Debriefing: sc targeted proteomics

Aedȋn Culhane

BIRS

sc targeted Proteomics; Application and Future

Increasing resolution, 3D, but limited # Proteins,

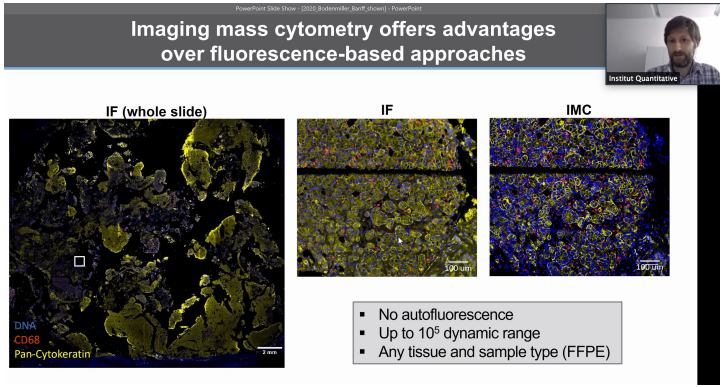
Good on FFPE so many samples available

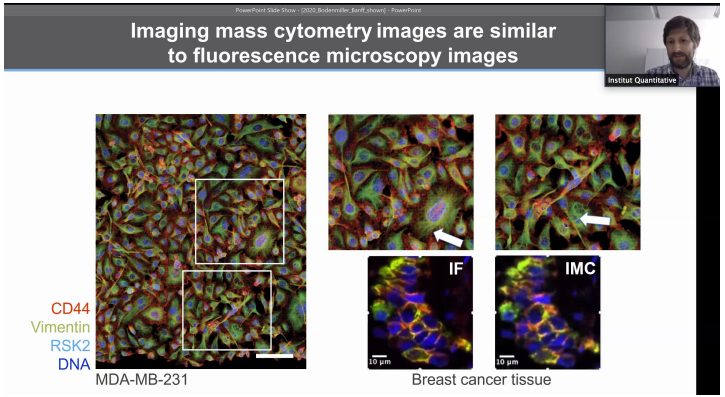
Spatial/co- expression of known molecular targets, small molecule inhibitor

Longitudinal Clinical/Biological

Widespread use depend on clinical utility and cost

- Cell properties- shape/size, type
- Cell type proportions /composition
- Cell location, co-location analysis
- Cell higher architecture (clusters of cells)
- Cell-cell interaction (eg tumor-immune)





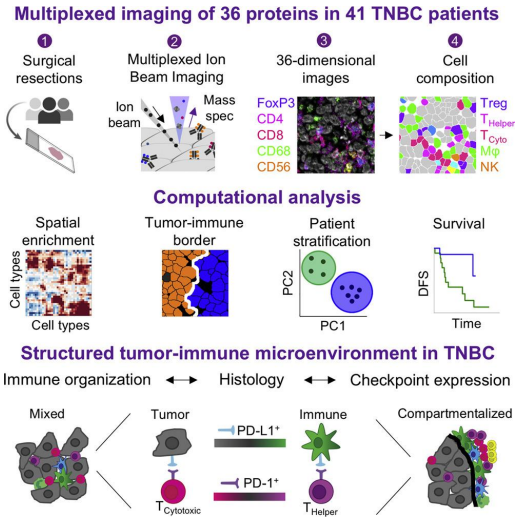
Data challenge

Single-cell targeted proteomics across technologies

Data from 2 studies characterizing the **breast cancer tumor immune microenvironment**, using:

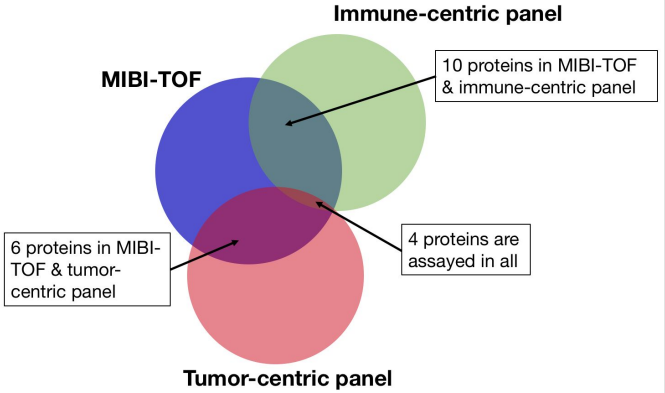
Multiplexed ion beam imaging by time-of-flight
(MIBI-TOF)

- 36 proteins
- 41 TN breast cancer patients



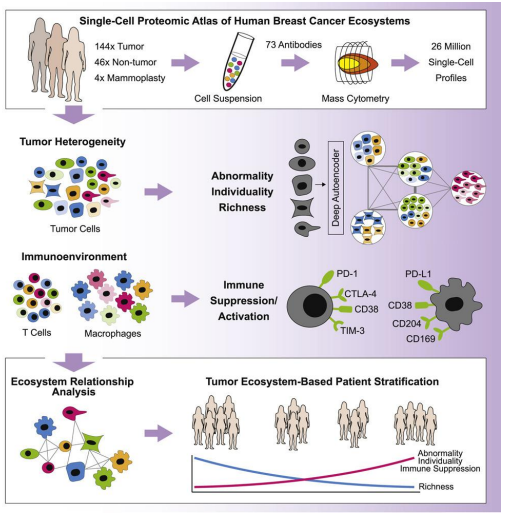
Keren, et al. (2018)

… with a total of 20 overlapping proteins



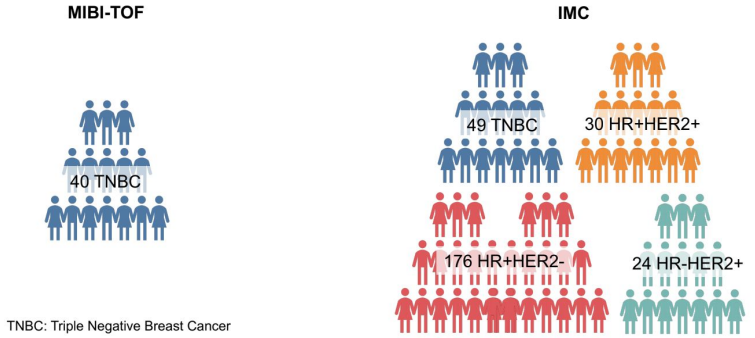
Mass cytometry
(aka mass tag, CyTOF)

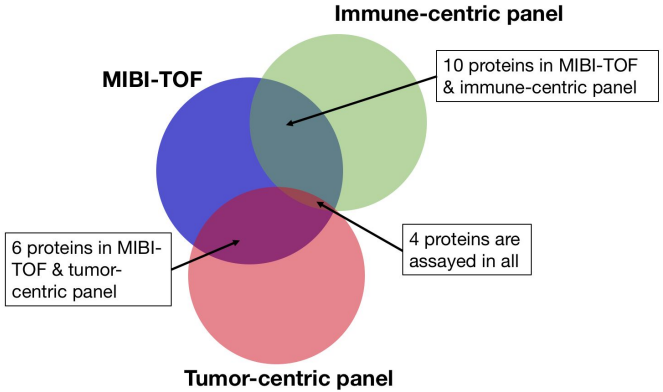
- 73 proteins across 2 panels: tumor and immune
- 140 breast cancer patients, 6 TN



Wagner, et al. (2019)

Data Challenge was challenging





- Limited overlap in phenotype: only 6 TN patients in the CyTOF dataset
- Limited overlap in proteins → some cells cannot be detected in one platform
- Few features. Can’t integrate features at Gene Ontology or Pathway level
- No obvious “ground truth” for validation

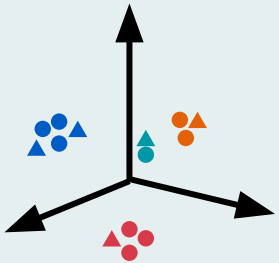
Hackathon Challenge Questions

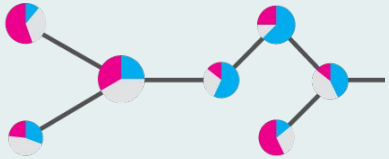
- How should we approach **integrating partially-overlapping** proteomic data collected on different patients with similar phenotypes?
- Can we **integrate** other 'omics datasets to support the results of these proteomic analyses?
- Without including the **spatial** x-y coordinate data, how well can we predict cell **co-location**?
- Can we **predict the spatial expression patterns** of proteins measured on mass-tag but not measured in the MIBI-TOF data?
- What additional information can we learn about the different **macrophage and immune populations** in breast cancer by conducting integrated analyses of these datasets?

Challenges

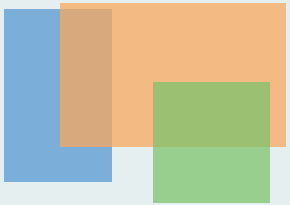
Integration of features

Cross-platform



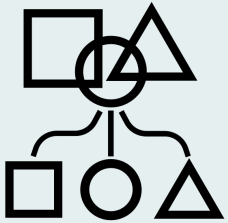


Transfer Imputing features across datasets



Prediction

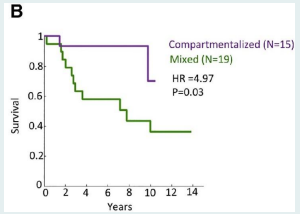
Transferring cell labels between datasets



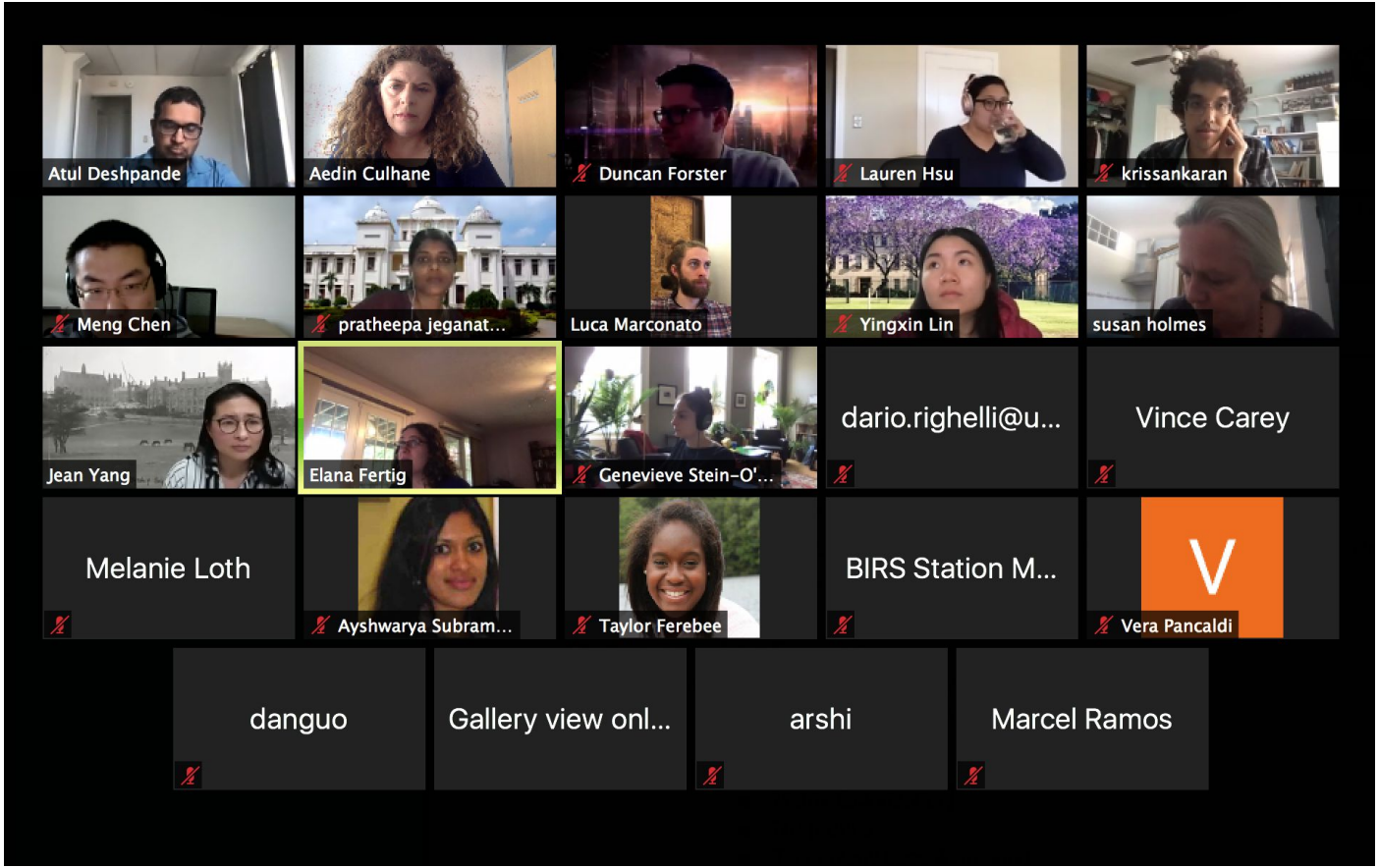
Spatial co-location, structure

Spatial structure co-location, autocorrelation

clinical significance



sc targeted Proteomics Brainstorm



Challenges in Data Integration

- Connecting partially overlapping datasets (feature)
- Gene Ontology (Duncan- cell type specific map)
- Combining different scales
- Normalization

- Scale - connecting partially overlapping datasets (cell)
- Coarse and fine grained resolution
    - (single cell-> bulk-> spatial->communities)
- Hierarchical , layers
- Integration at clinical level

Data Challenges: Normalization

Transform - log, arcsinh, sqrt?

Order of magnitude in scale

Questions about the validity of normalizing by Cell Size (neurons v leukocytes)

Cell type markers maybe only 1 cell type

Quantile -can’t assume composition or sum of expression uniform

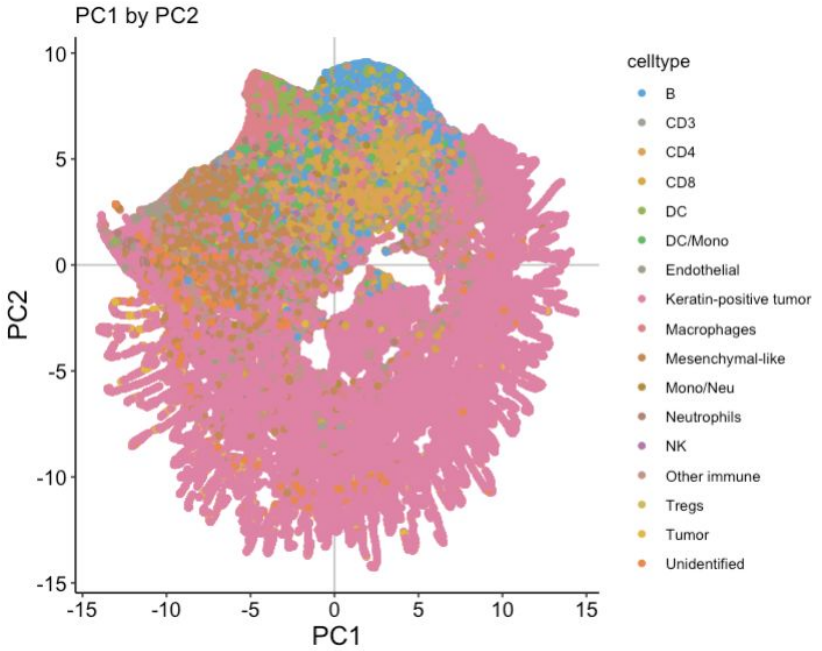
Different physical technologies, strong batch effects

Funky MIBI not mass Tag Umap - Why?

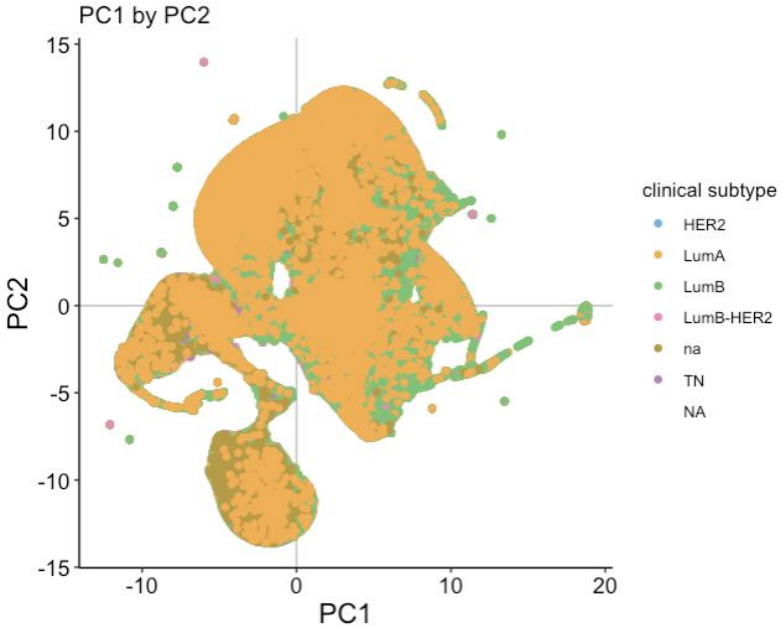
Outlier effect of cell type marker in low feature space, where markers unique to cell?

The UGLY MIBI UMAP!!!

mibi

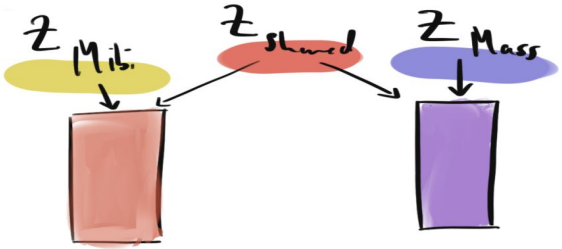


cyTOF



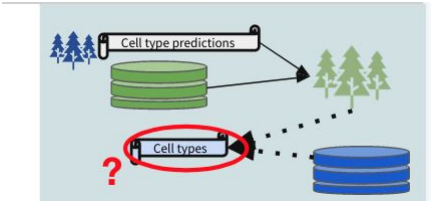
Connecting Partially overlapping Data

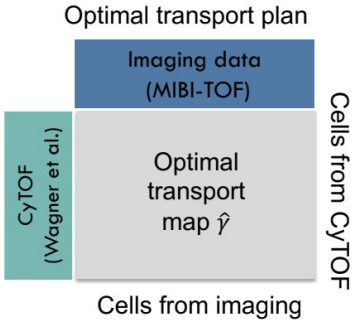
Cells/Clusters/Clinical - Two broad approaches Embedding or Graph - Taxonomy

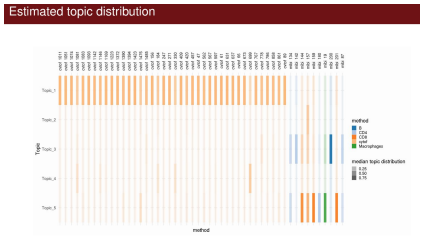


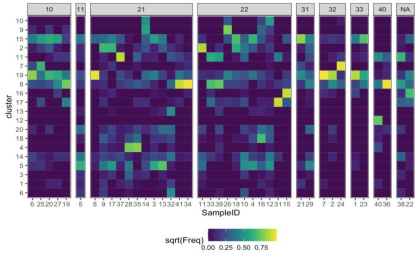
1. **Cluster**: Make clusters, from expression data
2. **Featurize**: Define spatial features
3. **Aggregate**: Data are at cell level, but we need summaries at sample level. So compute functions of spatial features / find cluster mixing %s.

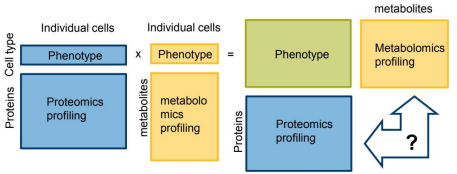
cluster











Bridge to connect data with partial/little overlap

Proteins as cell type or state marker

“BRIDGE”

- Presence/Absence of Cell States
- Cell compositions/proportions
- Spatial Structure (eg Morans)
- Biological Sample Phenotype
- Clinical features
- Survival (interval/right censored discussion)

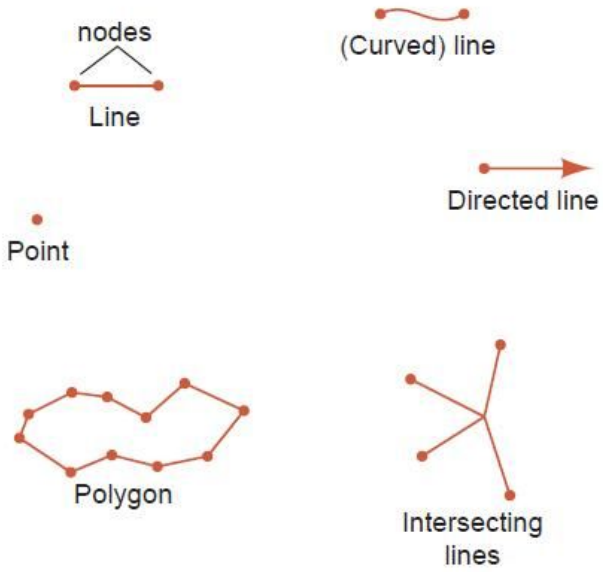
Integration of Spatial data - Scale, SWM

Scale and space in biological systems

How to define spatially- points, lines, polygons

XY centroid may poorly capture

- Cell size or shape (neuron)
- Neighbours may depend on cell size/shape
- Community
- Partially overlapping



I.16 Spatial objects

Spatial objects in a GIS can include points, lines of various types, intersecting lines, and polygons.

Integrating Spatial Data - Biological data have always been spatial

Spatial statistics already well established in other fields

- GIS, Ecology, Plant science (Taylor Ferebee), Weather (Elana Fertig)
- Statistics based on Euclidean distance (Sphere), graph triangulation (Moran, Gabriel), K-NN etc
- 15 years ago applied Ripley's K factor on biopsy (Susan Holmes)
- Test statistics to associate spatial patterns with covariates

Connecting features to spatial across platforms

Need to unify scaling (Vera Pancaldi)

Limited expected map of

- The physical structure
- Cell composition in structure
- Expected level of protein expression

How to connect Coarse v Fine resolution neighborhoods ?

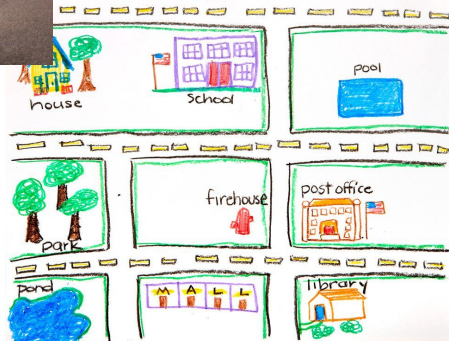
Connecting features to spatial across platforms

Need to unify scaling (Vera Pancaldi)

Limited expected map of

- The physical structure
- Cell composition in structure
- Expected level of protein expression

NEED



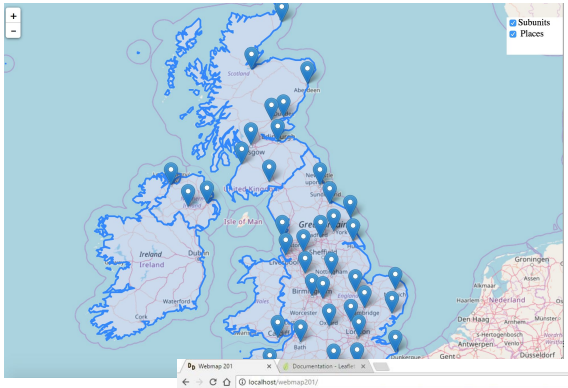
How to connect Coarse v Fine resolution neighborhoods ?

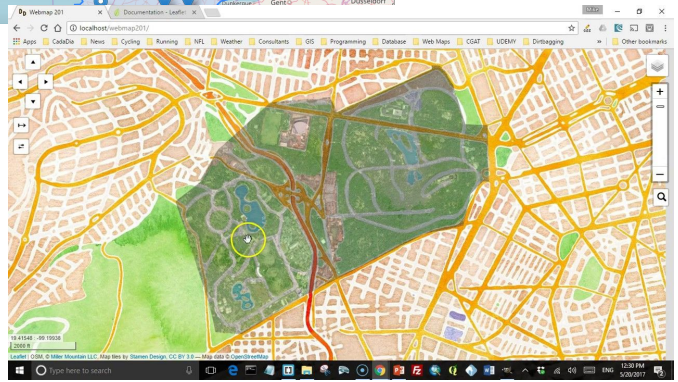
Have - Digital ‘Omics + Analog Map

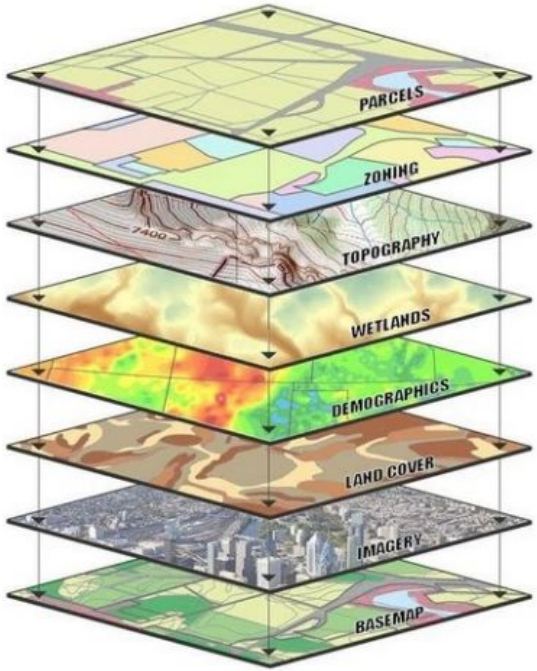
Library, Trinity College Dublin, Ireland

Will be “driving” blindly until we have a digital cell location map

Have layers in GIS data (Susan Holmes)







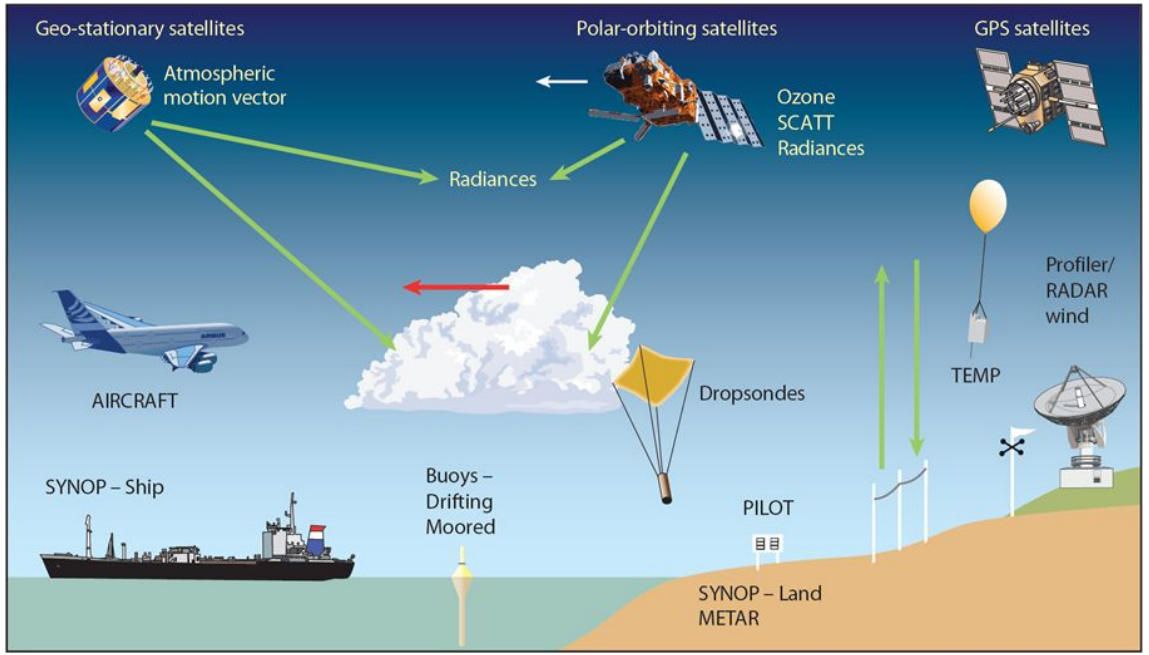
GIS DATA LAYERS

Many different types of data can be integrated into a GIS and represented as a map layer.

Examples can include: streets, parcels, zoning, flood zones, client locations, competition, shopping centers, office parks, demographics, etc.

When these layers are drawn on top of one another, undetected spatial trends and relationships often emerge. This allows us to gain insight about relevant characteristics of a location.

Weather layers: Satellites, Weather Balloons (Elana)



Single Cells ->Bulk -> Communities

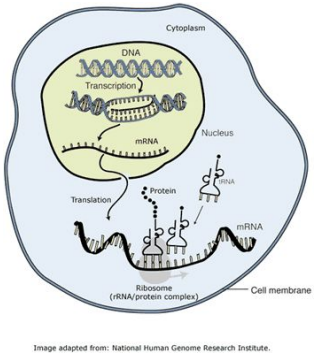
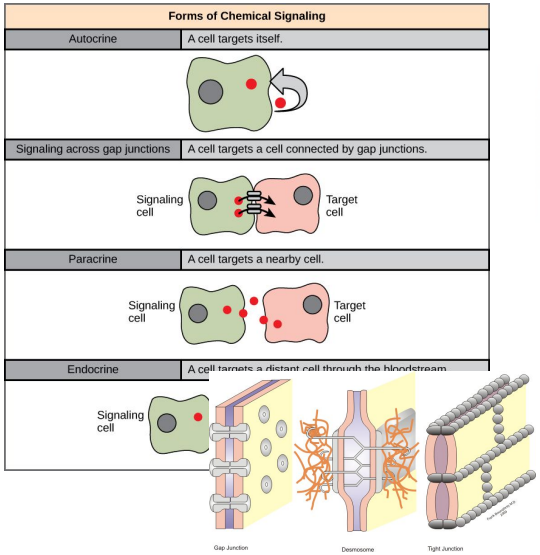


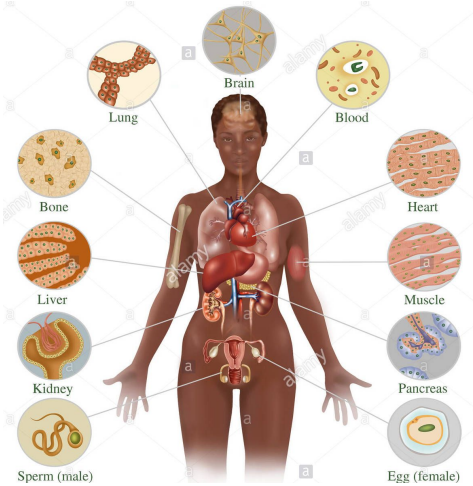
Image adapted from: National Human Genome Research Institute.

‘Omics DNA Chromatin
RNA Protein
Glycosylation-
metabolites etc



Connected by signnaling
(paracrine, endocine
Gap junctions, autocrine)

Composed of organized
Cells types, polarity



Human Phenotype
defined by
Systems,
Organs that are
composed of Cell
Communities

Spatial Proteomics Future Directions

- Need a digital map and to define variables of interest
- Can this be built from existing data
- New benchmark datasets
- Methods - leverage existing, build new
- spatial scale measures
- Other methods for validation (prediction strength, can we large number of “observations” cells)
- What are the limitations of the physical technologies and what directions are the technologies headed?
    - Limited channels

Output

- Table of Data types
- Table/Figure of Spatial measures
- Table of Taxonomy of methods
- (there was support for packaging up data/code vignettes in an R package)

Figures :

- Bridging Studies: mapping between datasets: pheno, cells, features
- Methods: Embed/Graph - Optimal Trans/Matrix Factor/Topic Modelling/Forests
- Communities Clusters/UMAP → Spatial Structure: XY- Line-Polygon. SWM (distance/grid/graph tri, neighbors, etc)
- Annotation (Duncan)

Table? Descriptors of different data types (sc'OMICS)

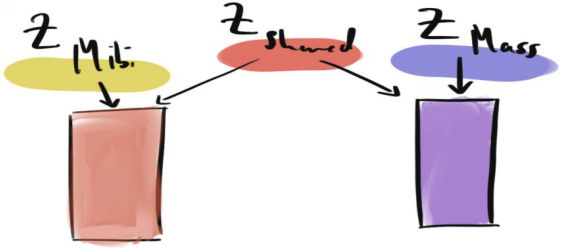
N Typical number of features (proteins/genes) measured
P Typical number of samples/cells measured
Distributions/orders of magnitude
Sample type (single-cell, multi-cell, bulky)
NA Missing values (%, the typical way to impute missing values)
Count data/intensity (continuous)
Omics : genomics/transcriptomics/proteomics/PTM/metabolomics
Typical normalization method (log transform)
State of the art publications
Clinical data -- potentially following clinical trials can help control for diversity of patients; also incorporating EHR
Spatial info (no, 2D, 3D)

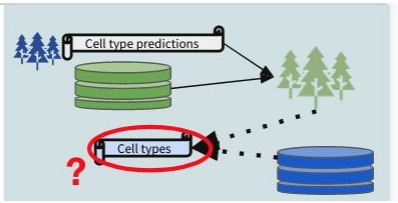
Table - “Taxonomy” of methods

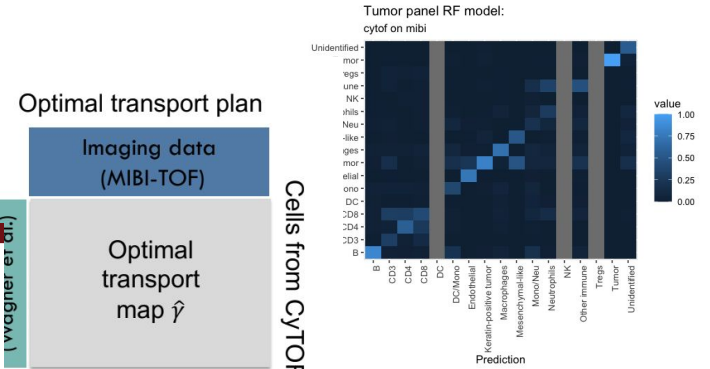
- Represent as a Table, or decision tree figure to manuscript?
- Matrix factorization
- NN graphs
- Technology specific methods
    - Still want consensus
    -
- Methods broadly divided into
        - Embedding
        - Graph
- Hierarchical Bayesian

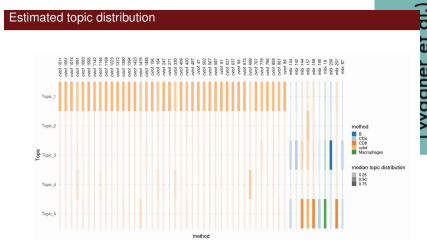
Figures Aligning data/Learning Across Studies

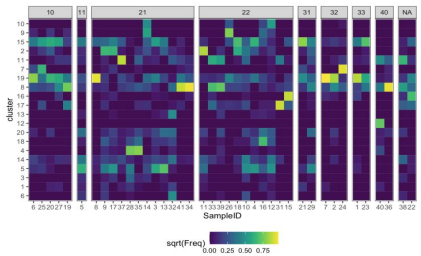
INSERT FIGURES????











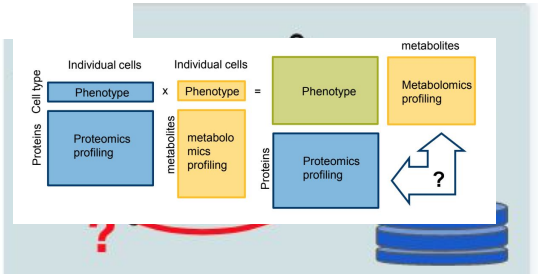
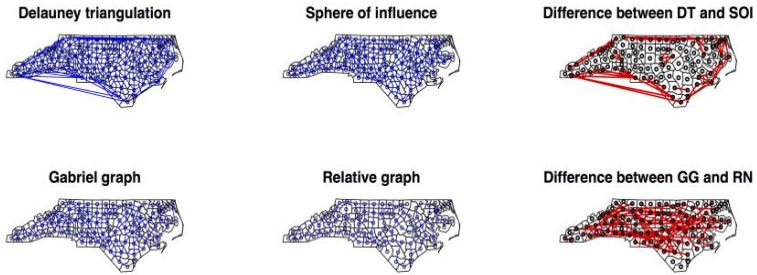


Figure: Spatial → Biology

Measures XY/Lines/Polygons

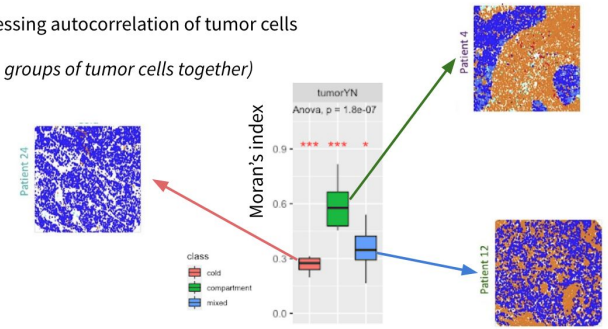
SWM Sphere (distance)/GRID/ Graph



Moran's index: Tumor cells

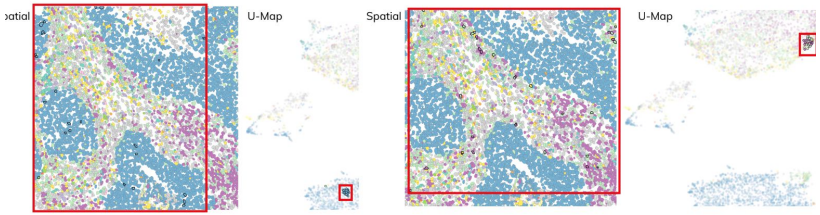
Assessing autocorrelation of tumor cells

(i.e., groups of tumor cells together)



Interactive Visualization

- Linked Brushing: Combine (literal) spatial map with abstract (U-)map
- Within cell-types, some U-Map clusters are spatially co-located, but far from universal



Examples where U-Map clusters are spatially diffuse. Immune cells highlighted in left pair, tumor on right.