



INTEGRATING MULTI-OMIC DATA ON THE CLOUD YIELDS INSIGHTS INTO RETROTRANSPOSON ACTIVITY IN CANCER

WILSON MCKERROW, PHD (NYU SCHOOL OF MEDICINE)

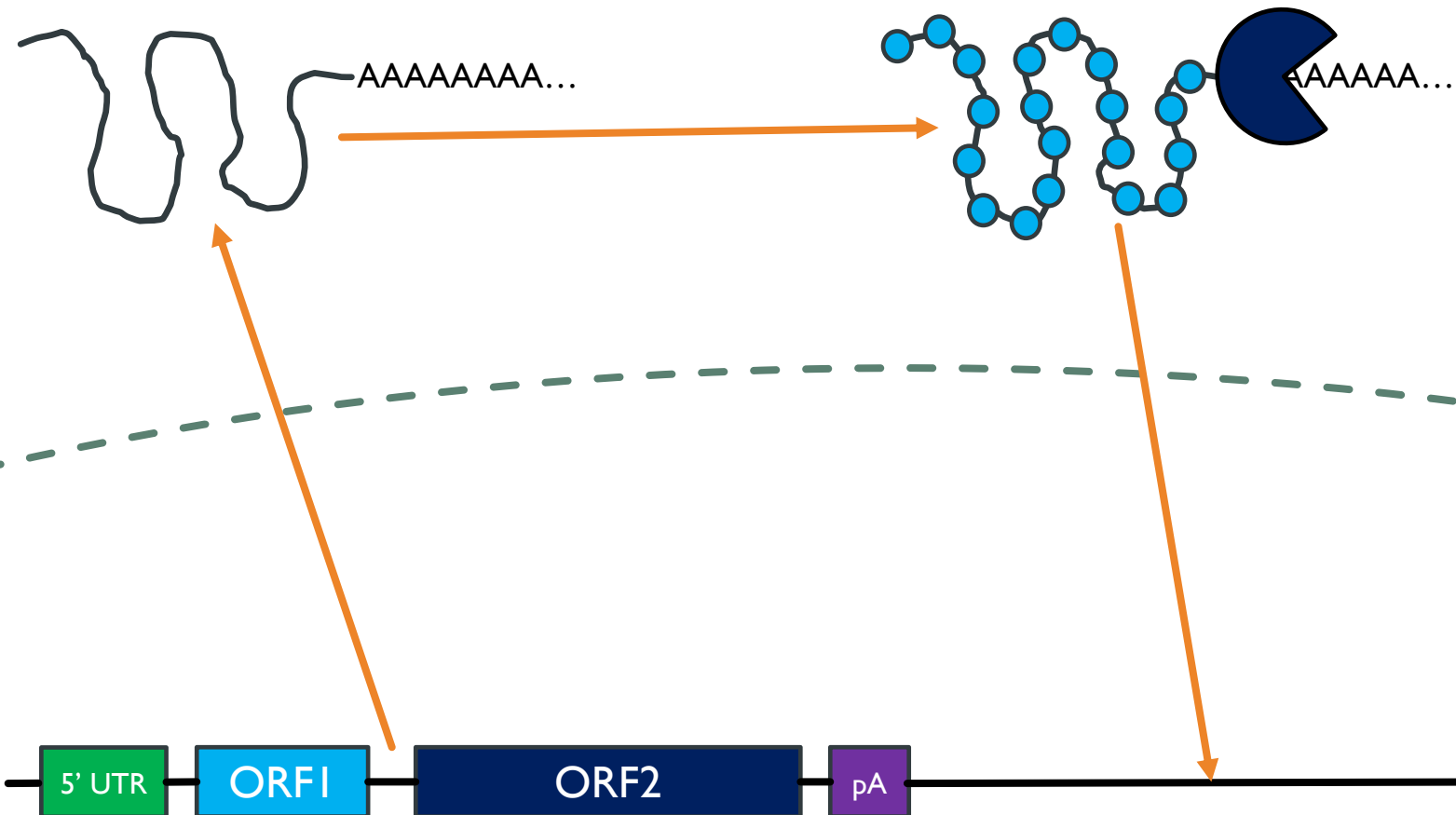
MANISHA RAY, PHD (SEVEN BRIDGES GENOMICS)



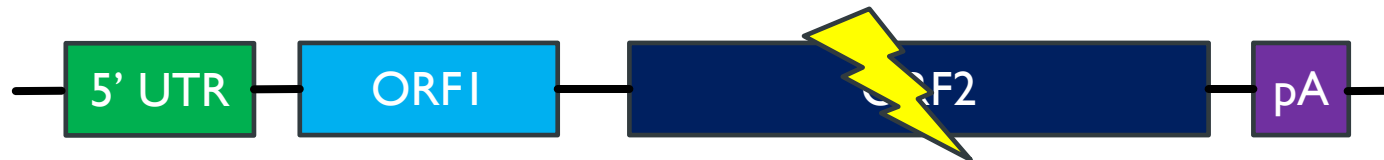
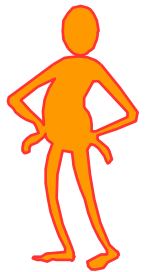
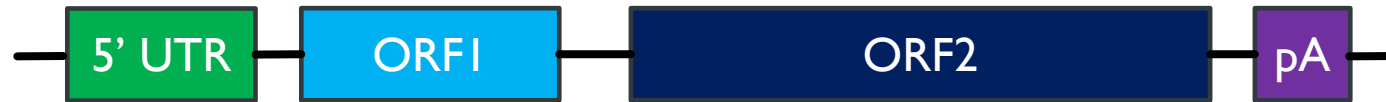
OUTLINE

- Background
- Methods
- Demo

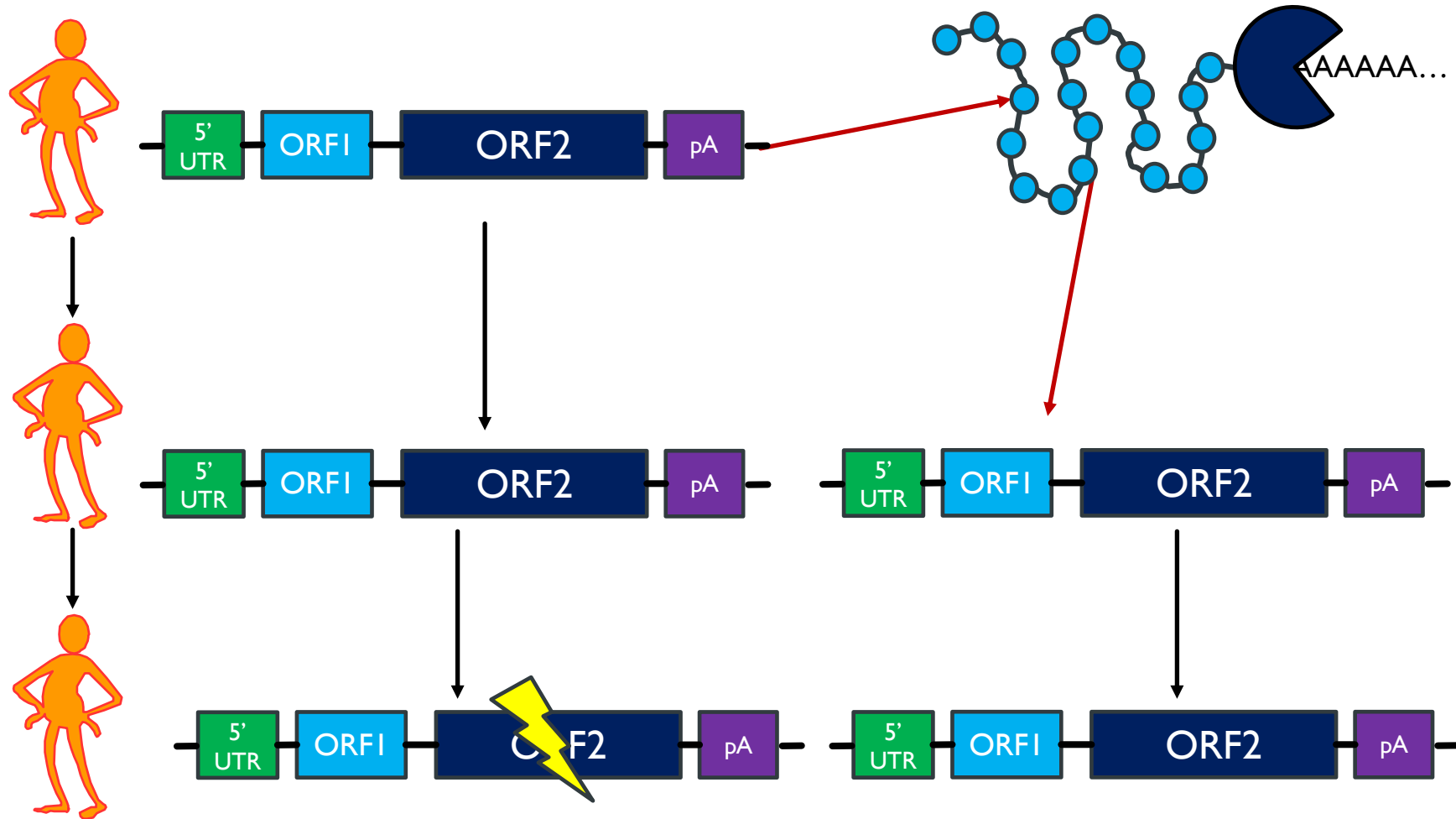
LINE-1: THE ONLY AUTONOMOUS RETROTRANSPOSON IN THE HUMAN GENOME



RETROTRANSPOSONS ACQUIRE SILENCING MUTATIONS

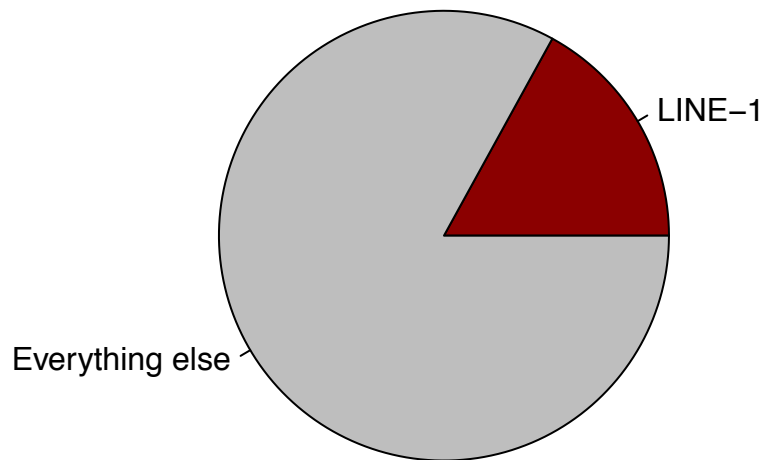


RETROTRANSPOSON MUST COPY THEMSELVES TO SURVIVE

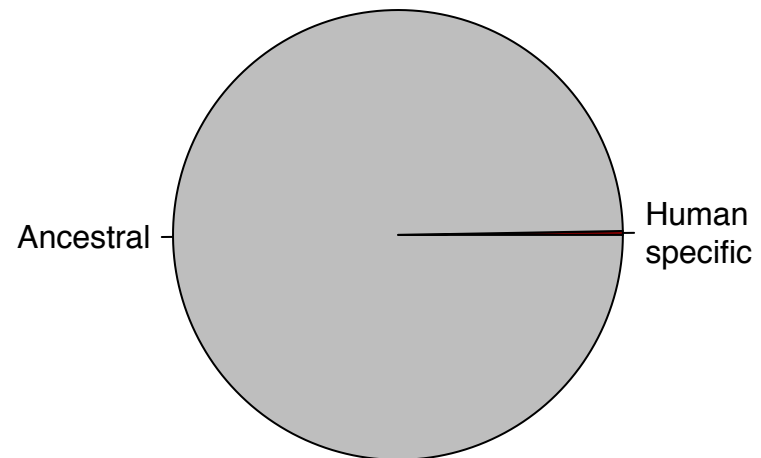


HUMAN LINE-1 ELEMENTS

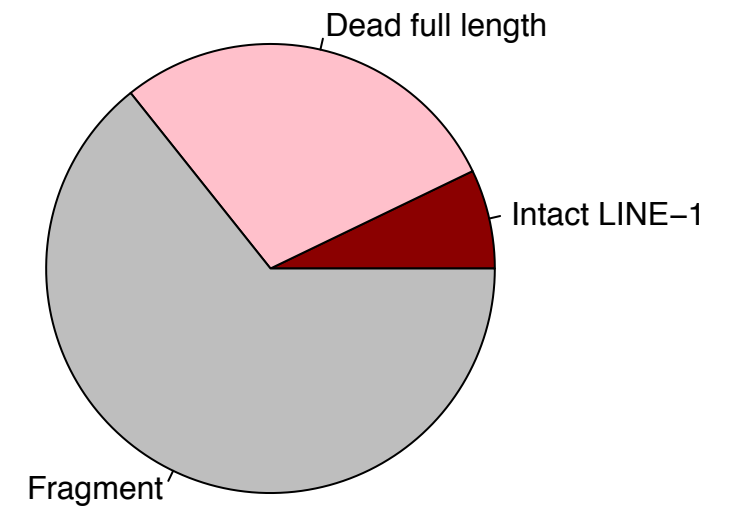
The human genome



L1 insertions



Human specific LINE-1



LINE-1 EXPRESSED IN MANY CANCER (ORFIP IHC)

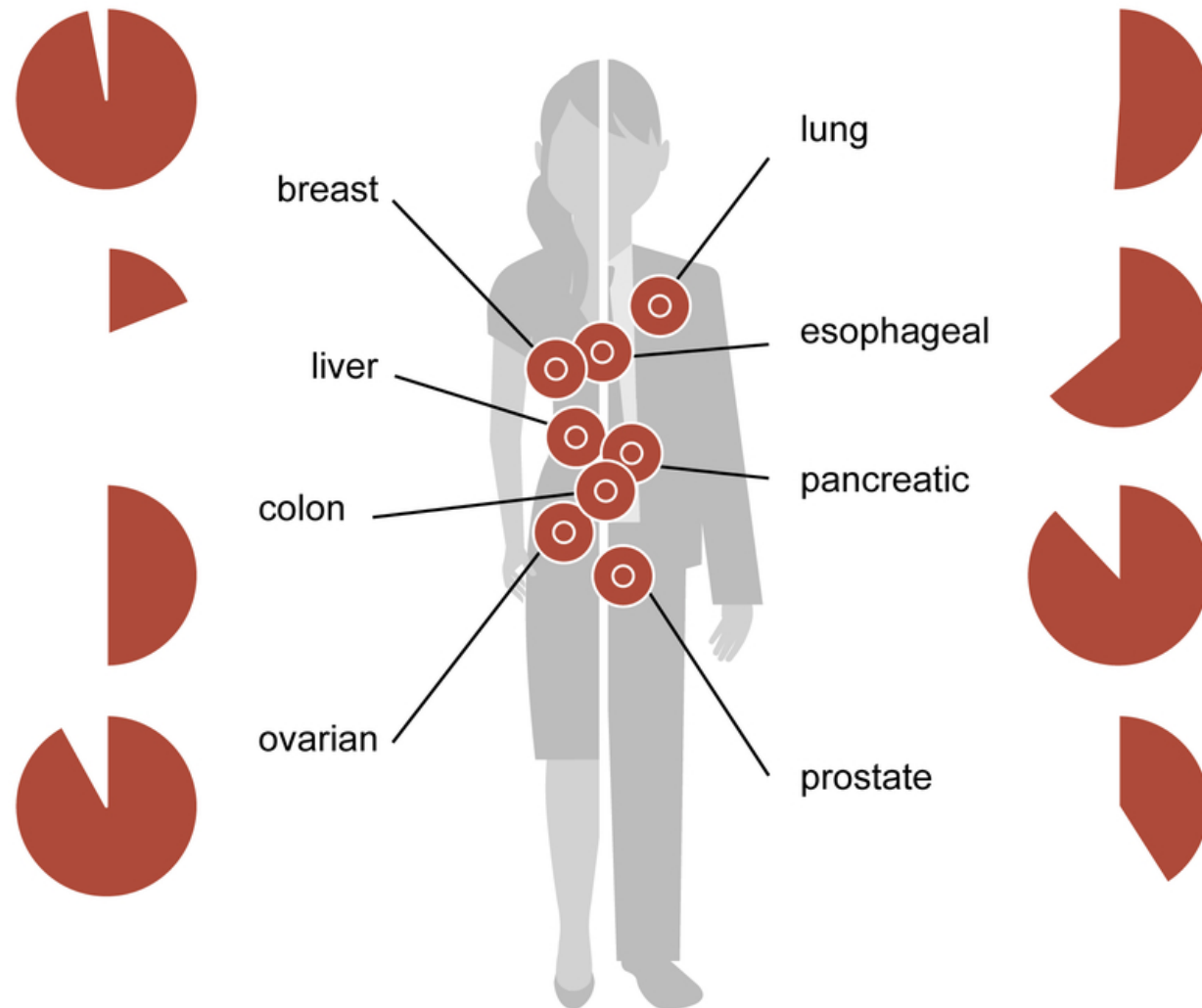
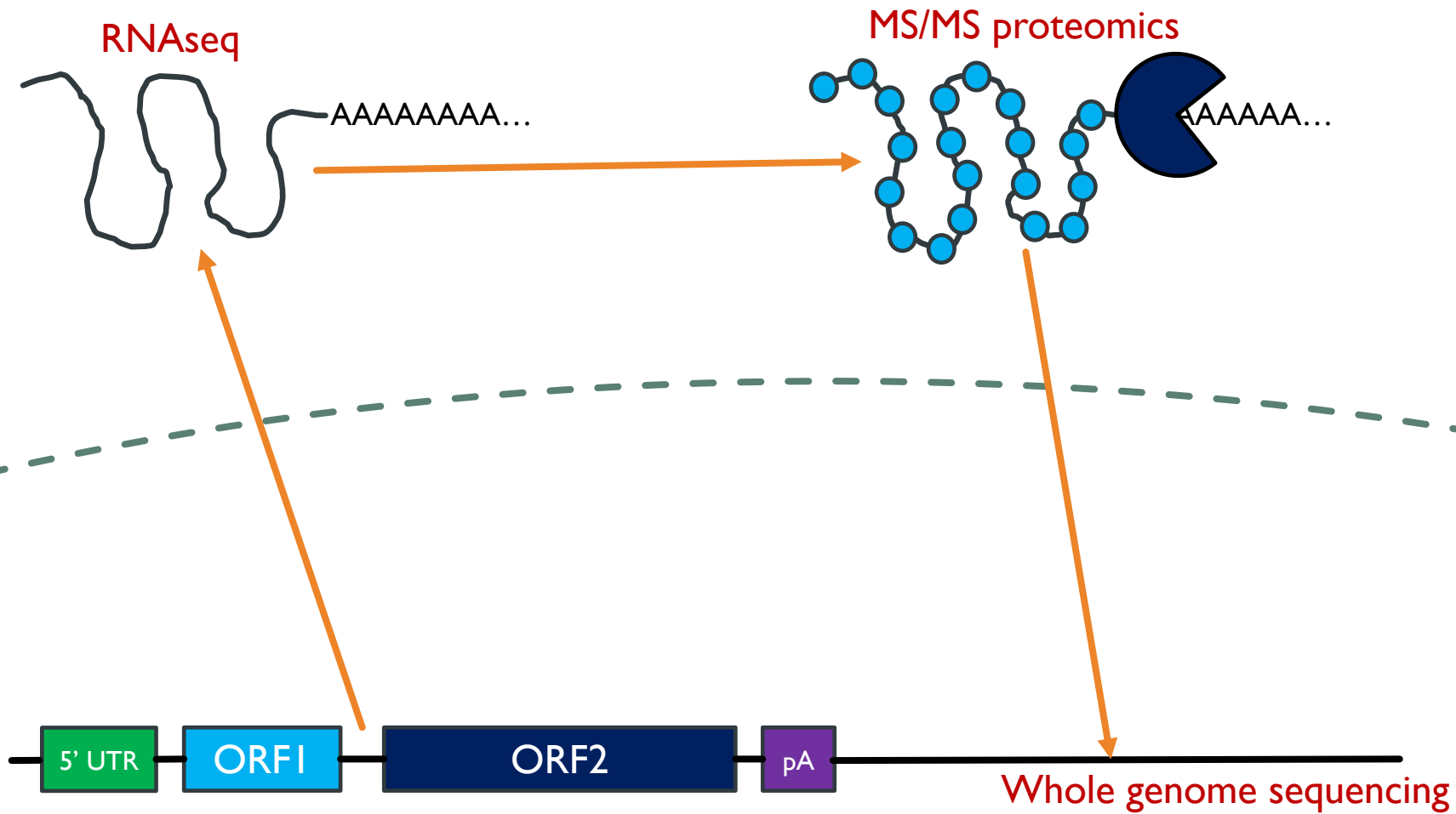


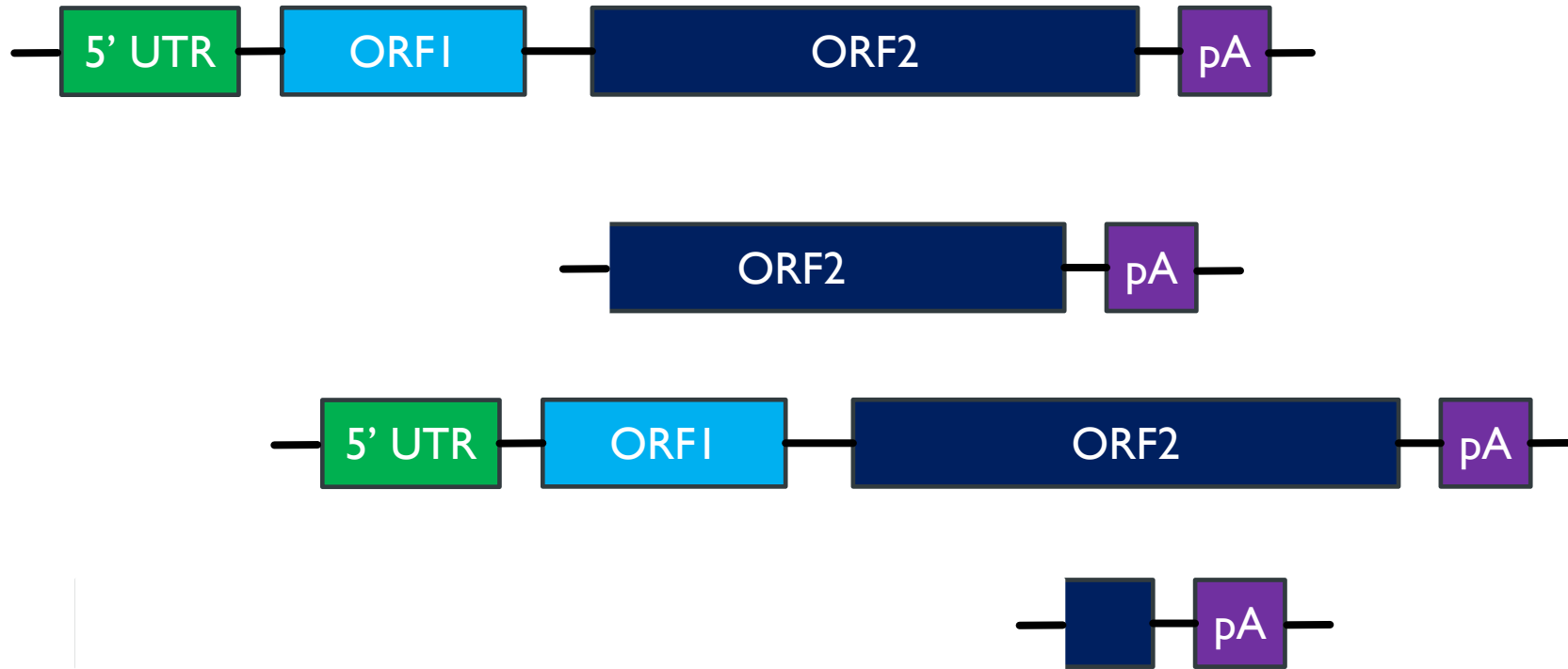
Figure: 28188229
Ardeljan, Taylor, Ting, Burns

FIRST GOAL: MEASURE LINE-1 AT EACH STAGE



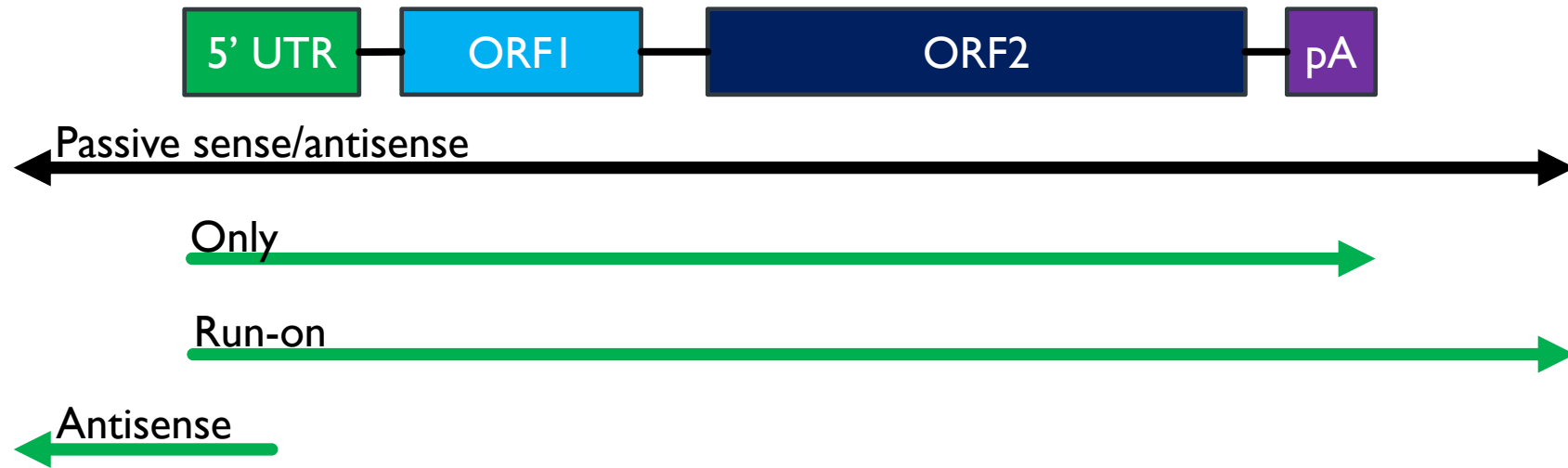
LINE-1 RNA QUANTIFICATION BY LIEM

I. Identify potentially transcribed loci



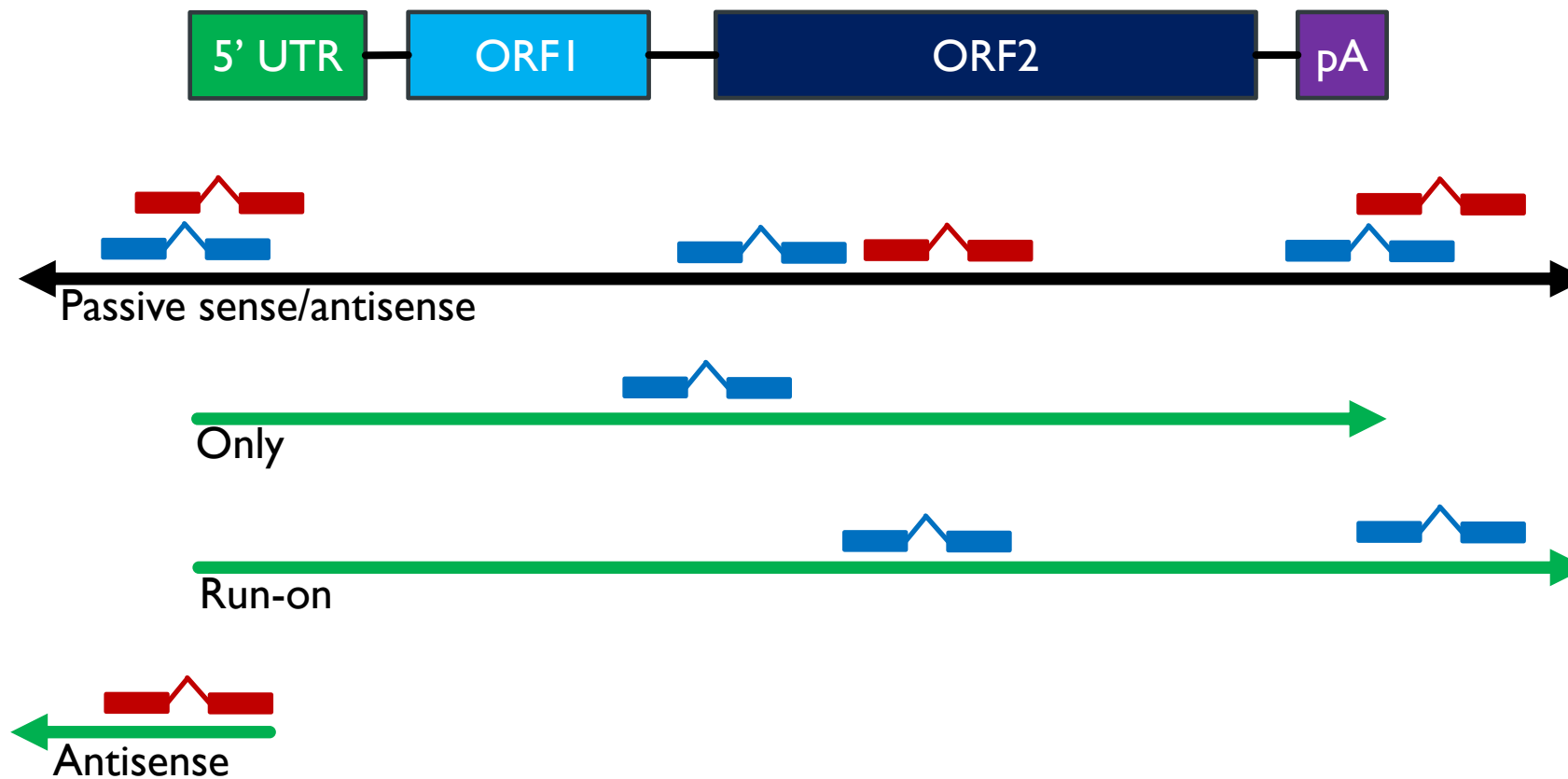
LINE-1 RNA QUANTIFICATION BY LIEM

2. Identify the ways those loci might be transcribed



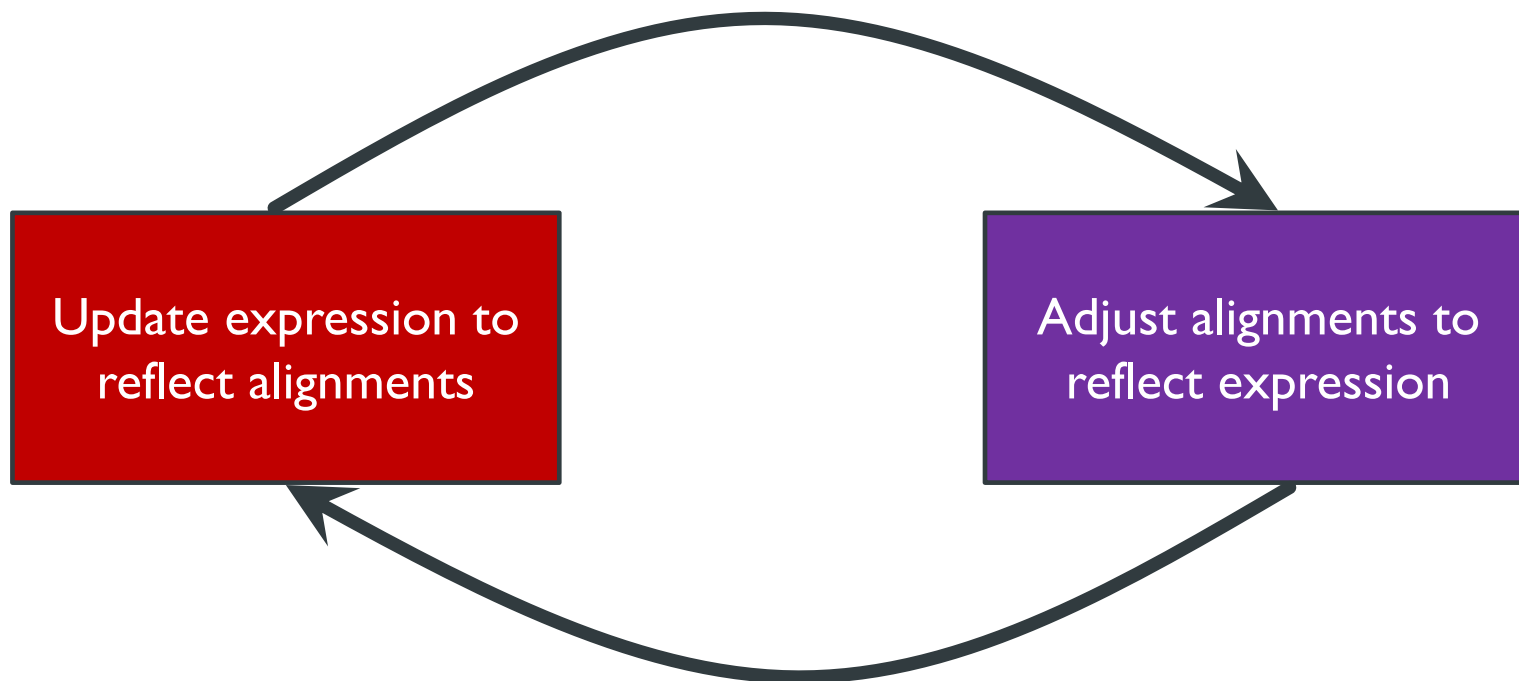
LINE-1 RNA QUANTIFICATION BY LIEM

3. Find the reads that might support those transcripts

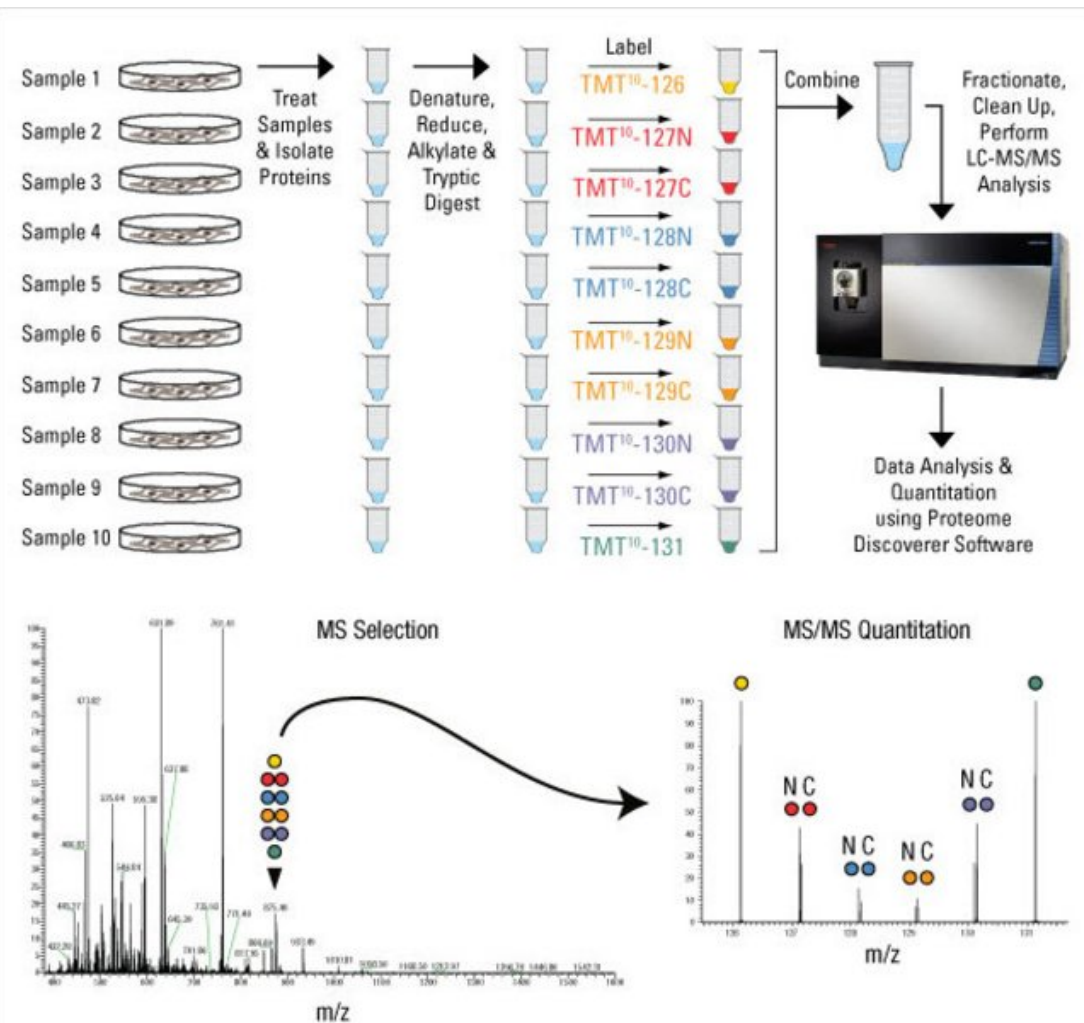


LINE-1 RNA QUANTIFICATION BY LIEM

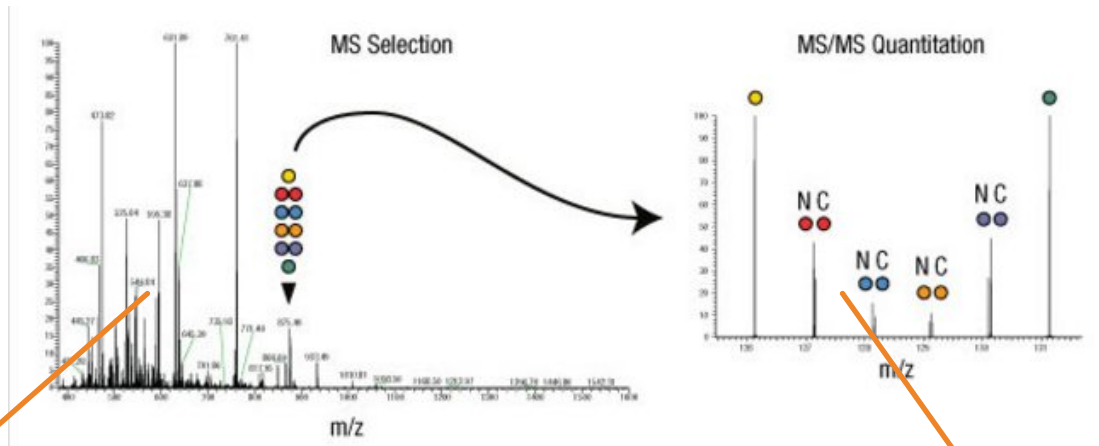
4. Quantify transcription by expectation maximization



LINE-1 ORF1P QUANTIFICATION FROM MS/MS



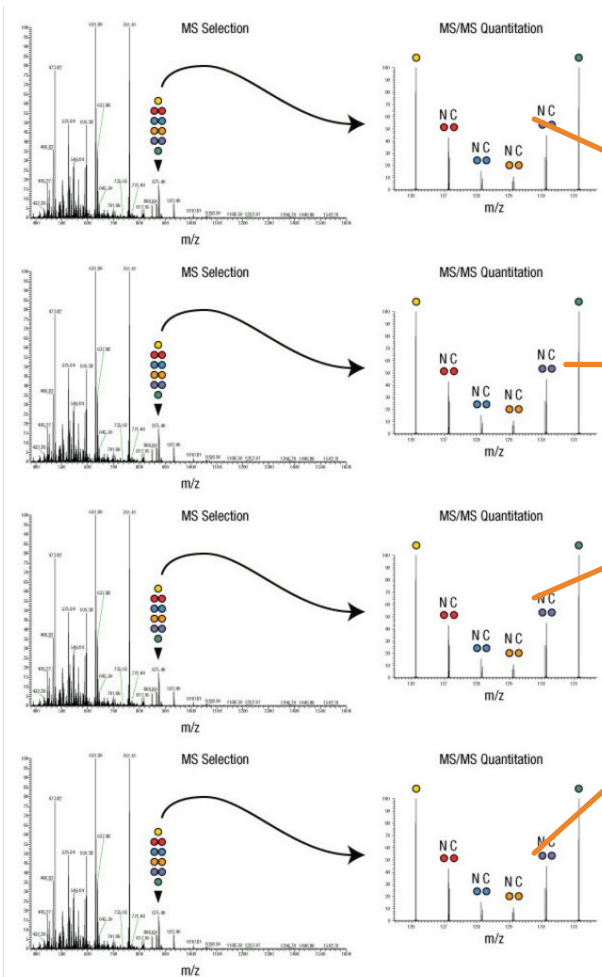
LINE-1 ORF1P QUANTIFICATION FROM MS/MS



Search peptide against human+LINE1 proteome using x!tandem

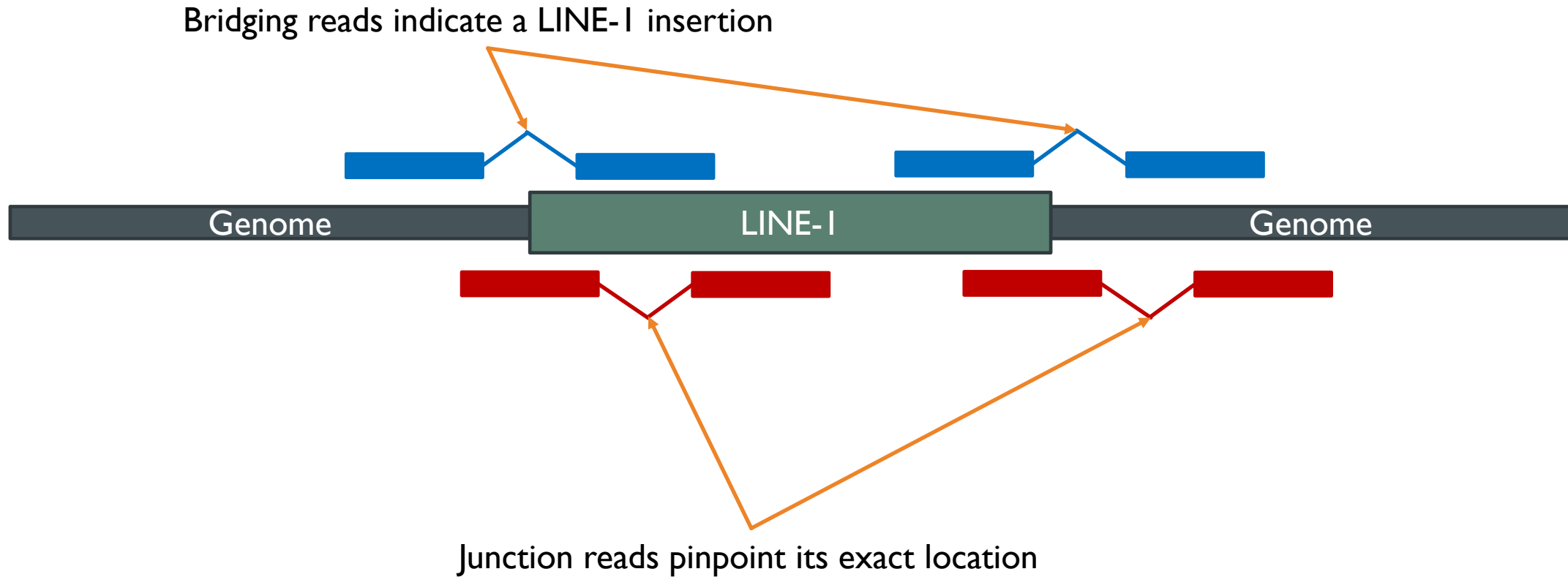
Relative quantification from peak height label

LINE-1 ORF1p QUANTIFICATION FROM MS/MS



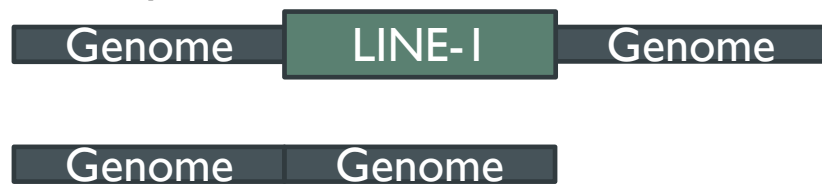
Quantify ORF1p from 10 diagnostic peptides

LINE-1 INSERTION QUANTIFICATION FROM WGS BY MELT

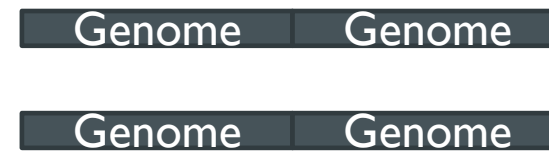


LINE-1 INSERTION QUANTIFICATION FROM WGS BY MELT

Primary Tumor



Blood Derived Normal



Somatic insertion

Two orange arrows originate from the bottom of the Primary Tumor and Blood Derived Normal diagrams and point towards the text "Somatic insertion".



DEMO TIME

ACKNOWLEDGMENTS

NYU School of Medicine

- David Fenyö
- Xuya Wang
- Mark Grivainis
- Jef Boeke
- Paolo Mita



Seven Bridges Genomics

- Manisha Ray
- Dave Roberson
- Ana Damljanovic



Case

Primary site
Uterus, NOS

Sample

Sample type
Primary Tumor

File

Data format
BAM

Experimental strategy
RNA-Seq

Analysis

Workflow type
STAR 2-Pass Genome

Add property

Properties and values

Find specific

Created datetime

Submitter ID

Updated datetime

Workflow link

Workflow version

Case 101 Sample 111 File 102 Analysis 102

Export Details Analytics

Analysis

0118a624-9f27-4c18-b563-e5930757340e
01411f97-06a2-463e-b16b-cd44905530ce
06626b34-8524-4690-b149-664ad9e52425
077-0015-1111-1111-1111-1111-1111-1111

Details for 0118a624-9f27-4c18-b563-e5930757340e

CPTAC-3

Created datetime 2019-03-29T15:17:40.227Z

Submitter ID 94b14697-41af-4573-af49-1ad776b4a8a6_star_2-pass_genome

Connections

Inbound: **File** 1

94b14697-41af-4573-af49-1ad776b4a8a6.ma_seq.genomic.gdc_realn.bam

Outbound:

No outbound connections

FILTERS

GENERAL

Study
 Select one or more studies

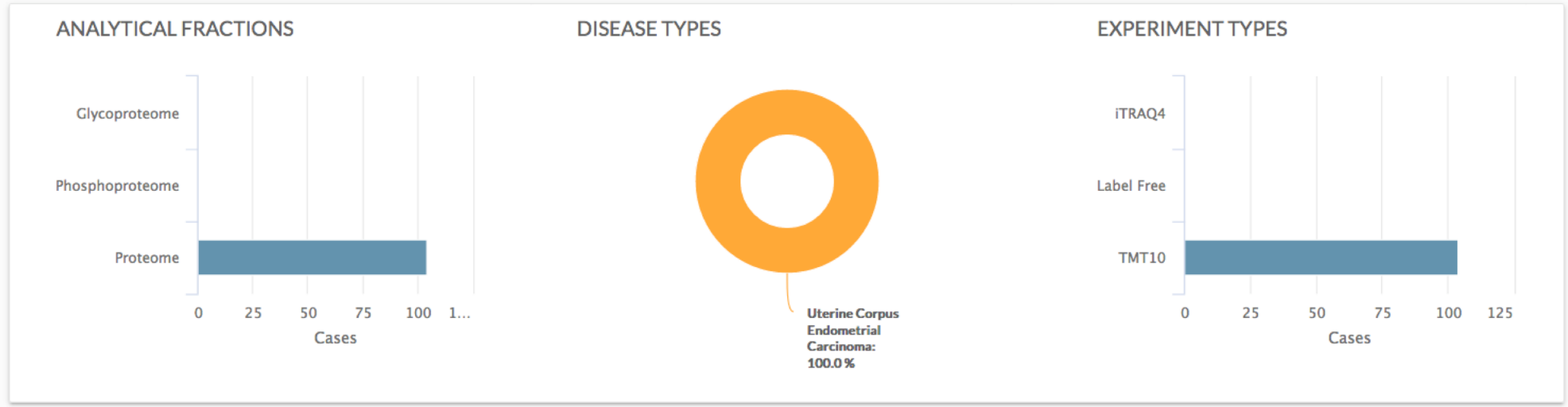
Data Category

Other Metadata (1 Study)
 Peptide Spectral Matches (1 Study)
 Processed Mass Spectra (1 Study)
 Protein Assembly (1 Study)
 Protein Databases (0 Study)
 Quality Metrics (1 Study)
 Raw Mass Spectra (1 Study)

FILE TYPE

Document (1 Study)
 Open Standard (1 Study)
 Proprietary (1 Study)
 Text (1 Study)
 Web (1 Study)

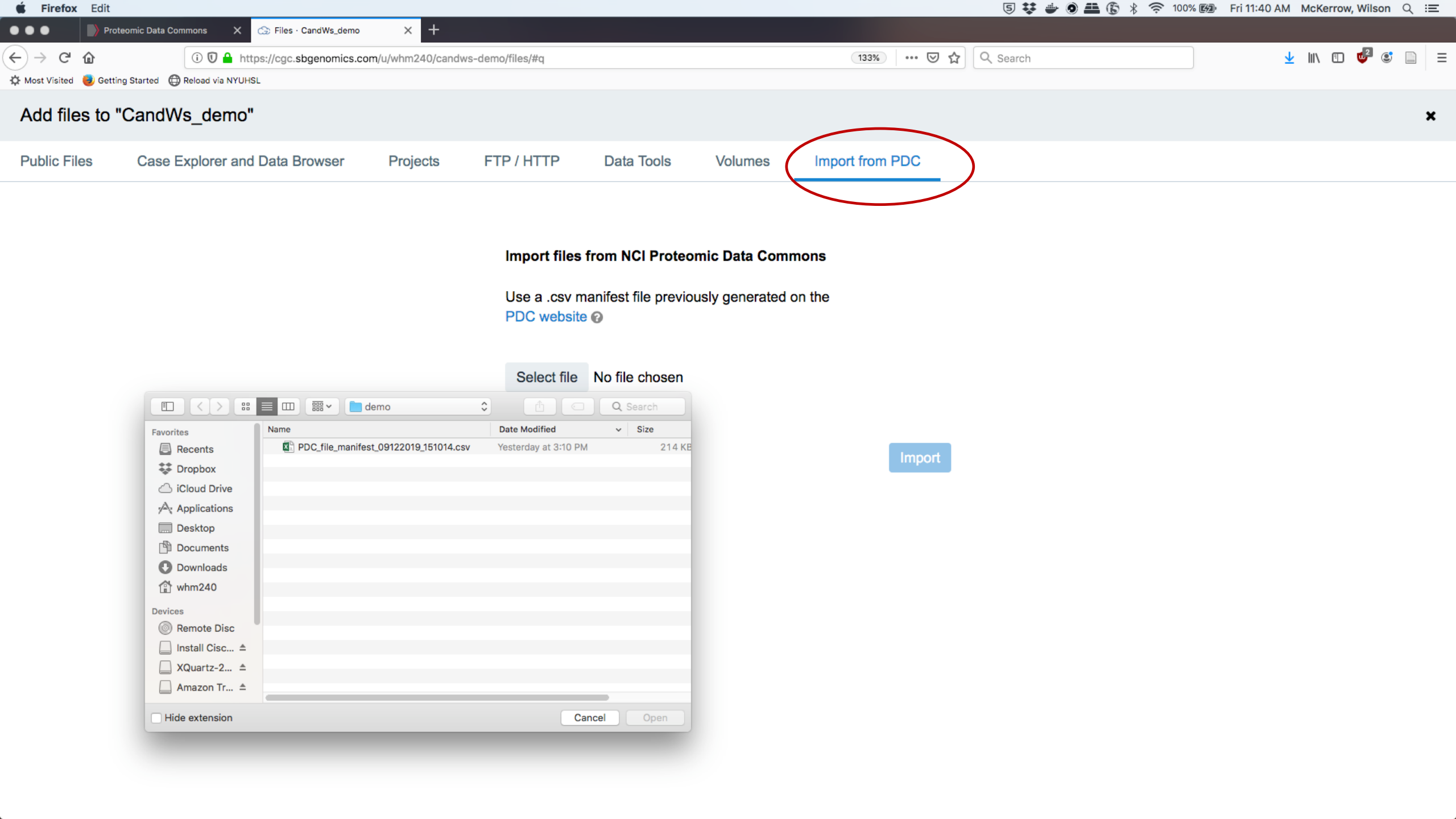
CLEAR project_name IS CPTAC3 Discovery AND primary_site IS Uterus, NOS AND analytical_fraction IS Proteome AND file_type IS Proprietary



Study (1) Biospecimens (135) Clinical (104) **Files (480)** Genes (10326)

Total files: 480
 Files data selected for download: 100 **Download File Manifest**

Study	File Name	Run Metadata ID	Project	Data Category	File Type	Access	File Size (MB)	Downloadable
CPTAC UCEC Discovery Study - Proteome	12CPTAC_UCEC_W_PNNL_20180222_B354_f18.raw	S043-1-14	CPTAC3 Discovery	Raw Mass Spectra	Proprietary	Open	889.323	Yes
CPTAC UCEC Discovery Study - Proteome	10CPTAC_UCEC_W_PNNL_20180222_B352_f16.raw	S043-1-12	CPTAC3 Discovery	Raw Mass Spectra	Proprietary	Open	856.096	Yes
CPTAC UCEC Discovery Study - Proteome	15CPTAC_UCEC_W_PNNL_20180503_B453_f17.raw	S043-1-18	CPTAC3 Discovery	Raw Mass Spectra	Proprietary	Open	773.441	Yes



Add files to "CandWs_demo"

- Public Files
- Case Explorer and Data Browser
- Projects
- FTP / HTTP
- Data Tools
- Volumes
- Import from PDC**

Import files from NCI Proteomic Data Commons

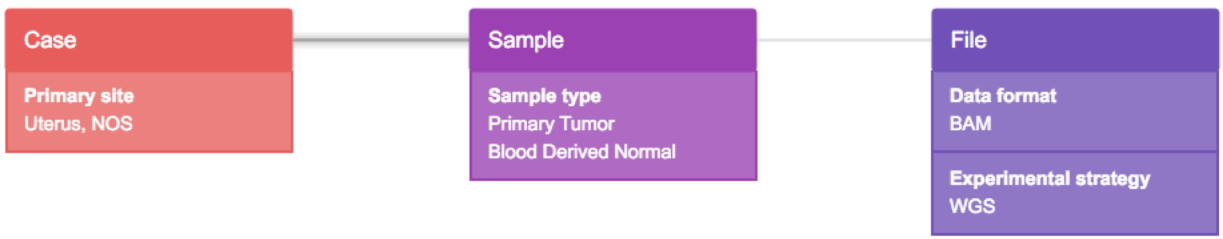
Use a .csv manifest file previously generated on the [PDC website](#)

Import

Select file No file chosen

Name	Date Modified	Size
PDC_file_manifest_09122019_151014.csv	Yesterday at 3:10 PM	214 KB

Cancel Open



Case 101 Sample 214 File 203

Export Details Analytics

Sample
020e5c88-a6ea-4d7b-a179-7b48030a8216
0255b77d-27b8-400f-8886-3fc4ac07f6f7
04df4034-7328-4b49-ab3c-33366a4d2ff4
0310040-00-1-11-000-110-70-1-70

Details for 020e5c88-a6ea-4d7b-a179-7b48030a8216

CPTAC-3

Created datetime 2017-09-04T15:34:12.411Z

Initial weight 0.415

Is FFPE false

CST embedded false

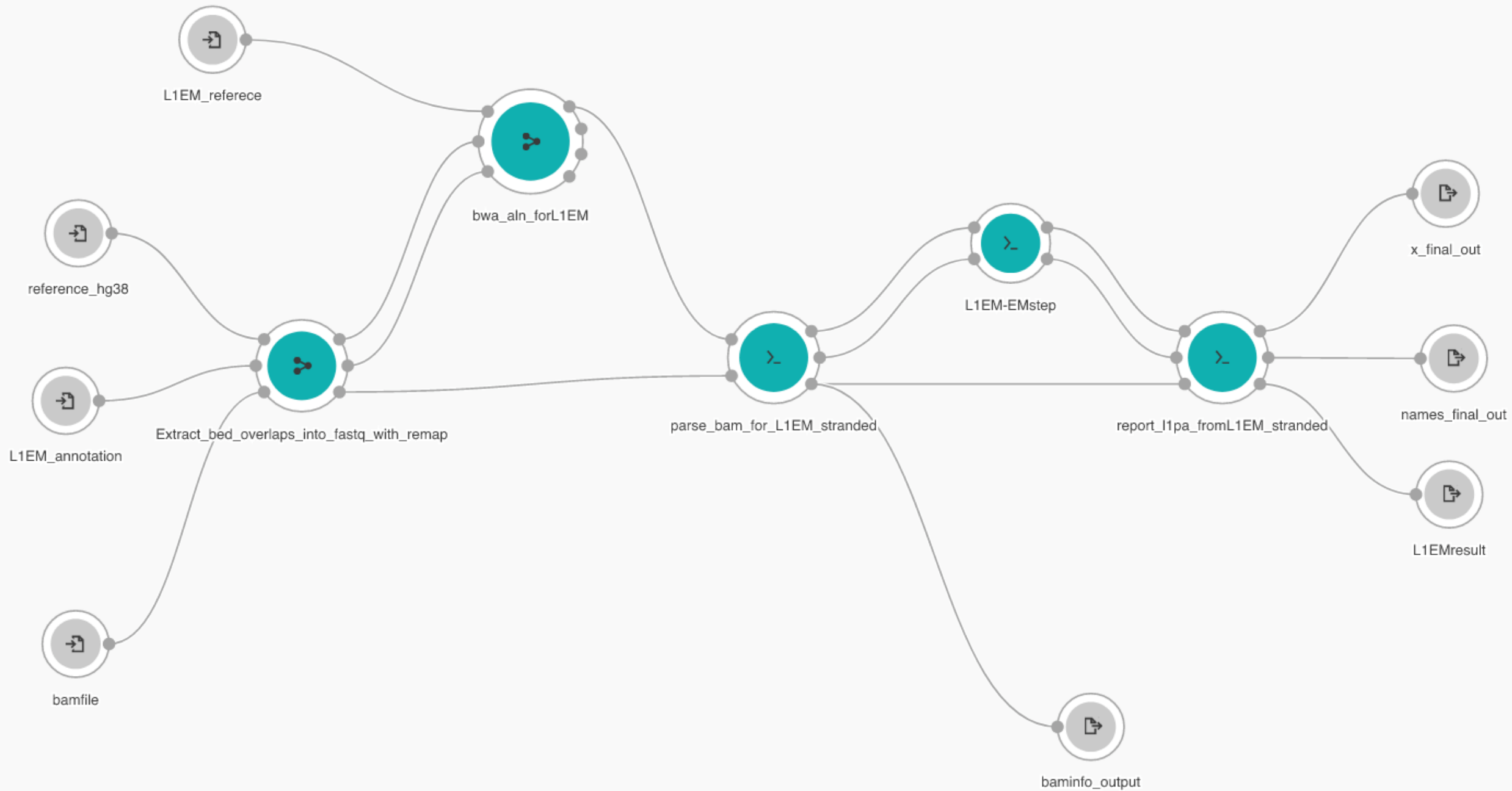
Connections

Inbound: Case 1

bc8f9b37-2f05-4508-830f-9356b99313dd

Outbound: File 1

d678d221-9dbd-4e9f-8f94-9aaef5f7092f_gdc_realn.bam



Firefox File Edit View History Bookmarks Tools Window Help

Proteomic Data Com X Apps · CandWs_demo X SB MELTCancerNormal X SB L1EM-WF-remap-str X Tasks · CandWs_demo X L1EM demo task · Tas X FenyoLab xtandem pi X MeltCancerNormal de X Explore Data | Data Bi X Fwd: Thought: A Jour X

https://cgc.sbggenomics.com/u/whm240/candws-demo/tasks/a8b5579f-dfe9-4651-81b7-3a43b709c477/#set-input-dat: 120% Search

Most Visited Getting Started Reload via NYUHSL

Projects Data Public Apps Public projects Developer whm240

Dashboard Files Apps **Tasks** **CONTROLLED** CandWs_demo Interactive Analysis Settings Notes

BATCH 102 **L1EM demo task**

Last update by whm240 on Sept. 13, 2019 09:44

App: L1EM-WF-remap-stranded - Revision: 0

Get support Discard **Run**

Task Inputs Execution Settings

Inputs

Batching On

▼ bamfile * Change selection

Batch by: File

This will create one task for each selected item.

- ▶ 019694c6-e25f-467c-ab7c-37c9dd33eb34.rna_seq.genomic.gdc_realn.bam (1 item)
- ▶ 045eced3-c565-4440-9d47-344b104e266c.rna_seq.genomic.gdc_realn.bam (1 item)
- ▶ 04bb5022-9b31-42dd-9d2d-0224e0abc2b7.rna_seq.genomic.gdc_realn.bam (1 item)
- ▶ 0573bb51-ce88-4bf3-9752-0e99b18a8b5e.rna_seq.genomic.gdc_realn.bam (1 item)
- ▶ 05ac292a-af75-43f6-860a-89ffbe8495a4.rna_seq.genomic.gdc_realn.bam (1 item)
- ▶ 06adca22-6143-4531-b4f1-628abf9a0af5.rna_seq.genomic.gdc_

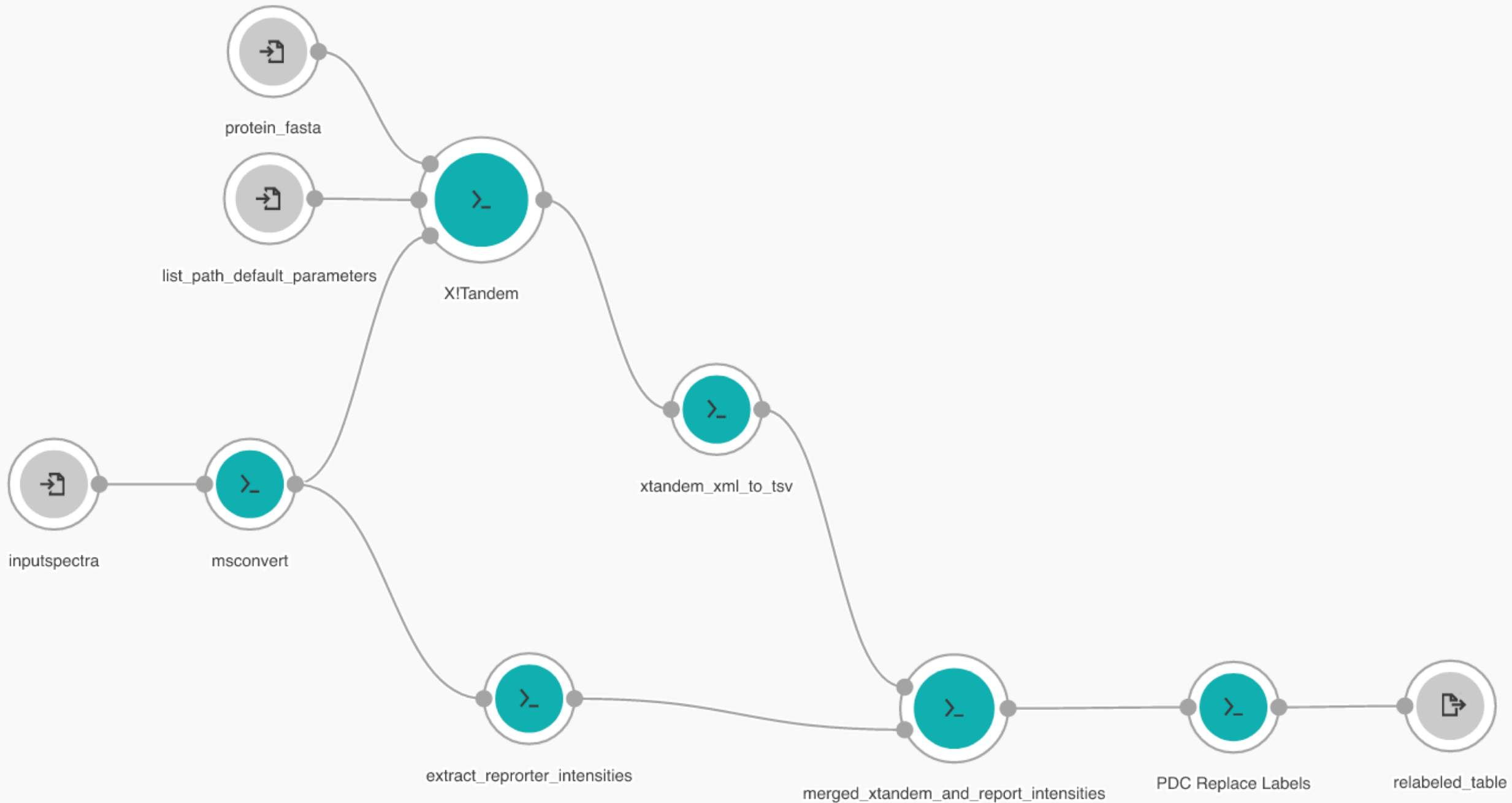
App Settings

Edit parameters Show editable ▼

There are no settings for this task

Outputs

L1EMresult	No value
baminfo_output	No value
names_final_out	No value
x_final_out	No value



Firefox File Edit View History Bookmarks Tools Window Help
Proteomic Data Common X Apps · CandWs_demo X SB MELTCancerNormal · Edit X SB L1EM-WF-remap-strande X Tasks · CandWs_demo X FenyoLab xtandem pipeli X MeltCancerNormal demo X Explore Data | Data Brow X Fwd: Thought: A Journal X

https://cgc.sbgenomics.com/u/whm240/candws-demo/tasks/460ee42d-173f-4cf2-82d2-77a4f71e8912/#set-input-data 120% Search

Most Visited Getting Started Reload via NYUHSL

Projects Data Public Apps Public projects Developer whm240

Dashboard Files Apps **Tasks** **CONTROLLED** CandWs_demo Interactive Analysis Settings Notes

BATCH 480 FenyoLab xtandem pipeline - demo task [Get support](#) [Discard](#) [Run](#)

Last update by whm240 on Sept. 13, 2019 12:02
App: FenyoLab xtandem pipeline - Revision: 0

Task Inputs Execution Settings

Inputs

Batching [?](#) On

inputspectra * [?](#) [Change selection](#)

Batch by: File

This will create one task for each selected item.

- ▶ 01CPTAC_CompRef_UCEC_W_PNNL_20170922_B1S5_f01.raw (1 item) ✕
- ▶ 01CPTAC_CompRef_UCEC_W_PNNL_20170922_B1S5_f02.raw (1 item) ✕
- ▶ 01CPTAC_CompRef_UCEC_W_PNNL_20170922_B1S5_f03.raw (1 item) ✕
- ▶ 01CPTAC_CompRef_UCEC_W_PNNL_20170922_B1S5_f04.raw (1 item) ✕
- ▶ 01CPTAC_CompRef_UCEC_W_PNNL_20170922_B1S5_f05.raw (1 item) ✕
- ▶ 01CPTAC_CompRef_UCEC_W_PNNL_20170922_B1S5_f06.raw

App Settings

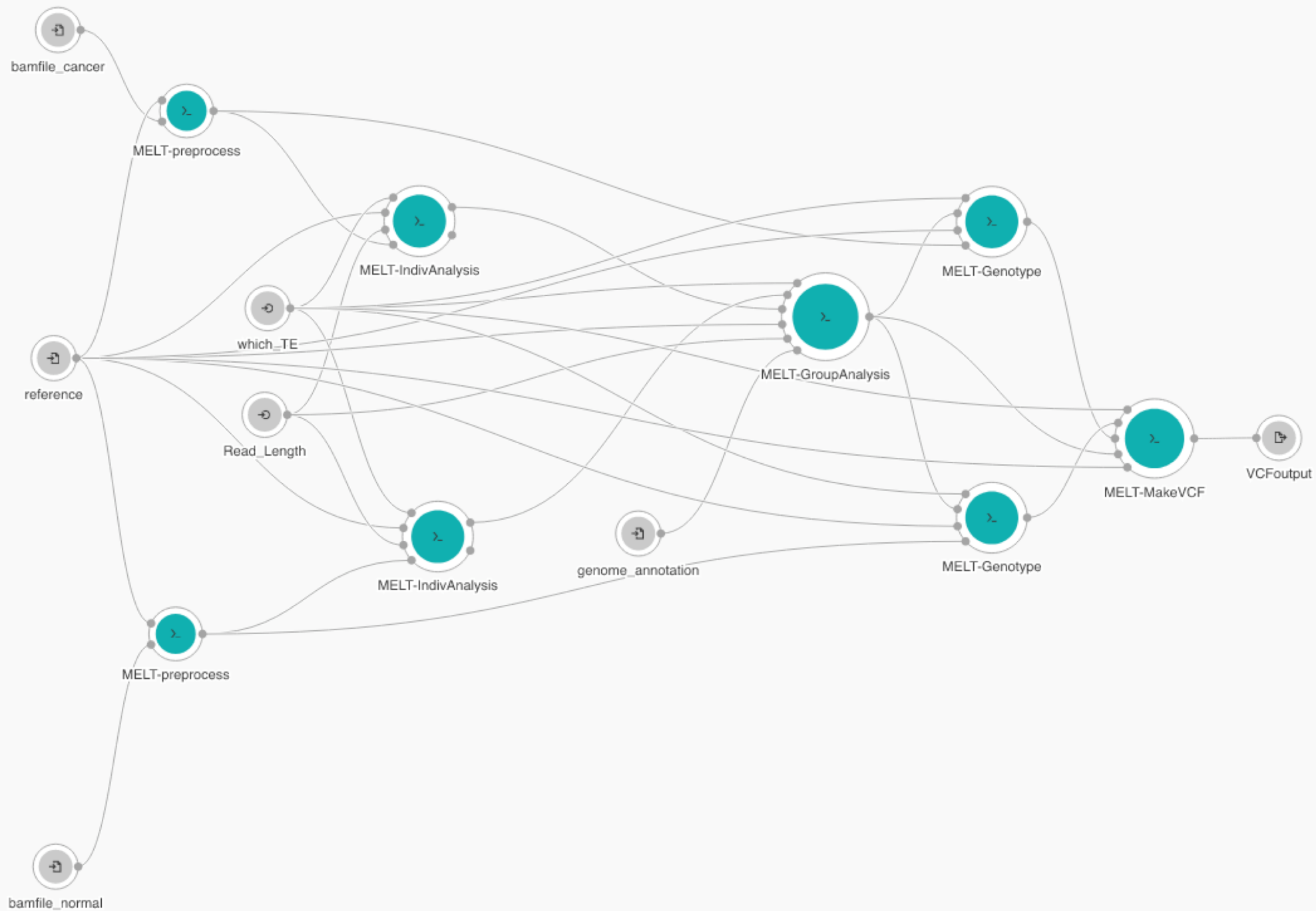
[Edit parameters](#) [Show editable](#)

- ▼ **XITandem** (#x_tandem_2013_06_15_1)
 - modification_mass**
57.021464@C,57.021464@U,229.162
 - prot_taxon** *
human_L1_HERVK
- ▼ **extract_reporter_intensities** (#extract_reporter_intensities)
 - type** *
TMT10

Outputs

relabeled_table [?](#) No value

[?](#)



Firefox File Edit View History Bookmarks Tools Window Help
Proteomic Data Commons X Apps · CandWs_demo X SB MELTCancerNormal · Editor X SB L1EM-WF-remap-stranded · E X Tasks · CandWs_demo X MeltCancerNormal demo task X Explore Data | Data Browser X Fwd: Thought: A Journal of Ph X +

https://cgc.sbgenomics.com/u/whm240/candws-demo/tasks/ffd11c1f-01d2-498a-88f6-decac605e82a/#set-input-data 120% Search

Most Visited Getting Started Reload via NYUHSL

Projects Data Public Apps Public projects Developer whm240

Dashboard Files Apps **Tasks** **CONTROLLED** CandWs_demo Interactive Analysis Settings Notes

BATCH 100 MeltCancerNormal demo task [Get support](#) [Discard](#) [Run](#)

Last update by whm240 on Sept. 13, 2019 12:03
App: MeltCancerNormal_allinone - Revision: 0

Task Inputs Execution Settings

Inputs

Batching [?](#) On

annotation * [?](#) [Change selection](#)

Batch by:

[hg38.genes.sort.bed](#)

bamfiles * [?](#) [Change selection](#)

Batch by: [?](#)

This task will be batched by file metadata (Case ID) and this will create 100 groups.

- ▶ C3I-00006 (2 items) ✕
- ▶ C3I-00008 (2 items) ✕
- ▶ C3I-00032 (2 items) ✕
- ▶ C3I-00090 (2 items) ✕
- ▶ C3I-00098 (2 items) ✕
- ▶ C3I-00136 (2 items) ✕
- ▶ C3I-00137 (2 items) ✕

App Settings

[Edit parameters](#) [Show editable](#)

coverage *

[?](#)

read_length *

[?](#)

which_TE *

[?](#)

Outputs

trans_vcl *No value*

[?](#)

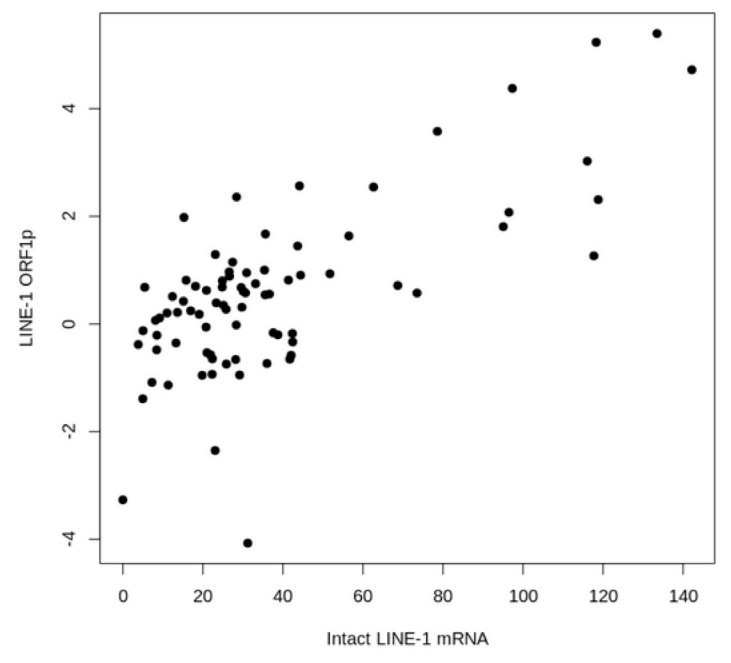
SAVED Demo2
Session started on Sept. 13, 2019 08:35 by whm240
Environment: JupyterLab | Price: \$0.72 | Duration: 1 hour, 43 minutes View Sessions

Copy Start

Files Settings

Analysis files:
Untitled.ipynb
Produced by this analysis
No files

```
In [1]: UCEC_table = read.table('/sbgenomics/project-files/UCEC_LINE.meta.mRNA.prot.somatic.UCECmeta.tsv',header=T,sep='\t')
In [2]: is_tum = UCEC_table["sample_type"] == 'Primary Tumor'
is_used = UCEC_table["Case_excluded"] == 'No'
orflp_measured = !is.na(UCEC_table["ORFlp"])
purity_measured = !is.na(UCEC_table["Purity_Cancer"])
pass_all = which(is_tumis_usedsorflp_measuredspurity_measured)
purity = UCEC_table[pass_all,"Purity_Cancer"]
plot(UCEC_table[pass_all,"both_FPM"]/purity,UCEC_table[pass_all,"ORFlp"]/purity,xlab='Intact LINE-1 mRNA',ylab='LINE-1 ORFlp',pch=19)
```



SAVED Demo2
Session started on Sept. 13, 2019 08:35 by whm240
Environment: JupyterLab | Price: \$0.72 | Duration: 1 hour, 43 minutes [View Sessions](#)

[Copy](#) [Start](#)

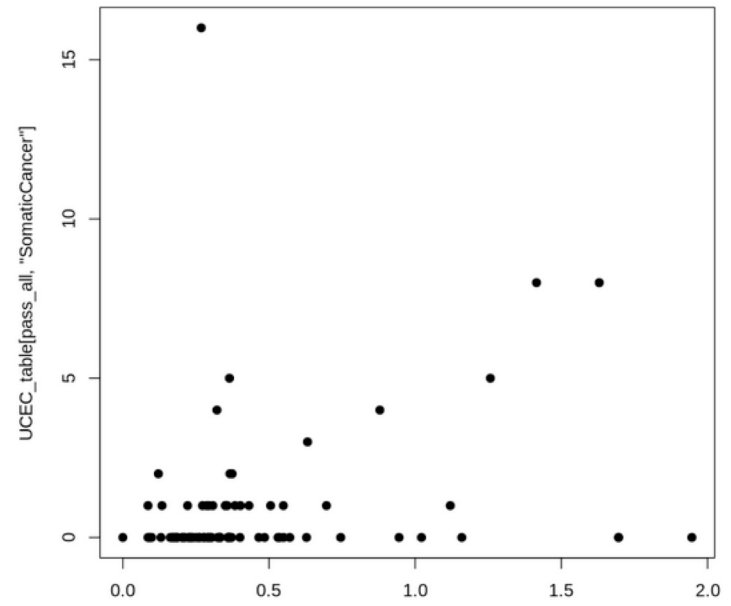
Analysis files:
[Untitled.ipynb](#)

Produced by this analysis
No files

```
In [3]: cor.test(UCEC_table[pass_all,'both_FPM']/purity,UCEC_table[pass_all,'ORF1p']/purity)
Pearson's product-moment correlation

data: UCEC_table[pass_all, "both_FPM"]/purity and UCEC_table[pass_all, "ORF1p"]/purity
t = 9.1904, df = 78, p-value = 4.612e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5956901 0.8120971
sample estimates:
cor
0.7210356

In [4]: LINE1_exp = sqrt((UCEC_table[pass_all,"both_FPM"]/max(UCEC_table[pass_all,"both_FPM"]))
          *(exp(UCEC_table[pass_all,"ORF1p"])/max(exp(UCEC_table[pass_all,"ORF1p"]),na.rm=T)))
plot((LINE1_exp/purity),UCEC_table[pass_all,"SomaticCancer"],pch=19)
```



SAVED Demo2
Session started on Sept. 13, 2019 08:35 by whm240
Environment: JupyterLab | Price: \$0.72 | Duration: 1 hour, 43 minutes [View Sessions](#)

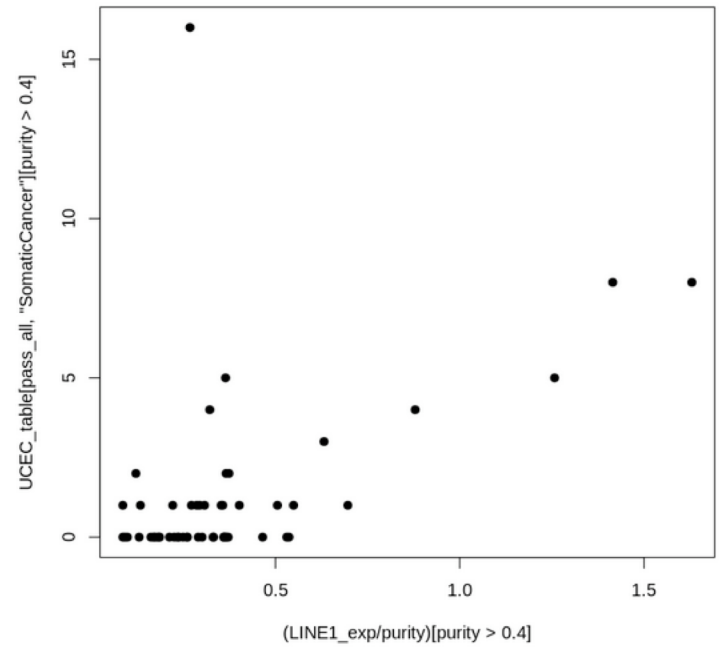
Copy Start

Files Settings

Analysis files:
[Untitled.ipynb](#)

Produced by this analysis
[No files](#)

```
In [5]: plot((LINE1_exp/purity)[purity>0.4],UCEC_table[pass_all,"SomaticCancer"][purity>0.4],pch=19)
```



```
In [6]: cor.test((LINE1_exp/purity)[purity>0.4],UCEC_table[pass_all,"SomaticCancer"][purity>0.4])  
  
Pearson's product-moment correlation  
  
data: (LINE1_exp/purity)[purity > 0.4] and UCEC_table[pass_all, "SomaticCancer"][purity > 0.4]  
t = 4.2639, df = 52, p-value = 8.49e-05  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
0.2792782 0.6935774  
sample estimates:  
cor
```

SAVED Demo2
Session started on Sept. 13, 2019 08:35 by whm240
Environment: JupyterLab | Price: \$0.72 | Duration: 1 hour, 43 minutes [View Sessions](#)

[Copy](#) [Start](#)

Analysis files:
[Untitled.ipynb](#)

Produced by this analysis
[No files](#)

```
In [7]: UCBC_table_filtered = UCBC_table[pass_all, ]

L1_pten_p53 = LINE1_exp[UCBC_table_filtered["TP53_TP53"]==1 & UCBC_table_filtered["PI3K_PTEN"]==1]
L1_pten_only = LINE1_exp[UCBC_table_filtered["TP53_TP53"]==0 & UCBC_table_filtered["PI3K_PTEN"]==1]
L1_p53_only = LINE1_exp[UCBC_table_filtered["TP53_TP53"]==1 & UCBC_table_filtered["PI3K_PTEN"]==0]
L1_no_pten_no_p53 = LINE1_exp[UCBC_table_filtered["TP53_TP53"]==0 & UCBC_table_filtered["PI3K_PTEN"]==0]

x = data.frame(cbind(c(L1_pten_p53,L1_pten_only,L1_p53_only,L1_no_pten_no_p53),
  c(rep('pten, p53',length(L1_pten_p53)),
    rep('pten only',length(L1_pten_only)),
    rep('p53 only',length(L1_p53_only)),
    rep('neither',length(L1_no_pten_no_p53)))))
x[,1] = as.numeric(as.character(x[,1]))
plot(x[,2],x[,1],xlab='Mutation Status',ylab='LINE-1 expression',main='CPTAC3 endometrial cancer')
```

CPTAC3 endometrial cancer

