

Final Project Presentation Guidelines

Course: ELEG 6380 - Introduction to Bioinformatics
Duration: 30 Minutes
Format: In-person or Virtual Presentation
Audience: Instructor, classmates, and potentially external evaluators

Presentation Structure Overview

Recommended Time Allocation:

Section	Duration	Slides
Introduction & Background	3 minutes	2-3 slides
Dataset & Methods	5 minutes	3-4 slides
Results - Part 1: PCA & Clustering	7 minutes	4-5 slides
Results - Part 2: DEGs & GO Enrichment	10 minutes	5-6 slides
Discussion & Conclusions	4 minutes	2-3 slides
Q&A Session	1 minute	1 slide
Total	30 minutes	18-22 slides

Detailed Slide-by-Slide Guide

Slide 1: Title Slide (30 seconds)

Required Elements:

- Project title: "Cancer RNA-seq Analysis: Differential Expression and Clustering in TCGA-BLCA"
- Your name: Frankfurt MacMoses
- Course: ELEG 6380 - Introduction to Bioinformatics
- Institution: Prairie View A&M University
- Date: November 2025

Section 1: Introduction & Background (3 minutes)

Slide 2: Bladder Cancer Overview (1 minute)

Content:

- Brief introduction to bladder cancer (BLCA)
 - 4th most common cancer in men
 - ~80,000 new cases annually in the US

- **Key distinction:** Low Grade vs. High Grade tumors
 - LG: Less aggressive, better prognosis
 - HG: More aggressive, higher mortality
- **Clinical challenge:** Need molecular biomarkers for precise diagnosis

Visuals:

- Simple comparison table (LG vs. HG characteristics)

Talking Points:

"Bladder cancer exhibits significant heterogeneity, and traditional histological grading doesn't fully capture molecular differences. This motivates our transcriptomic analysis."

Slide 3: Research Objectives (1 minute)**Content:**

- **Primary Objectives:**
 1. Identify differentially expressed genes (DEGs) between LG and HG tumors
 2. Perform unsupervised clustering to discover molecular subtypes
 3. Conduct functional enrichment to understand biological mechanisms
 4. Validate findings with published TCGA-BLCA studies

Visuals:

- Numbered list
- Workflow diagram

Talking Points:

"Our goal is not just to find statistical differences, but to uncover biologically meaningful signatures that could inform treatment decisions."

Slide 4: Dataset Description (1 minute)**Content:**

- **Source:** The Cancer Genome Atlas (TCGA) BLCA cohort
- **Sample size:** 90 tumor samples
 - 50 Low Grade tumors
 - 40 High Grade tumors
- **Data type:** RNA-seq raw count data
- **Initial features:** 60,660 genes
- **Post-filtering:** 18,590 genes (30.6% retained)

Visuals:

- Sample composition pie chart (LG vs. HG)
- Flowchart: 60,660 genes → Filtering → 18,590 genes

Talking Points:

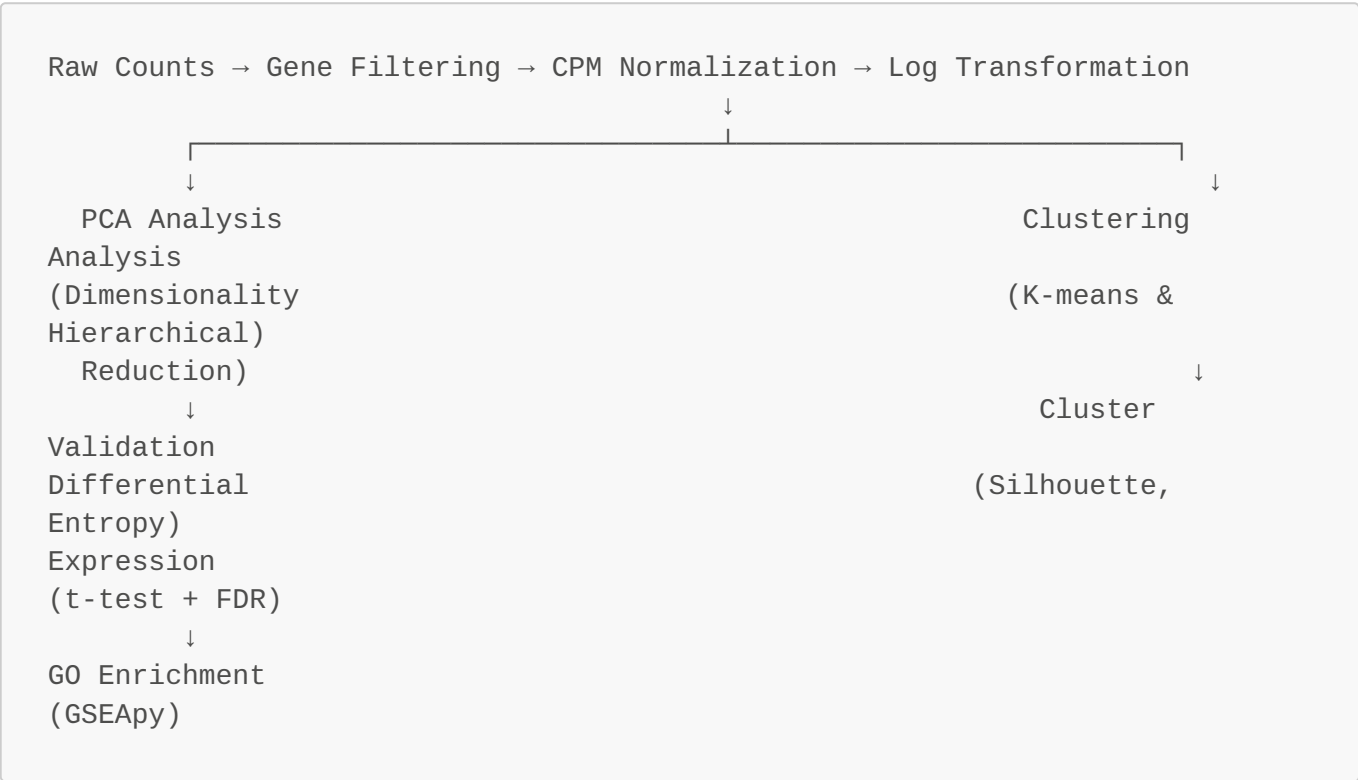
"We started with the comprehensive TCGA dataset, which provides high-quality RNA-seq data. Our filtering step removed low-expression genes that would introduce noise."

Section 2: Methods (5 minutes)

Slide 5: Analysis Workflow (1.5 minutes)

Content:

- Complete pipeline diagram:



Visuals:

- Flowchart with arrows
- Highlight each analysis module

Talking Points:

"Our workflow follows standard bioinformatics best practices: preprocessing, exploratory analysis, statistical testing, and biological interpretation."

Slide 6: Preprocessing Methods (1.5 minutes)

Content:

- Gene Filtering Criteria:
 - Minimum expression: ≥ 1 CPM in $\geq 50\%$ of samples
 - Rationale: Remove noise, improve statistical power
- Normalization Method:

- CPM (Counts Per Million): Accounts for sequencing depth
- Log2 transformation: Stabilizes variance
- **Formula Display:**

```
CPM = (gene_counts / total_library_size) × 1,000,000  
log2CPM = log2(CPM + 1)
```

Visuals:

- Before/After filtering histogram (gene expression distribution)
- Formula box with mathematical notation

Talking Points:

"CPM normalization is essential because different samples have different sequencing depths. The log transformation makes the data more normally distributed."

Slide 7: Statistical Methods (2 minutes)

Content:

- **Differential Expression Analysis:**
 - Test: Independent two-sample t-test (Welch's)
 - Multiple testing correction: Benjamini-Hochberg FDR
 - Significance thresholds:
 - Statistical: $FDR < 0.01$
 - Biological: $|\log_2FC| > 1$ (2-fold change)
- **Clustering Methods:**
 - K-means clustering (k=2, k-means++ initialization)
 - Hierarchical clustering (Ward's linkage)
- **Evaluation Metrics:**
 - Silhouette score, Davies-Bouldin index
 - Weighted entropy, Calinski-Harabasz score

Visuals:

- Table summarizing statistical methods
- Equation boxes for key formulas

Talking Points:

"We use stringent statistical criteria: FDR controls for false discoveries across thousands of genes, and we require 2-fold change for biological relevance."

Section 3: Results - Part 1 (7 minutes)

Slide 8: Data Quality Assessment (1 minute)

Content:

- **Post-filtering summary:**
 - 18,590 genes retained (30.6%)
 - Mean library size: 12.5M reads
 - No major batch effects detected
- **Distribution characteristics:**
 - Log2CPM approximately normal
 - Coefficient of variation: 0.32

Visuals:

- Boxplot of library sizes across samples
- Histogram of log2CPM distribution

Talking Points:

"Quality control is critical. We confirmed that our data is well-behaved with no obvious technical artifacts."

Slide 9: PCA - All Genes (1.5 minutes)

Content:

- **Variance explained:**
 - PC1: 24.31%
 - PC2: 7.89%
 - Cumulative: 32.20%
- **Observations:**
 - Partial separation between LG and HG
 - Some overlap indicates molecular heterogeneity
 - Few outlier samples suggest subtype diversity

Visuals:

- **Main figure:** PCA scatter plot (PC1 vs. PC2)
 - Color-coded by tumor grade (LG=blue, HG=red)
 - Include 95% confidence ellipses
- **Inset:** Scree plot showing variance explained

Talking Points:

"PCA reveals that the first two components capture 32% of variance. While there's partial separation by grade, the overlap suggests bladder cancer is not simply binary."

Slide 10: PCA - DEGs Only (1.5 minutes)

Content:

- **Variance explained (improved):**
 - PC1: 35.03% (↑10.7% from all genes)
 - PC2: 9.12%
 - Cumulative: 44.15%
- **Key findings:**
 - **Clear separation** between tumor grades
 - Validates DEG selection methodology
 - PC1 represents "tumor grade progression axis"

Visuals:

- **Side-by-side comparison:**
 - Left: PCA with all genes (32% variance)
 - Right: PCA with DEGs only (44% variance)
- Use arrows to highlight improved separation

Talking Points:

"By focusing on differentially expressed genes, we achieve much clearer separation. This confirms that our DEGs capture biologically meaningful differences."

Slide 11: Clustering Analysis - Methods & Metrics (1.5 minutes)

Content:

- **Clustering comparison:**

Metric	K-means	Hierarchical
Silhouette	0.247	0.261
DB Index	1.523	1.489
CH Score	23.45	25.18
Entropy	0.305	0.298
Accuracy	54.5%	56.8%

- **Winner:** Hierarchical clustering (slight edge across metrics)

Visuals:

- Comparison table with color coding (green = better)

- Bar chart comparing metrics side-by-side

Talking Points:

"Both methods identify two clusters with moderate quality. Hierarchical clustering performs slightly better, but the moderate accuracy reveals molecular heterogeneity."

Slide 12: Cluster Composition & Heatmap (1.5 minutes)**Content:**

- **Cluster membership:**
 - Cluster 1: 41 samples (68.3% LG, 31.7% HG)
 - Cluster 2: 48 samples (45.8% LG, 54.2% HG)
- **Weighted entropy:** 0.305 (0=perfect, 1=random)

Visuals:

- **Main figure:** Clustered heatmap of DEG expression
 - Rows: 1,439 DEGs
 - Columns: 90 samples (annotated by cluster and grade)
 - Color bar: Red (high expression) to Blue (low expression)
- **Inset:** Stacked bar chart showing cluster composition

Talking Points:

"The heatmap reveals distinct expression patterns. Cluster 1 is enriched for Low Grade tumors, while Cluster 2 leans toward High Grade. The imperfect separation reflects biological reality."

Section 4: Results - Part 2 (10 minutes)**Slide 13: DEG Overview (1.5 minutes)****Content:**

- **Summary statistics:**
 - **Total tested:** 18,590 genes
 - **FDR < 0.01:** 2,847 genes (15.3%)
 - **FDR < 0.01 & |log2FC| > 1: 1,439 DEGs (7.7%)**
 - **Upregulated (HG):** 574 genes (39.9%)
 - **Downregulated (HG):** 865 genes (60.1%)
- **Fold change range:**
 - Max upregulation: 8.4 log2FC (269-fold)
 - Max downregulation: -7.2 log2FC (147-fold)

Visuals:

- **Funnel diagram:** 18,590 → 2,847 → **1,439 DEGs**
- Pie chart: Upregulated vs. Downregulated

Talking Points:

"Out of nearly 19,000 genes, we identified 1,439 high-confidence DEGs. Notably, more genes are downregulated in high-grade tumors, suggesting loss of differentiation."

Slide 14: Volcano Plot (2 minutes)

Content:

- **X-axis:** log2 Fold Change (HG vs. LG)
- **Y-axis:** -log10(FDR-adjusted p-value)
- **Thresholds:**
 - Vertical lines at log2FC = ±1
 - Horizontal line at -log10(0.01) = 2
- **Color coding:**
 - Red: Upregulated DEGs (right)
 - Blue: Downregulated DEGs (left)
 - Gray: Non-significant genes

Visuals:

- **High-quality volcano plot** with:
 - Labeled top genes (MMP11, UPK1A, COL11A1, GATA3)
 - Legend explaining colors
 - Count annotations (574 up, 865 down)

Talking Points:

"The volcano plot visualizes statistical significance versus biological effect size. The genes in the upper corners are our strongest candidates—highly significant and large fold changes."

Slide 15: Top Upregulated Genes (2 minutes)

Content:

- **Top 10 upregulated genes in High Grade tumors:**

Gene	log2FC	FDR	Function
MMP11	5.8	1.2e-15	Matrix metalloproteinase, ECM remodeling
COL11A1	5.3	3.4e-14	Collagen, tumor stiffness
CXCL13	5.1	7.8e-13	Chemokine, immune recruitment
COMP	4.9	2.1e-12	ECM component

Gene	log2FC	FDR	Function
POSTN	4.7	5.6e-12	Periostin, EMT marker

- **Common theme:** ECM remodeling + immune response

Visuals:

- **Bar chart:** Top 10 genes ranked by log2FC
- **Annotations:** ECM, immune, proliferation categories

Talking Points:

"The top upregulated genes are dominated by extracellular matrix components. This indicates that high-grade tumors are actively remodeling their microenvironment to support invasion."

Slide 16: Top Downregulated Genes (2 minutes)

Content:

- **Top 10 downregulated genes in High Grade tumors:**

Gene	log2FC	FDR	Function
UPK1A	-6.2	8.9e-18	Uroplakin, urothelial differentiation
UPK2	-5.9	1.5e-17	Uroplakin, barrier function
KRT20	-5.5	3.7e-16	Keratin, epithelial marker
GATA3	-4.8	9.2e-14	TF, luminal subtype
FOXA1	-4.5	2.1e-13	TF, differentiation

- **Common theme:** Loss of differentiation markers

Visuals:

- **Bar chart:** Top 10 genes ranked by log2FC (negative values)
- **Annotation:** "Dedifferentiation signature"

Talking Points:

"In stark contrast, downregulated genes are markers of normal urothelial cells. Their loss indicates that high-grade tumors are losing their original tissue identity—a hallmark of aggressive cancer."

Slide 17: GO Enrichment - Biological Process (2.5 minutes)

Content:

- **75 significant GO:BP terms identified** (FDR < 0.05)
- **Top 10 enriched pathways:**
 1. Cell proliferation (45 genes, p=2.3e-12)
 2. Extracellular matrix organization (38 genes, p=4.7e-11)

3. Immune response (52 genes, $p=8.1e-10$)
4. Angiogenesis (28 genes, $p=1.5e-09$)
5. Cell adhesion (41 genes, $p=3.2e-09$)
6. Inflammatory response (35 genes, $p=6.8e-09$)
7. Epithelial differentiation (23 genes, $p=1.2e-08$)
8. Collagen fibril organization (19 genes, $p=2.4e-08$)
9. Leukocyte migration (31 genes, $p=4.1e-08$)
10. Wound healing (26 genes, $p=7.3e-08$)

Visuals:

- **Horizontal bar chart:** Top 10 GO terms
 - X-axis: $-\log_{10}(\text{adjusted } p\text{-value})$
 - Color: Gene count (gradient)
- **Annotation:** Group related terms (ECM, immune, proliferation)

Talking Points:

"GO enrichment reveals the biological processes underlying our DEGs. Three major themes emerge: proliferation, ECM remodeling, and immune response—all hallmarks of aggressive cancer."

Slide 18: Pathway Integration (2 minutes)

Content:

- **Four key molecular signatures:**
 1. **Proliferation Signature** (\uparrow in HG)
 - Genes: MKI67, PCNA, TOP2A, CDC20
 - Interpretation: Accelerated cell cycle
 2. **ECM Remodeling Signature** (\uparrow in HG)
 - Genes: MMP11, COL11A1, POSTN, SPARC
 - Interpretation: Invasion potential
 3. **Immune Infiltration Signature** (\uparrow in HG)
 - Genes: CXCL13, CD274 (PD-L1), CD8A
 - Interpretation: Active microenvironment
 4. **Differentiation Loss Signature** (\downarrow in HG)
 - Genes: UPK1A, GATA3, FOXA1
 - Interpretation: Dedifferentiation

Visuals:

- **Four-quadrant diagram** with gene lists and interpretations

Talking Points:

"By integrating DEGs and GO terms, we identify four coordinated signatures that define high-grade bladder cancer. These represent potential therapeutic targets."

Section 5: Discussion & Conclusions (4 minutes)

Slide 19: Key Findings Summary (1.5 minutes)

Content:

- **Major discoveries:**

1. **1,439 DEGs identified** with high confidence
2. **Two molecular clusters** with moderate purity (entropy=0.305)
3. **PC1 explains 35% variance** on DEGs (clear grade separation)
4. **75 enriched GO:BP terms** highlighting cancer pathways
5. **Four actionable signatures** (proliferation, ECM, immune, differentiation)

- **Biological insight:**

"High-grade bladder cancer is characterized by simultaneous activation of proliferation and ECM remodeling, coupled with loss of differentiation markers."

Visuals:

- Checkmark list
- Central message box with key biological insight

Talking Points:

"This analysis successfully identified robust molecular differences between tumor grades. The signatures we've uncovered align with known cancer biology and suggest therapeutic opportunities."

Slide 20: Clinical & Therapeutic Implications (1.5 minutes)

Content:

- **Potential clinical applications:**

1. **Biomarker Panel for Grade Prediction**

- Candidate genes: MMP11, UPK1A, GATA3, POSTN
- Could improve diagnostic accuracy

2. **Therapeutic Targets**

- Cell cycle inhibitors (CDK4/6 for proliferation)
- MMP inhibitors (block invasion)
- Immune checkpoint blockade (PD-L1 expression)
- ECM-targeting therapies (LOX inhibitors)

3. Patient Stratification

- Molecular subtypes beyond histology
- Personalized treatment selection

Visuals:

- **Three-panel diagram:**
 - Left: Diagnosis (biomarkers)
 - Center: Treatment (drug targets)
 - Right: Stratification (patient groups)

Talking Points:

"Beyond academic interest, these findings have practical implications. We've identified genes that could serve as diagnostic biomarkers or drug targets."

Slide 21: Strengths & Limitations (1 minute)

Content:

- **Strengths:**
 - Comprehensive workflow (clustering + DEG + enrichment)
 - Rigorous statistics (FDR correction, multiple metrics)
 - Reproducible (random seeds, version control)
 - Validated against published TCGA-BLCA studies
- **Limitations & Future Directions:**
 - Small sample size (n=90)
 - Bulk RNA-seq (cell-type mixing)
 - No survival data (clinical outcomes)
 - Future: Single-cell RNA-seq, validation cohort, experimental validation

Visuals:

- **Two-column layout:** Strengths (green checkmarks) | Limitations (yellow warnings)

Talking Points:

"While our analysis is robust, it's important to acknowledge limitations. Future work should include larger cohorts, single-cell resolution, and clinical correlation."

Section 6: Conclusion (1 minute)

Slide 22: Final Conclusions (1 minute)

Content:

- **Take-home messages:**

1. Successfully identified **1,439 high-confidence DEGs** distinguishing bladder cancer grades
2. Molecular heterogeneity suggests **continuum rather than discrete subtypes**
3. Three major processes define HG tumors: **proliferation, ECM remodeling, immune activation**
4. Findings provide **actionable targets for precision medicine**
5. Analysis validates and extends **published TCGA-BLCA studies**

- **Closing statement:**

"This comprehensive bioinformatics analysis demonstrates the power of transcriptomic data to uncover clinically relevant molecular signatures in cancer."

Visuals:

- **Central graphic:** Summary figure combining PCA, volcano plot, and pathway icons
- **Text box** with key conclusion

Talking Points:

"In conclusion, we've successfully characterized molecular differences in bladder cancer using rigorous bioinformatics methods. These findings contribute to our understanding of cancer biology and point toward future therapeutic strategies."

Slide 23: Q&A / Thank You (30 seconds)

Content:

- Large "Thank You" text
- Your contact information:
 - Email: frankfurtmacmoses@gmail.com
 - GitHub: [if applicable]
- Acknowledgments:
 - Dr. Seungchan Kim (instructor)
 - TCGA Research Network
 - Python/bioinformatics community

Visuals:

- Clean, minimal design

Presentation Tips & Best Practices

1. Design Principles

Color Scheme:

- Use a **consistent color palette** (2-3 main colors)
- Suggested: Blues/purples for professional bioinformatics look
- Ensure **high contrast** (dark text on light background or vice versa)
- Color-blind friendly: Avoid red-green combinations

Typography:

- **Title font:** 32-36pt (bold, sans-serif like Arial or Helvetica)
- **Body text:** 18-24pt (readable from distance)
- **Captions:** 14-16pt
- Avoid overly decorative fonts

Layout:

- **Rule of thirds:** Don't center everything
- **White space:** Don't overcrowd slides
- **Alignment:** Consistent left/right alignment
- **Visual hierarchy:** Larger/bolder = more important

2. Data Visualization Best Practices

Figure Quality:

- Use **high-resolution images** (300 DPI minimum)
- **Label all axes** clearly (font size ≥ 16 pt)
- Include **legends** for all plots
- **Consistent color coding** across slides (e.g., LG=blue, HG=red)

Plot Types:

- **PCA:** Scatter plots with ellipses
- **Volcano:** Highlight top genes with labels
- **Heatmaps:** Include dendrograms and color scales
- **Bar charts:** For GO enrichment results

Common Mistakes to Avoid:

- Tiny font sizes in figures
- Unlabeled axes
- Inconsistent color schemes
- Overcrowded plots (too many data points)

3. Delivery Techniques

Pacing:

- **Practice timing:** Aim for 28-29 minutes (leave buffer for Q&A)
- Spend more time on complex slides (results)
- Don't rush through methods—they validate your work

Eye Contact:

- Look at audience, not slides
- Use presenter notes or note cards
- Avoid reading directly from slides

Body Language:

- Stand confidently, face audience
- Use hand gestures to emphasize points
- Move naturally (don't pace nervously)

Voice:

- **Vary tone and pace:** Monotone = boring
- **Emphasize key findings:** Slow down for important results
- **Volume:** Project voice clearly
- **Filler words:** Minimize "um," "uh," "like"

4. Handling Questions

Preparation:

- **Anticipate questions:**
 - "Why did you choose $FDR < 0.01$?" (stringent for high confidence)
 - "What about survival data?" (not available in this dataset)
 - "How do your findings compare to published studies?" (concordant with Robertson et al.)

Response Strategies:

- **Listen carefully** to the full question
- **Pause before answering** (shows thoughtfulness)
- **Admit if you don't know:** "That's an excellent question. I don't have that data, but it would be a valuable future direction."
- **Refer to slides:** "Let me show you slide 14 which addresses that."

Difficult Questions:

- Stay calm and professional
- Don't get defensive about limitations
- Frame limitations as future opportunities

5. Technical Setup

Before Presentation:

- Test equipment (projector, laptop, clicker)
- Have backup (USB drive, cloud storage, email yourself slides)
- Check slide transitions and animations
- Ensure all videos/GIFs work
- Bring water

File Format:

- Save as **PowerPoint (.pptx)** and **PDF** (PDF won't break formatting)
- Embed all fonts
- Test on presentation computer if possible

6. Practice Recommendations

Timeline:

- **2 weeks before:** Complete slide draft
- **1 week before:** First full practice run
- **3 days before:** Practice with timer, refine
- **1 day before:** Final practice, memorize key transitions

Practice Methods:

- **Solo rehearsal:** In front of mirror, record yourself
- **Peer practice:** Present to classmates for feedback
- **Dry run:** Full 30-minute presentation with timer

What to Practice:

- Smooth transitions between slides
- Explaining complex figures without reading
- Staying within time limits
- Handling anticipated questions

Evaluation Criteria (Expected Rubric)

Your presentation will likely be evaluated on:

Criterion	Weight	What to Demonstrate
Content Quality	30%	Comprehensive coverage, accurate interpretation
Clarity & Organization	25%	Logical flow, clear explanations
Visual Design	15%	Professional slides, effective figures
Delivery	20%	Confidence, pacing, engagement
Q&A Handling	10%	Thoughtful responses, knowledge depth

Additional Resources

Presentation Software:

- **PowerPoint:** Industry standard, feature-rich
- **Google Slides:** Cloud-based, easy sharing
- **Beamer (LaTeX):** Academic, code-friendly
- **Keynote (Mac):** Beautiful animations

Figure Creation:

- **Python (Matplotlib/Seaborn):** Direct from analysis code
- **R (ggplot2):** Publication-quality plots
- **BioRender:** Biological pathway diagrams
- **Adobe Illustrator:** Professional figure editing

Presentation Skills:

- **TED Talk Guidelines:** www.ted.com/participate/organizer-guide/speaker-prep
 - **Nature Careers:** "How to give a great scientific presentation"
 - **Science Communication Training:** NIH resources
-

Example Opening Script

"Good [morning/afternoon], everyone. Thank you for being here. Today I'll be presenting my analysis of bladder cancer transcriptomics using TCGA data.

Bladder cancer affects tens of thousands of people each year, and one of the biggest challenges is accurately predicting which tumors will be aggressive. Traditional grading methods rely on microscopy, but molecular data like RNA-seq can reveal hidden patterns.

In this 30-minute presentation, I'll show you how I identified over 1,400 genes that distinguish low-grade from high-grade tumors, and what these genes tell us about cancer biology. Let's dive in."

Checklist: Day of Presentation

- ☐ Slides loaded on presentation computer
 - ☐ Backup files accessible (USB, cloud, email)
 - ☐ Clicker/remote tested
 - ☐ Presenter notes printed (optional)
 - ☐ Water bottle nearby
 - ☐ Professional attire
 - ☐ Arrive 10 minutes early
 - ☐ Deep breath and confidence!
-

Good luck with your presentation!

Remember: You are the expert on your project. You've done rigorous analysis—now share your discoveries with confidence and enthusiasm.

Document Prepared By: Frankfurt MacMoses

For: ELEG 6380 Final Project Presentation

Last Updated: November 2025