr7	27224795	27226329	175244	1.482110129	+	42.98119375	41	0	HOKA11
chr5	2753335	2755361	160233	1.65830007	+	42.74901651	36	0	C5orf8
chr15	89949617	89953370	70551	1.984483112	+	41.67414535	21	0	LINC00925
chr17	4665164	46656660	86619	1.768230747	+	38.90107842	22	0	HOKB4
chr11	86382900	86383940	38865	1.806421409	+	37.93484958	21	0	ME3
chr13	37034536	37030299	54130	1.769526523	+	37.16215586	39	0	CCNA1
chr7	27195602	27198025	175227	1.675759994	-	36.86671986	27	0	HOKA7
chr14	57275867	57278669	60073	1.81847604	-	36.38952081	20	0	OTX2
chr3	147109428	147113092	38699	1.45206299	-	36.30015748	25	0	ZIC4
chr1	91182138	91186922	10129	1.273858492	-	35.7150519	26	0	BARHL2
chr6	30094947	30099802	163322	1.177714111	+	35.33142333	30	0	TRIM40

Step 3 - Identify and plot regulatory TFs whose expression associates with TF binding motif DNA methylation (ELMER)



New Features in Galaxy

- Easy fetching of TCGA data (Methylation, RNA, Clinical)
- Ability to recognize .RData files and use them within Galaxy as a History dataset
- Future: galaxy-epigenetics-workbench - a Galaxy Docker repository for epigenetic data analysis

Want to contribute?

- Issues and tool requests can be filed at github.com/galaxyproject/tools-iuc/issues
- Pull requests are welcome github.com/galaxyproject/tools-iuc
- github.com/galaxyproject

Acknowledgements

- Dataset Collections (John Chilton)
- PileOMeth (Bjoern Gruening)
- Galaxy Flavor Generator (Martin Page, Bjoern Gruening)
- Galaxy Team