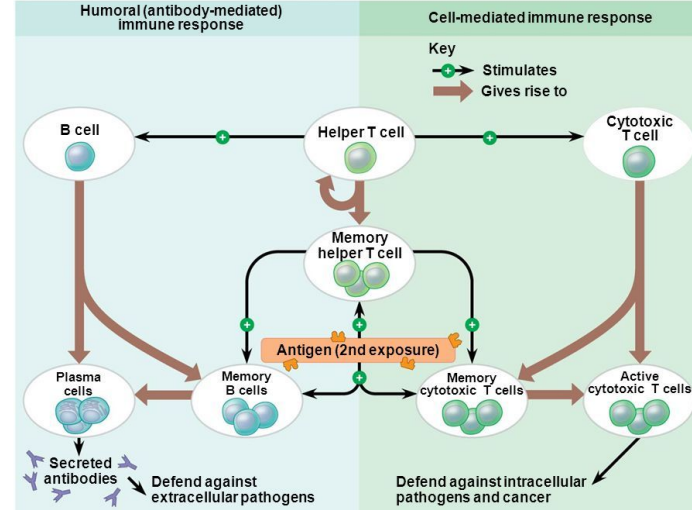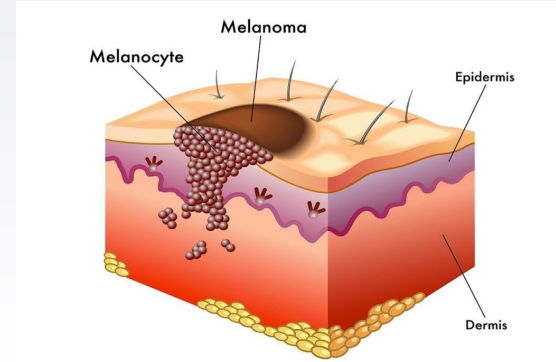# Application of JMP to Analyze Gene Expression Single-Cell RNA-seq Data in Melanoma Cells

**Catherine Zhou**

Lynbrook High School

# Introduction

- Melanoma - most aggressive type of skin cancer
  - Represents 65% of all deaths from skin cancer
- T cells
  - Type of white blood cell
  - Different types →
  - Mount response against tumor, but become ineffective

# Single-cell RNA-sequencing

- ▸ Alternative splicing
- ▸ Bulk RNA sequencing vs single-cell RNA-sequencing (scRNA-seq)
- ▸ Issues arise when studying AS in scRNA-seq data



Brulois et al., Nature (2020)
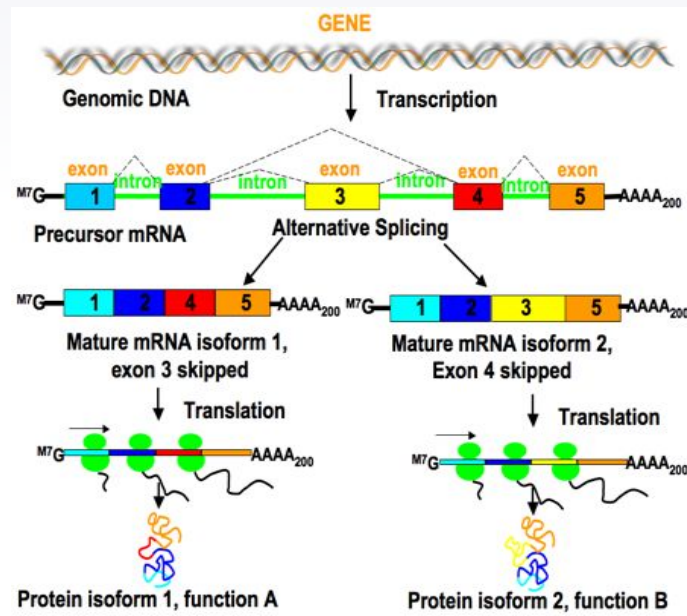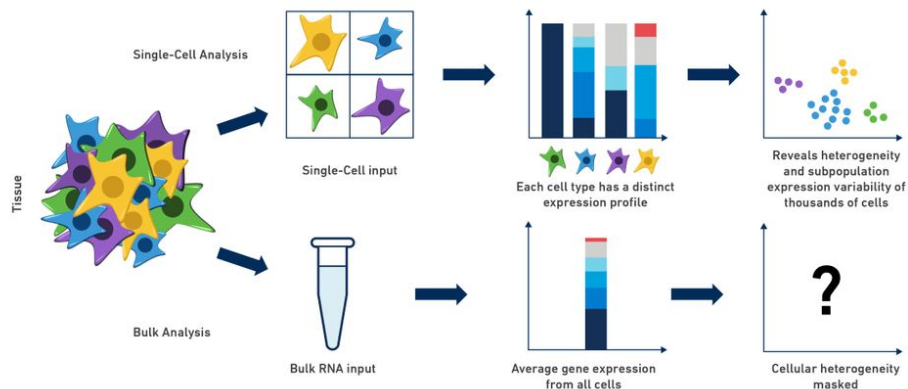
# Presentation Overview

▸ Question: Using scRNA-seq data, how can we find the most significant alternative splicing (AS) events differentiating cancerous T cells from healthy lymph node T cells?

  ▸ Data processing in R
  ▸ **JMP Pro:** Predictive modeling, clustering, visualization
  ▸ Results: Initial analysis of scRNA-seq dataset of T cells in murine melanoma sample
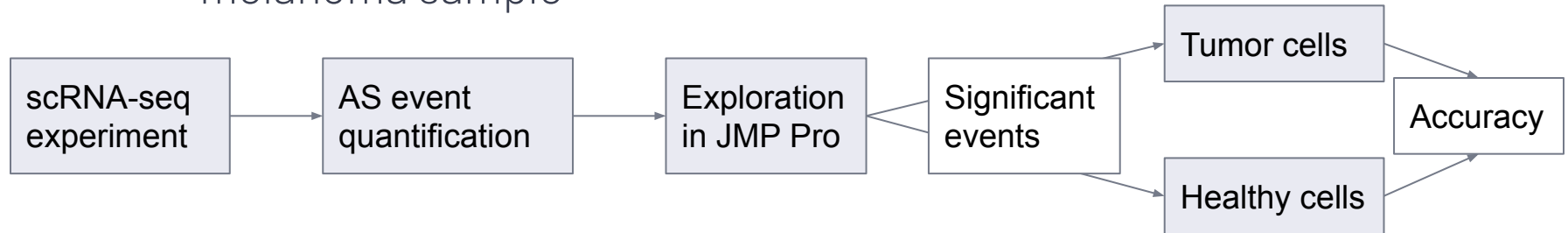
# Dataset Preparation and Processing

- Used an scRNA-seq dataset of cells in lymph nodes with melanoma in mice[1]
- Read alignment with STAR
- Detection of AS events with pipeline derived from rMATS
  - Generalized linear mixed model
  - Single sample with each cell
- Quantified exon skipping events with lJC and SJC
- Created matrix of all exons, cell barcodes, and lJC and SJC counts



junction reads of inclusion isoform
exon reads of inclusion isoform
junction reads of skipping isoform
exon reads of skipping isoform

[1]Davidson, Sarah et al. *Cell reports* 31,7 **2020**

# Dataset Processing

- Used cell labels previously defined in dataset
- Filtering AS events with R
  - At least 10 reads per junction
  - AS event has to be detected in >10 cells
  - AS event cannot have no variability across cells
  - Removed ~75% of exons
- Calculated PSI value = IJC/(IJC+SJC)
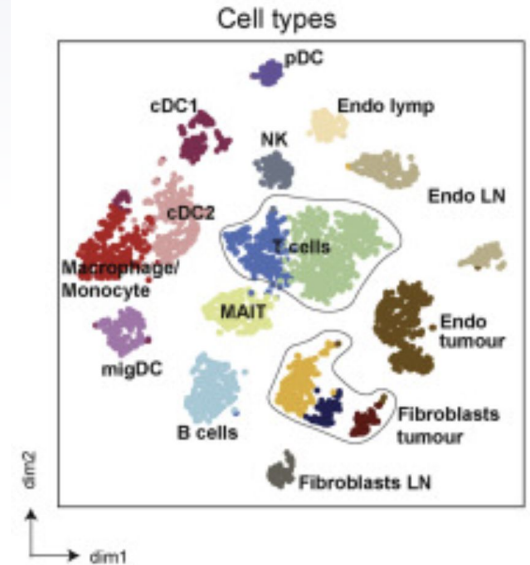
# Application of Predictive Modeling in JMP

- Success of previous studies and models
- Advantages of using predictive modeling
  - Extract meaningful features
  - Classify data and predict outcomes with supervised learning
  - Recognize underlying relationship of data
- JMP Pro provides great interface for exploring different models
  - Also helps reduce "black box" effect
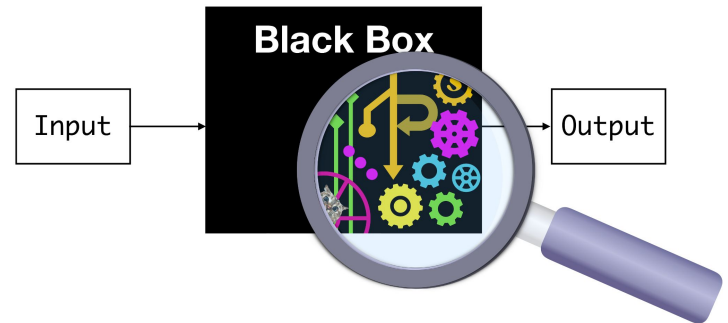
# Predictive Modeling - Bootstrap forest

- Dataset size: 8044 variables (exons), 1014 rows (sample cells)
- Difficult to run analysis with large number of columns
- Bootstrap forest for variable selection
  - Predictor screening - same function, slower
- Used tuning design table
  - Computationally efficient
  - Tuned number of trees (doesn't make huge difference)
- Created validation column
- Took top 100 exons, 50 exons, 10 exons

# Bootstrap Forest - Accuracy

## Tuning # of trees

**Model Validation-Set Summaries**

The fit below was the best of these models fit.

| N Terms | N Trees | N Trees Specified | Entropy RSquare | Misclassification Rate | Avg -Log p | RMS Error | Avg Abs Error |
|---|---|---|---|---|---|---|---|
| 3529 | 50 | 50 | 0.6258 | 0.0788 | 0.2318 | 0.2554 | 0.1758 |
| 3529 | 100 | 100 | 0.5982 | 0.0837 | 0.2489 | 0.2633 | 0.1881 |
| 3529 | 200 | 200 | 0.6083 | 0.0788 | 0.2426 | 0.2593 | 0.1809 |
| 3529 | 300 | 300 | 0.6073 | 0.0690 | 0.2432 | 0.2591 | 0.1842 |
| 3529 | 400 | 400 | 0.6045 | 0.0739 | 0.2450 | 0.2591 | 0.1850 |
| 3529 | 500 | 500 | 0.6100 | 0.0788 | 0.2416 | 0.2576 | 0.1838 |

**Specifications**

| | | | |
|---|---|---|---|
| Target | cell_type | Training Rows: | 608 |
| Validation Column: | Validation | Validation Rows: | 203 |
| | | Test Rows: | 203 |
| Number of Trees in the Forest: | 50 | Number of Terms: | 8044 |
| Number of Terms Sampled per Split: | 3529 | Bootstrap Samples: | 608 |
| | | Minimum Splits per Tree: | 10 |
| | | Minimum Size Split: | 5 |

## Variable selection (validation set metrics)

**All exons**

| N Terms | N Trees | N Trees Specified | Entropy RSquare | Misclassification Rate | Avg -Log p | RMS Error | Avg Abs Error |
|---|---|---|---|---|---|---|---|
| 3529 | 50 | 50 | 0.6042 | 0.0690 | 0.2452 | 0.2587 | 0.1885 |

**Top 100**

| N Terms | N Trees | N Trees Specified | Entropy RSquare | Misclassification Rate | Avg -Log p | RMS Error | Avg Abs Error |
|---|---|---|---|---|---|---|---|
| 68 | 30 | 30 | 0.6310 | 0.0788 | 0.2285 | 0.2520 | 0.1547 |

**Top 50**

| N Terms | N Trees | N Trees Specified | Entropy RSquare | Misclassification Rate | Avg -Log p | RMS Error | Avg Abs Error |
|---|---|---|---|---|---|---|---|
| 37 | 40 | 40 | 0.6472 | 0.0640 | 0.2185 | 0.2474 | 0.1559 |

**Top 10**

| N Terms | N Trees | N Trees Specified | Entropy RSquare | Misclassification Rate | Avg -Log p | RMS Error | Avg Abs Error |
|---|---|---|---|---|---|---|---|
| 8 | 30 | 30 | 0.6491 | 0.0837 | 0.2173 | 0.2537 | 0.1481 |

# Model Screening

- Ran top 10 exons on all models
  - Resulted in high accuracy across the board
- Prediction profiler
- Demo (Very quick!)

# Clustering

- ▶ K-means clustering of all cells
  - ▷ More accurately classifies tumor vs normal cells
- ▶ UMAP on all cells
  - ▷ Shows potential new subsets
  - ▷ Created using JMP add-in
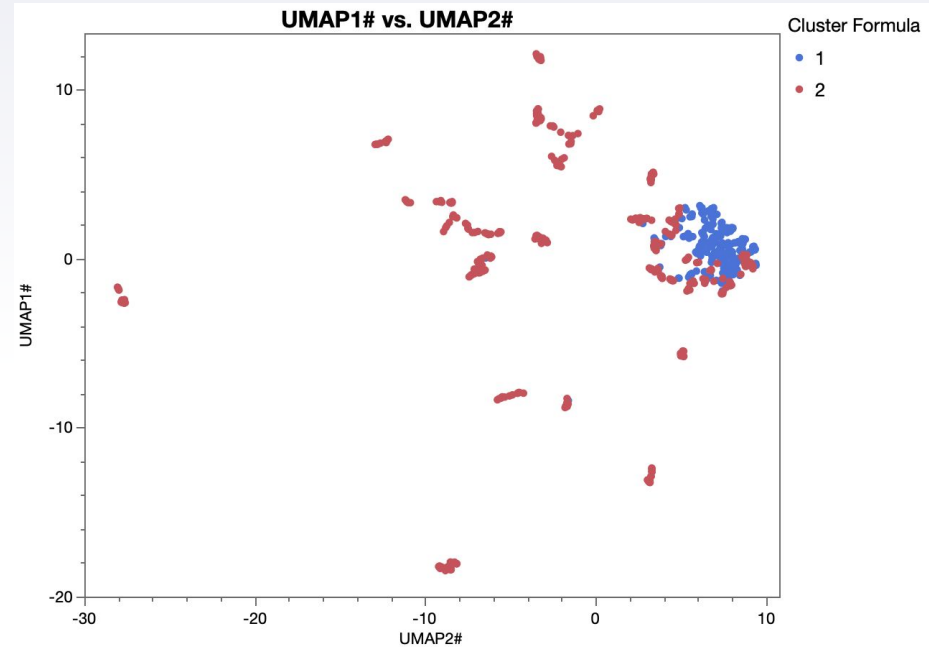
# Gene enrichment analysis

| Index | Name | P-value | Adjusted p-value | Odds Ratio | Combined score |
|-------|------|---------|------------------|------------|----------------|
| 1 | RNA import into mitochondrion (GO:0035927) | 0.002248 | 0.05604 | 624.59 | 3808.55 |
| 2 | regulation of leukocyte degranulation (GO:0043300) | 0.004043 | 0.05604 | 312.23 | 1720.62 |
| 3 | regulation of myeloid leukocyte mediated immunity (GO:0002886) | 0.004940 | 0.05604 | 249.76 | 1326.33 |
| 4 | regulation of mast cell activation involved in immune response (GO:0033006) | 0.004940 | 0.05604 | 249.76 | 1326.33 |
| 5 | positive regulation of protein maturation (GO:1903319) | 0.005388 | 0.05604 | 227.05 | 1185.99 |
| 6 | regulation of mast cell degranulation (GO:0043304) | 0.007178 | 0.05735 | 166.47 | 821.79 |
| 7 | protein import into mitochondrial matrix (GO:0030150) | 0.008519 | 0.05735 | 138.70 | 660.97 |
| 8 | positive regulation of protein processing (GO:0010954) | 0.008966 | 0.05735 | 131.39 | 619.44 |
| 9 | regulation of protein processing (GO:0070613) | 0.01120 | 0.05735 | 103.99 | 467.16 |
| 10 | protein transmembrane import into intracellular organelle (GO:0044743) | 0.01431 | 0.05735 | 80.48 | 341.79 |

Malignant tumors selectively retain mitochondrial genome and ETC function

T-cells are a type of leukocyte

# Gene enrichment analysis
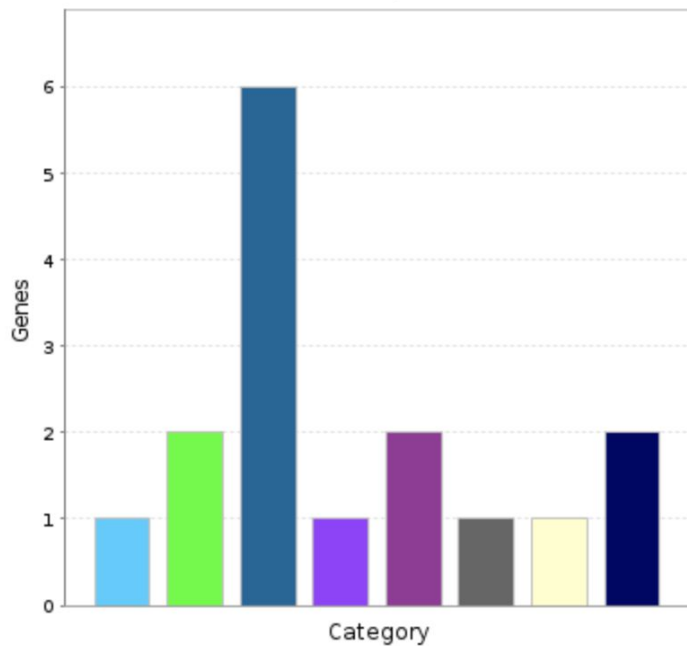
# Notable genes

- Stpbx2
  - Involved in intracellular trafficking, control of SNARE (soluble NSF attachment protein receptor) complex assembly, and the release of cytotoxic granules by natural killer cells
- Slamf9
  - Encodes a member of the signaling lymphocytic activation molecule family, transmembrane
- Wars
  - Tryptophanyl-tRNA synthetase (WARS) catalyzes the aminoacylation of tRNA(trp) with tryptophan and is induced by interferon

# Analysis of tumor gene expression over time



Mean(3110053B16Rik_chr12_-_20964759_20964842) & 9 more vs. time

- Mean(3110053B16Rik_chr12_-_20964759_20964842)
- Mean(Apobr_chr7_+_126187068_126187334)
- Mean(Nkd2_chr13_-_73971390_73971475)
- Mean(Pcdh9_chr14_-_93797911_93798013)
- Mean(Slamf9_chr1_+_172303699_172304041)
- Mean(Stxbp2_chr8_+_3682499_3682600)
- Mean(Tomm20_chr8_-_127667928_127667975)
- Mean(Wars_chr12_-_108850190_108850274)
- Mean(Wars_chr12_-_108854237_108854474)
- Mean(Zfp125_chr12_-_20964759_20964842)

▸ Tomm20

  ▷ Inclusion of exon greatly increases over time

  ▷ Translocase of the outer mitochondrial membrane complex subunit 20 (TOMM20) facilitates cancer aggressiveness and therapeutic resistance in chondrosarcoma.

  ▷ Roche ME, *et al*. Biochim Biophys Acta Mol Basis Dis, 2020 Dec 1. PMID 32920118

# Conclusion

- Exploration of dataset facilitated by JMP demonstrates role of genes/exons in T cells in melanoma
- Code can be replicated in Python for more robust/detailed analysis
- Potential genes can aid in the discovery of novel and personalized approaches to cancer treatment
- Can perform in-vitro testing on top genes