Easy Data Dissemination

Sam Payne and Caleb Lindgren
Brigham Young University
sam_payne@byu.edu

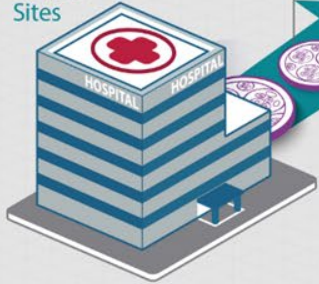https://payne.byu.edu   github.com/PayneLab   @byu_sam

CPTAC
PIPELINE

Tumor Characterization Program (Proteogenomic Tumor Atlas)

Translational Research Program (in partnership with NCI-sponsored clinical trials)

Biospecimen Core Resource (Pathology / Molecular QC)

Tissue Source Sites

Tissue samples & clinical data

Primary and uniform genomic and proteomic data

Proteogenomic Data Analysis Centers

Clinical, pathology & metadata

Analyzed results

CPTAC Data Coordinating Center

Proteogenomic Translational Research Centers (Understanding drug responses and resistance)

Primary genomic and proteomic data

Protein, DNA, RNA analytes

Proteome Characterization Centers (DNA and RNA characterized in partnership with TCGA)

Primary genomic and proteomic data

Harmonized protected & public data

Primary and uniform genomic and proteomic data; Analyzed results

Research Community

Endometrial Cancer

Gather clinical cohort

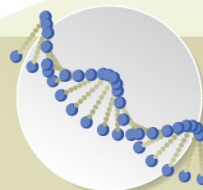Multi-omics of tumor samples

CPTAC Endometrial Carcinoma Cohort

87 endometrioid tumors
13 serous tumors
49 normal uterine samples
  • 18 normal endometrium
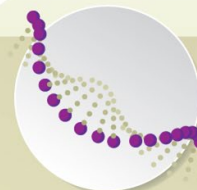 • 31 mixed endometrium-myometrium

| Whole genome and exome sequencing	RNA sequencing	MS protein analysis
Somatic mutation	Gene expression	Protein
Copy number variation	Gene fusion	Protein phosphorylation
Structure variation	Splice variant	Protein acetylation
MSI status	miRNA expression	

Clinical Data

Sample_ID	Patient_ID	Proteomics_Tumor_Normal		Country	Histologic_Grade_FIGO	Myometrial_invasion_Specify	Histologic_type	Treatment_naive	Tumor_purity	Path_Stage_Primary_Tumor-pT	Path_Stag
S001	C3L-00006	Tumor	United States		FIGO grade 1	under 50 %	Endometrioid	YES	Normal	pT1a (FIGO IA)	
S002	C3L-00008	Tumor	United States		FIGO grade 1	under 50 %	Endometrioid	YES	Normal	pT1a (FIGO IA)	
S003	C3L-00032	Tumor	United States		FIGO grade 2	under 50 %	Endometrioid	YES	Normal	pT1a (FIGO IA)	
S005	C3L-00090	Tumor	United States		FIGO grade 2	under 50 %	Endometrioid	YES	Normal	pT1a (FIGO IA)	
S006	C3L-00098	Tumor	United States		NaN	under 50 %	Serous	YES	Normal	pT1a (FIGO IA)	
S007	C3L-00136	Tumor	United States		FIGO grade 1	under 50 %	Endometrioid	YES	Normal	pT1a (FIGO IA)	
S008	C3L-00137	Tumor	Other_specify		FIGO grade 2	under 50 %	Endometrioid	YES	Normal	pT1a (FIGO IA)	
S009	C3L-00139	Tumor	United States		NaN	50 % or more	Serous	YES	Normal	pT3a (FIGO IIIA)	
S010	C3L-00143	Tumor	United States		FIGO grade 1	Not identified	Endometrioid	YES	Normal	pT1 (FIGO I)	
S011	C3L-00145	Tumor	United States		FIGO grade 1	under 50 %	Endometrioid	YES	Normal	pT1a (FIGO IA)	

Genomics Data

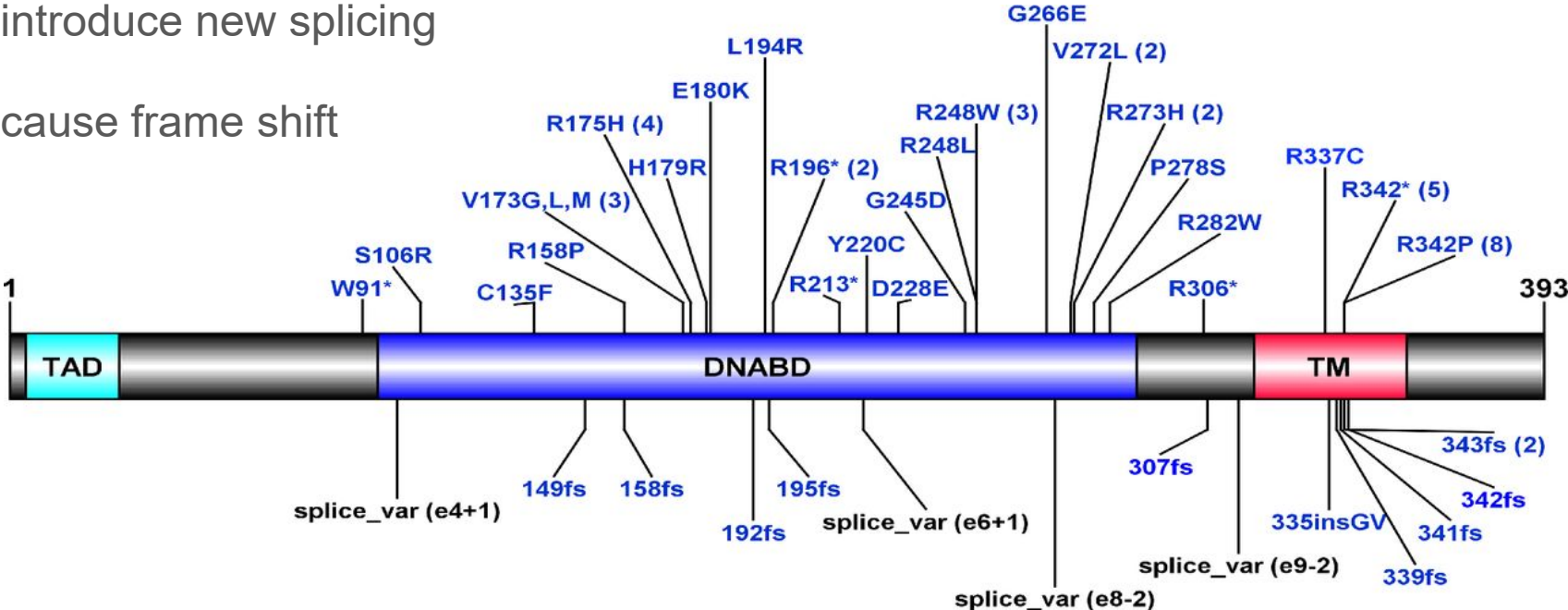Somatic mutation - changes to DNA after embryo, present only in some cells
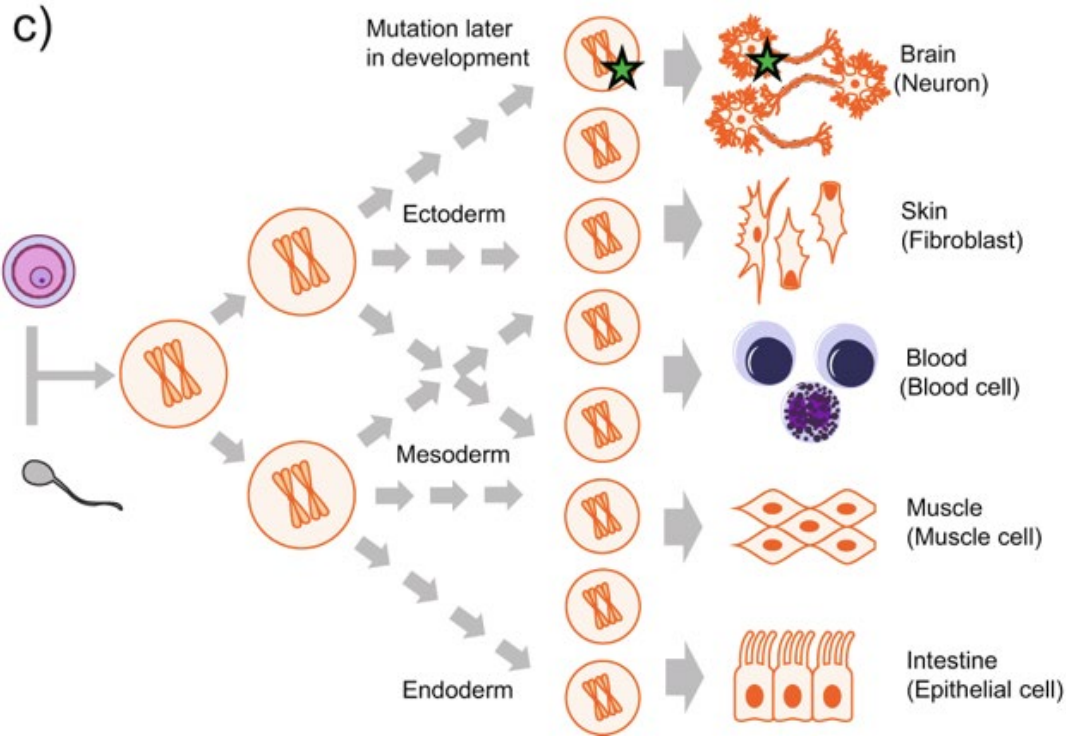
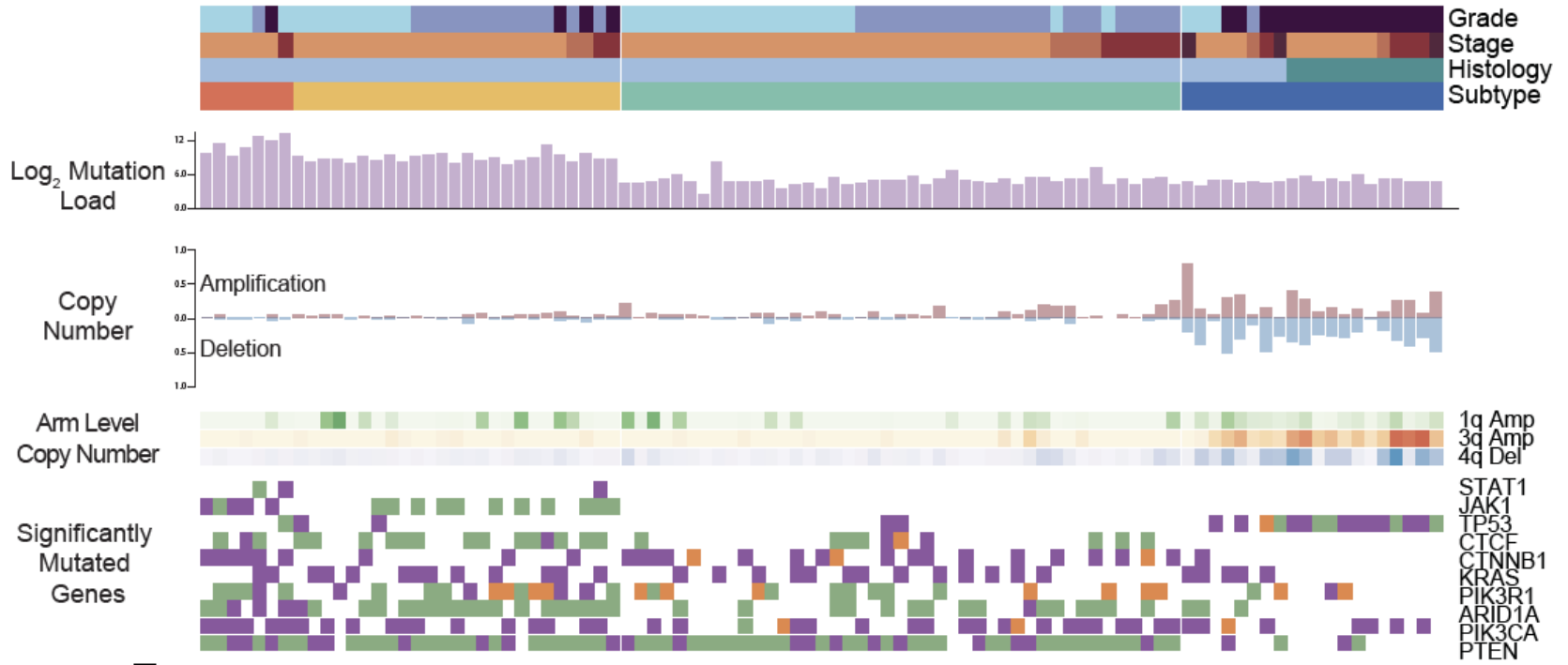# Single Nucleotide Mutations

change an amino acid

introduce new splicing

cause frame shift

Exploring Data

Look at clinical and genomics data to find mutations that correlate with histology

Mutation and clinical data

Sample_ID	Histologic_type	TP53_Mutation	TP53_Location	TP53_Mutation_Status	Sample_Status
S001	Endometrioid	Missense_Mutation	p.R248W	Single_mutation	Tumor
S002	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor
S003	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor
S005	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor
S006	Serous	Missense_Mutation	p.S241C	Single_mutation	Tumor
S007	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor
S008	Endometrioid	Missense_Mutation	p.S127Y	Single_mutation	Tumor
S009	Serous	Nonsense_Mutation	p.R196*	Single_mutation	Tumor
S010	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor
S011	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor
S012	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor
S014	Endometrioid	Wildtype_Tumor	No_mutation	Wildtype_Tumor	Tumor

Multi-omics data

DNA, RNA, Proteins and cellular state
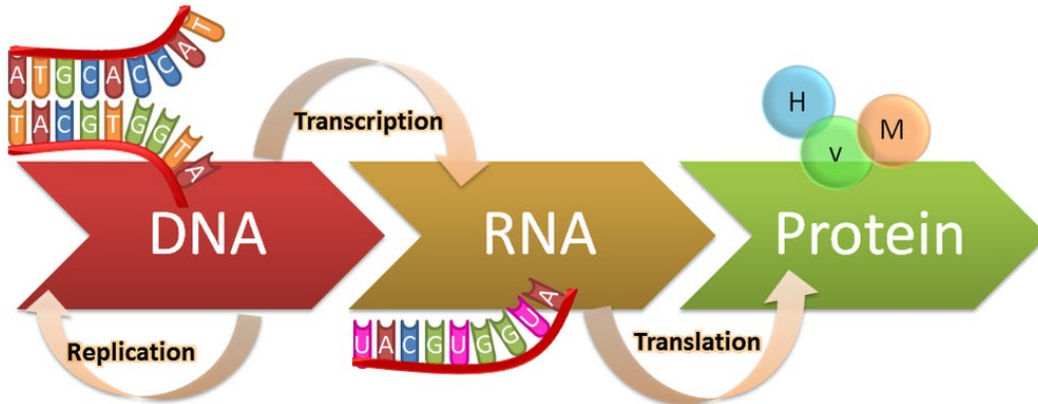
DNA, RNA measured with 'sequencing'

Protein measured with mass spectrometry

	CTNNB1	JAK1	PIK3CA	PTEN	TP53
Sample_ID					
S001	0.80300	-0.38200	-0.4500	-0.5260	0.295000
S002	0.42600	-0.39300	-0.2150	-0.8300	0.277000
S003	0.00891	-0.09580	-0.1780	-0.9410	-0.871000
S005	0.78800	-0.62700	-0.3920	0.7300	-0.343000
S006	-0.40200	0.14100	0.2870	-0.3790	3.010000
S007	1.43000	-0.22900	-0.3490	0.0293	-0.148000
S008	0.97100	-0.54800	-0.3170	-1.0100	0.441000
S009	-0.25200	-0.36000	0.2820	0.1300	-1.220000
S010	0.00979	0.30400	-0.0714	0.3900	-0.082500
S011	0.59100	0.44100	0.0640	0.0301	0.181000

Multi-omics data integration

B. Protein abundance, mutation status, cancer subtype

E. Protein abundance, phosphorylation abundance, histology
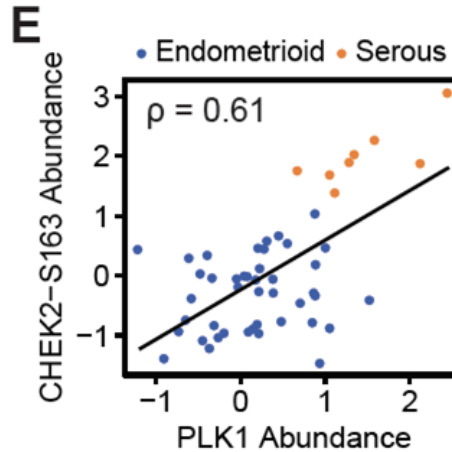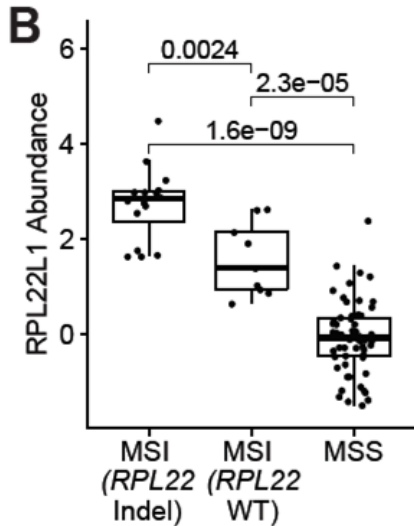
cptac package

Python package for Data as a Service

data is 'in code, not an excel file'

```
Below are the dataframes contained in this dataset:
        acetylproteomics
                Dimensions: (144, 10862)
        circular_RNA
                Dimensions: (109, 4945)
        clinical
                Dimensions: (144, 26)
        CNV
                Dimensions: (95, 28057)
        derived_molecular
                Dimensions: (144, 125)
        experimental_setup
                Dimensions: (144, 26)
        miRNA
                Dimensions: (99, 2337)
        phosphoproteomics
                Dimensions: (144, 73212)
        phosphoproteomics_gene
                Dimensions: (144, 8466)
        proteomics
                Dimensions: (144, 10999)
        somatic_mutation
                Dimensions: (52560, 3)
        somatic_mutation_binary
                Dimensions: (95, 51559)
        transcriptomics
                Dimensions: (109, 28057)
```

Interactive demo of a notebook

View demo at

https://colab.research.google.com/drive/17SMHMDeOUMdkJAd2Xc3Hgvvyqv_WEW4r

How to work with imaging data

imaging features as a table (nucleus_major_axis, nucleus_aspect_ratio, etc.)

correlating imaging features with clinical, omics, etc.

Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence