Lecture Notes - BIOC 412

Class: #BIOC412

Wednesday, Sep 13, 2023

Topic: **Data Sets**

Todo/Assignments

- ✅ ~~Look at Assignment Part 2 ✅ 2023-09-19~~
- ☐ Ask Dr. Murch about her Turkish Coffee

Notes

Datasets

Wine

Top Middle and Bottom Replicates → Pseudoreplicate
Hypothesis:
What can we use this data to ask?
Compare years 2005 vs. 2006?
Compare Wineries
Top vs. Middle. vs. Bottom OF EACH DIFFERENT WINERY
Alcohol level vs. Chemical Diversity

Coffee

Big dataset!
Different brands.
Is prep (extraction temps) contributing to diversity.

Cannabis

White Widow and Big Bud with different plant replicates.

Assignment

What do I want to compare?
Data quality characteristics!

- High-risk vs. low-risk
  - 3 Replicate → features detected, how many replicates does it need to be present in to be **my data?**
  - All 3 can throw out good data (Type 2 - False Negative);
  - Only in 1 increases risk of artifacts (Type 1 - False Positive).
  - **No Real Wrong Decision**
  Find out the quality of the data - do some testing of the water.
- ☐ What constitutes a feature in the data?

Types of Experimental Design for Metabolomics Data Analysis

Complete Randomized Block

Simplest design

	Group 1	Group 2
Treatment 1	X	X
Treatment 2	X	X
Treatment 3	X	X

Stats to Compare
Compare treatments
Compare groups
Compare treatment x group interactions

Use ANOVA

Nested Design

Two conditions
(Pretend there are two)

Treatment	Group 1	Group 2
1	X	X
2	X	X
3	X	X

Use MANOVA

Time Study

Treatment	Start	Time Interval 1	2	3	n
1	init	Replicates	...	...	...

Treatment	Start	Time Interval 1	2	3	n
2	init	Replicates	...	...	...
3	init	Replicates	...	...	...

Stats to Compare

Compare Treatments

Compare Time Points

Compare Treatment x Time

Regression analysis

Multivariate analysis

Replicating Blocks

Experiment	Dependent	Independent
1	Treatment 1	Replicate Tests
1	Treatment 1	Replicate Tests
1	Treatment 1	Replicate Tests
2	Treatment 1	Replicate Tests
2	Treatment 1	Replicate Tests
2	Treatment 1	Replicate Tests

ANOVAs within each block

- Hope you get the same answer
- If not, try again
- Not a MANOVA

Data Minimization

Pick treatments and use ML to predict the space between treatments.

DOE - Design of Experiments

- Originally developed to optimize manufacture of ball bearings!
- System of selected treatments and comparisons to extrapolate relationships.
- Method of simplifying and reducing sample numbers.
- Determines relationships BETWEEN treatments and responses as a geometric pattern.
- Makes predictions for responses.

1. One factor x
2. Two factor xy

3. Three factor xyz
4. Four factor - orthogonal projection
5. Nine factor cube projection
6. Ten dimensional cube projection

DesignExpert
jmp Statistical Discovery
R...?

☐ DO R IN CHATGPT. JUST TRY IT.

Ellistat is FRUSTRATING AND HARD TO USE

Give ranges, program makes a small set of tests

Do it at the start, do the minimum amount of work and get the maximum amount out of it.

Youden Experiment - Type of DOE

Older, no cool fancy tools in R.

Investigate 7 factors in one experiment requiring only 8 determinations.

1. Choose 7 factors
2. Choose a high and low value for each
3. run experiments so that all are covered
4. stats calc the effect of each factor
5. plot factors along a line according to relative weight
6. identify factors that matter.

$$\frac{X_1+X_2+X_3+X_4}{4} - \frac{X_5+X_6+X_7+X_8}{4} = J$$

Mass Spec

m/z can be used for:
Compound ID
Checking Isotopes

Let's Build a Mass Spec

1. Sample injection
   1. Chromatography (HP-LC)
   2. Heat (GC-MS)
   3. Paper (Airport)
   4. Laser Ablation

     5. MALDI

     6. ...

2. Ionizer
    1. ESI (Electrospray Ionization)
        1. COULOMBIC EXPLOSION
3. Focuser
    1. Lens, Z-Stack, Filter, basically a copper coil
    2. Lower charges are slower
    3. Smaller molecules are faster
4. Quadrupole
    1. Alternating charges metal rods
    2. Spins 'em around
5. Archetype
    1. **Single quadrupole**
    2. **Time of Flight**
        1. Out of Q1 into literally a box
        2. Pushers and Pullers - Back and Forth
        3. V and M modes
        4. Smallest and Most Charged Leave First
        5. m/z directly proportional to time spent in the trap.
    3. **Triple Quadrupole - MS-MS (Tandem)**
        1. Q1 - First Quadrupole
        2. Q2 - Collision Cell
            1. Has Poles
            2. In a box
            3. Put in some Argon gas 'cuz it's inert
            4. High voltage to fragment
        3. Q3 - Put fragments in a nice straight line
        4. Single Reaction Monitoring - Not everything will fragment in Q2
            1. MRM, SRM, SRI
        5. Only measure the fragments you optimize for
    4. **Orbitrap**
        1. Modified version of a ToF MS.
        2. Stuff never has to leave
            1. You can pick when to let stuff leave
        3. You can also fragment everything
        4. Data-Dependent Analysis
            1. Pick one m/z to fragment and detect
        5. Data-Independent Analysis

1. Fragment EVERYTHING
2. MSDIAL puts everything back together from fragments.

Monday, Sep 18, 2023

Topic: Metabolomics Workflow: Data Acquisition Guest Lecture

Lecture Link: N/A

Todo/Assignments

- ☐ Assignment 2 📅 2023-09-27

Notes

TMIC - The Metabolomics Innovation Center

- ☐ Check out TMIC web page

hmdb
Many services for doing stuff below.

NMR Metabolomics

- Non-destructive
- Robust instruments
- Minimal instrument downtime
- Simple sample prep
- No chromatography
- No chemical derivatization
- Spectra are predictable
- Allows for precise structure determination
- Inherently quantitative
- Easily automated
  But...
- Poorly sensitive
- Modest metabolite coverage
- Expensive instruments
- Large instrument footprint
- Needs cryogens (He (l))
- Need to maintain
- Small spectral databases
- Few software resources

GCxGC-MS and GC-MS Metabolomics

GCxGC-MS

- Excellent sensitivity
- Excellent separation
- High peak capacity
- 2D separation plane
- More information per unit time
  But..
- Limited availability of fast detectors
- Maximum allowable operating temp
- vast amount of data

GC-MS

- Sensitive
- Excellent separation
- Comprehensive DBs for identification
  But...
- Requires derivatization to make things volatile
  - E.g. Sugars $\rightarrow$ acetylation
- Fit for non-targeted screening of volatile compounds

LC-MS Metabolomics

Non-volatiles

- Most popular
- Lots of options for detectors, chromatography, derivatization, etc.
- Targeted or untargeted
- Less clear of pros and cons.
  - Can get expensive

Metabolomics Toolbox

Experimental Design

Cassette model with complete randomized standards
Sample replicates with automated data collection and integration

Validated Metabolomics Methods

- Validation Standards
- Alignment Standards
- Targeted Standards

Statistical Models and Scripts

- Eliminate false discoveries
- Discover new metabolites

Logical Algorithms and Biotransformations

- Discover new pathways
- Discover metabolomic responses
- Discover metabolite families
  Murch is interested in this.

pnnl-comp-mass-spec.github.io

Wednesday, Sep 20, 2023

Topic: dd

Lecture Link:

Todo/Assignments

- [ ]

Notes

Guest Lecture from Concordia on Friday. Metabolomics in the medical system. 1pm!!!
Be there!!!

Case Study: Thidiazuron (TDZ)

TDZ is a Herbicide

- Sprayed on cotton fields
- Chemically synthesized
  - Diurea derivative (thiadiazole and phenyl)
- All the leaves fall off the plants
- Plant growth regulator
- Mediator of endogenous plant growth regulators

- Sold as DROPP
  - Made cotton cheap!
  - Sprayed 5 days before harvest
  - Leaves drop in 3 days
    - Unique to Malvaceae
  - Leaves are green and turgid

Many papers published

- *In vivo* propagation
  - Cotton defoliations
  - Bud breaking of apple trees
  - Greenhouse regeneration
- *In vitro* propagation
  - Organogenesis
  - Somatic embryogenesis
    - Basically budding, African violet.
      - Murch part of breeding program.
      - Horsters greenhouse in Ontario.
      - Undifferentiation from somatic cell, rearranges its identity, and redifferentiates into an embryo.

FT-MS: Fourier Transform MS

- Quadropole
- Detector collects all signals off of very long poles
- Math uses FFT to transform signal to *m/z*s
- Developed by Comisarow M. was from UBC Vancouver - Fourier transform ion cyclotron resonance spectroscopy. He invented it..
- UVIC has a 14T
- Custom built instruments

Hypotheses:

Thidiazuron forms oligomers in solutions and plant tissues.

Obtained m/z to 6 decimal places, highest you can get!

- Vast majority of the time you don't have a separation system before collection.
- You can predict a formula from deconvolution. How?
  - 19774 lines of data
    - 3 modes and 3 extractions
      - +, -, neutral modes

- EtOH, Water, Hexanes
- 9 Treatments
  - 2 Replicates
    - 3 treatments
    - 6 samples
    - $15,000
- **Deconvoluting m/z Signals**
  - Prediction algorithms
  - You really need at least 4 decimals to tell things apart.

TDZ Oligomers in stock solutions

- Old solution worked better than a brand new solution.
- Murch Review 1997
  - 220, 440, 660, 880, 1100, 1320 peaks
- Diels-Alder!!!
  - Diene and Dienophile
- Tetramer is very structural, maybe it is docking specifically somewhere and doing something.
  - Monomer is really not that active.

Thidiazuron is metabolized by plant cells to release bioavailable sulfur and nitrogen.

Looking at breakdown products, there are compounds that could give N or S. Glutathione is higher in lower [TMZ] treatment.

TDZ increases uptake and catabolism of 5C and 6C sugars from the culture medium.

Holds the glucose transporter open sterically.
(Slide | Gray dots decreased, pink dots increased).

- Investigate pathway increases
- TMD changes how plants move sugars.
- Shut down chlorophyll metabolism; porphyrin.

Mummichog vs. GSEA

TDZ forms conjugates with molecules in plant cells.

Look at slides.

TDZ inhibits biosynthesis of diterpene-derived metabolites and enhances synthesis of sesquiterpenes and triterpenes.

Growth regulators

- Absiscic acid decreased completely!
    - Is responsible for leaves staying on

TDZ Affects the Shikimate Pathway

Kynurenine stuff

- Oxidation product of tryptophan

On TDZ

~4.5 million kg of TDZ used per year IN THE US

- We barely know anything about it
- It does a lot in plants
- What does it do in us?
- We spray on:
    - Apples (Bloop),
    - Canola,
    - Cotton,
    - Pears (Bloom).

Purpose

We will be making something just like this.

- ☐ Double check the case study slides

Assignment 1, Part B

How to count features:

- Count signals in each column (treatment across replicates).

Add and subtract things that enzymes can do.
Pick a molecule, and check out how it is affected.

- Look for its modification
- ☐ Make a script to add and subtract masses to stuff

Enzymes are super simple. They do serial things - add or take *something*.

- Predictable by change in mass: (±)
  - NH2
  - Carboxy
  - Acetyl
  - Proton
  - Hydroxy
  - Glucose
  -
- Isomerization can be looked at using same mass at different retention times.
- ☐ Look at Wikipathways or Kegg for pathways
- ☐ Actually look at the data!!!

NMR Based

Shipley sometime soon.