Example Prompts for Spatial MCP POC

Bioinformatics Workflow Investigation Guide

Version: 1.0
Date: October 24, 2025
Purpose: Example prompts demonstrating how bioinformaticians can interact with the Spatial MCP POC

Stage 1: Data Ingestion & Quality Control

Example 1.1: Initial FASTQ Quality Assessment

> "I have a new spatial transcriptomics sample with files spatial_sample_001_R1.fastq.gz and spatial_sample_001_R2.fastq.gz. Can you validate these FASTQ files using FGbio, check the quality metrics, and tell me if the data is suitable for downstream analysis? I'm particularly concerned about the read quality scores and whether the UMI barcodes are properly formatted."

MCP Servers Used: mcp-FGbio, mcp-spatialtools
Expected Actions:

- Validate FASTQ file integrity
- Extract quality metrics (Q30 scores, read length distribution)
- Check UMI barcode format and spatial barcode structure
- Provide quality assessment report

Example 1.2: Quality Filtering with Custom Thresholds

> "I need to filter my spatial transcriptomics data from sample_breast_cancer_01 to remove low-quality reads. Please use a minimum Q-score of 30 and filter out any spatial barcodes that have fewer than 100 UMIs. Also, extract the UMIs and show me the duplication rate before and after filtering."

MCP Servers Used: mcp-FGbio, mcp-spatialtools
Expected Actions:

- Apply quality filtering (Q30 threshold)

- Filter by UMI count per barcode
- Calculate UMI duplication rates
- Generate filtered FASTQ files
- Provide QC statistics comparison

Stage 2: Spatial Segmentation & Histology Integration

Example 2.1: Histology-Guided Tissue Segmentation

"I have H&E stained images for my spatial sample (tissue_section_A_HE.tif). Can you register this histology image to my spatial barcode coordinates, segment the tissue into distinct morphological regions (tumor, stroma, necrotic areas), and then split my spatial transcriptomics data by these regions? I want to analyze each region separately."

MCP Servers Used: mcp-openImageData, mcp-spatialtools, mcp-deepcell
Expected Actions:

- Load and process H&E image
- Register image to spatial coordinates
- Perform tissue segmentation
- Split spatial data by morphological regions
- Create region-specific data subsets

Example 2.2: Region of Interest Analysis

"Looking at my spatial transcriptomics sample, I notice some interesting expression patterns in the upper-left quadrant of the tissue. Can you help me define a region of interest (coordinates approximately x: 1000-3000, y: 500-2500) and extract all spatial barcodes from this region? Also, show me the histology image overlay for this area and tell me what tissue structures are present."

MCP Servers Used: mcp-spatialtools, mcp-openImageData
Expected Actions:

- Define spatial ROI based on coordinates
- Extract barcodes within ROI
- Retrieve corresponding histology image section
- Overlay spatial data on image
- Provide tissue structure description

Stage 3: Sequence Alignment & Reference Mapping

Example 3.1: Standard Alignment with Quality Metrics

"I need to align my filtered spatial transcriptomics reads to the human reference genome (hg38). Please fetch the reference genome, align my reads using STAR with 8 threads, and provide comprehensive alignment statistics. I want to know the uniquely mapped rate, multi-mapping rate, and how many reads were unmapped. If the alignment rate is below 80%, suggest potential causes."

MCP Servers Used: mcp-FGbio, mcp-spatialtools, mcp-seqera
Expected Actions:

- Fetch hg38 reference genome
- Build/use STAR index
- Perform alignment
- Calculate alignment metrics
- Diagnose low alignment rates if applicable

Example 3.2: Nextflow Pipeline for Batch Processing

"I have 12 spatial transcriptomics samples that all need to be processed through the same alignment pipeline. Can you use Seqera Platform to launch an nf-core/rnaseq workflow for all samples simultaneously? Use the hg38 reference, STAR aligner, and generate MultiQC reports. Set up the workflow to run on AWS with 16-core instances. Monitor the progress and alert me if any samples fail."

MCP Servers Used: mcp-seqera, mcp-FGbio, mcp-spatialtools
Expected Actions:

- Configure nf-core/rnaseq pipeline
- Set up batch processing for 12 samples
- Deploy to AWS compute environment
- Monitor workflow execution
- Generate consolidated QC reports

Stage 4: Expression Quantification & Cell Segmentation

Example 4.1: UMI-Based Gene Expression Matrix

"Now that I have aligned BAM files, I need to generate a gene expression matrix. Please count unique UMIs per gene per spatial barcode, apply UMI deduplication, and create a genes × spots matrix. Also, perform basic filtering to remove spots with fewer than 200 detected genes and genes detected in fewer than 3 spots. Show me summary statistics about the resulting matrix dimensions and sparsity."

MCP Servers Used: mcp-spatialtools, mcp-FGbio
Expected Actions:

- Extract UMIs from aligned reads
- Count UMIs per gene per barcode
- Deduplicate UMIs
- Generate expression matrix
- Apply quality filters
- Provide matrix statistics

Example 4.2: Deep Learning Cell Segmentation with Expression

"I want to perform single-cell resolution analysis on my spatial data. Can you use DeepCell to segment individual cells from my H&E images, assign spatial barcodes to each segmented cell, and then aggregate the expression data at the single-cell level instead of spots? I need a cells × genes expression matrix where each row represents one segmented cell with its transcriptomic profile."

MCP Servers Used: mcp-deepcell, mcp-openImageData, mcp-spatialtools
Expected Actions:

- Load histology images
- Run deep learning cell segmentation
- Map spatial barcodes to segmented cells
- Aggregate expression by cell
- Generate single-cell expression matrix

Stage 5: Analysis & Integration

Example 5.1: Cell Type Identification with Foundation Models

> "I have a spatial expression matrix from a brain tissue sample. Can you use Hugging Face's Geneformer model to predict cell types for each spot based on the gene expression profiles? I expect to see neurons, astrocytes, microglia, and oligodendrocytes. After prediction, create a spatial map showing where each cell type is located in the tissue and calculate the proportion of each cell type."

MCP Servers Used: mcp-huggingFace, mcp-spatialtools, mcp-openImageData
Expected Actions:

- Load Geneformer model from Hugging Face
- Generate embeddings for expression profiles
- Predict cell types
- Map predictions to spatial coordinates
- Visualize spatial distribution
- Calculate cell type proportions

Example 5.2: Comparative Analysis with TCGA and Clinical Integration

> "My spatial transcriptomics sample is from a breast cancer patient (invasive ductal carcinoma). Can you compare the gene expression profile from the tumor regions of my sample against TCGA breast cancer cohorts? Identify genes that are differentially expressed compared to the TCGA average. Then, link this sample to the mock patient record in Epic (patient ID: MOCK_12345) to see if there are any clinical features (age, stage, treatment history) that might explain the expression patterns. Finally, run a pathway enrichment analysis on the differentially expressed genes."

MCP Servers Used: mcp-tcga, mcp-mockEpic, mcp-spatialtools, mcp-seqera
Expected Actions:

- Query TCGA breast cancer expression data
- Perform differential expression analysis
- Retrieve mock patient clinical data
- Correlate expression with clinical features
- Run pathway enrichment (via Nextflow pipeline)
- Generate integrated report

Cross-Stage Workflows

Example 6.1: End-to-End Automated Pipeline

> "I have raw spatial transcriptomics data for a lung cancer sample. Can you execute the complete analysis pipeline from start to finish? Starting with FASTQ validation, quality filtering, alignment to hg38, UMI counting, cell segmentation, cell type prediction using DNABERT-2, comparison to TCGA lung cancer cohorts, and finally generate a comprehensive report with spatial visualizations. Use Seqera Platform to orchestrate the workflow and optimize for speed."

MCP Servers Used: All 8 servers
Expected Actions:

- Orchestrate multi-stage workflow
- Execute each pipeline stage sequentially
- Handle data flow between stages
- Generate comprehensive analysis report
- Provide spatial visualizations

Example 6.2: Troubleshooting Failed Analysis

> "My spatial transcriptomics alignment failed with only 45% of reads mapping to the reference genome. Can you help me diagnose the problem? Check if the reference genome version matches my sample species, validate the FASTQ files for corruption, check if there's high rRNA contamination, and look at the read length distribution. If you find issues, suggest corrections and re-run the alignment with optimized parameters."

MCP Servers Used: mcp-FGbio, mcp-spatialtools, mcp-seqera
Expected Actions:

- Diagnose low alignment rate
- Check reference genome compatibility
- Validate FASTQ integrity
- Analyze read composition
- Suggest parameter adjustments
- Re-run alignment with fixes

Advanced Analysis Scenarios

Example 7.1: Spatial Gene Expression Patterns

> "I want to identify spatially variable genes in my tissue sample. Can you run a spatial autocorrelation analysis (Moran's I or Geary's C) on the top 2000 most variable genes to find which genes show significant spatial patterns? Then, cluster the spots based on these spatially variable genes and visualize the spatial domains. Also, use DNABERT to generate sequence embeddings for the spatially variable genes to see if there are any sequence motifs associated with spatial patterning."

MCP Servers Used: mcp-spatialtools, mcp-huggingFace, mcp-seqera
Expected Actions:

- Calculate spatial autocorrelation statistics
- Identify spatially variable genes
- Perform spatial clustering
- Generate spatial domain maps
- Create sequence embeddings
- Identify sequence motifs

Example 7.2: Multi-Sample Integration with Clinical Outcomes

> "I have spatial transcriptomics data from 8 different patients with varying treatment responses (4 responders, 4 non-responders). Can you integrate all samples, perform batch correction, identify differentially expressed genes between responders and non-responders, and correlate with clinical data from mock Epic records? Use TCGA data as a reference to contextualize the findings. Launch this as a Nextflow pipeline on Seqera Platform and generate a clinical report highlighting biomarkers associated with treatment response."

MCP Servers Used: All 8 servers (multi-sample workflow)
Expected Actions:

- Load multiple spatial samples
- Perform batch correction/integration
- Differential expression between groups
- Query clinical data for all patients

- Compare against TCGA cohorts
- Run workflow via Seqera
- Generate clinical biomarker report

Exploratory & Interactive Analysis

Example 8.1: Interactive Gene Expression Exploration

"I'm interested in the expression of the gene 'EPCAM' across my spatial sample. Can you show me: (1) the spatial distribution of EPCAM expression overlaid on the H&E image, (2) which cell types express EPCAM most highly according to the Geneformer predictions, (3) how EPCAM expression in my sample compares to TCGA data, and (4) what other genes are co-expressed with EPCAM in the same spatial regions?"

MCP Servers Used: mcp-spatialtools, mcp-openImageData, mcp-huggingFace, mcp-tcga
Expected Actions:

- Extract EPCAM expression values
- Create spatial heatmap
- Overlay on histology image
- Correlate with cell type predictions
- Compare to TCGA expression
- Find co-expressed genes

Example 8.2: Hypothesis Generation from Spatial Patterns

"Looking at my spatial transcriptomics data from a tumor sample, I notice interesting expression patterns at the tumor-stroma boundary. Can you help me generate hypotheses about what biological processes might be occurring here? Specifically: (1) identify genes highly expressed at the boundary, (2) use pathway enrichment to find what processes they're involved in, (3) check TCGA data to see if these boundary genes are prognostic markers, and (4) use the Hugging Face models to predict potential cell-cell interactions at this interface."

MCP Servers Used: mcp-spatialtools, mcp-seqera, mcp-tcga, mcp-huggingFace
Expected Actions:

- Identify boundary regions

- Extract boundary-enriched genes
- Run pathway analysis
- Query TCGA survival data
- Predict cell-cell interactions
- Generate biological hypotheses

Quality Control & Validation

Example 9.1: Comprehensive QC Report

> "Before I proceed with downstream analysis, I want a comprehensive quality control report for my spatial transcriptomics dataset. Can you generate metrics covering: FASTQ quality scores, alignment rates, UMI duplication levels, number of genes detected per spot, spatial distribution of read counts, tissue coverage, and comparison of my QC metrics to typical values from published datasets? Flag any concerning metrics and suggest whether I should re-process the sample."

MCP Servers Used: mcp-FGbio, mcp-spatialtools, mcp-tcga
Expected Actions:

- Aggregate QC metrics from all stages
- Calculate comprehensive statistics
- Compare to reference benchmarks
- Generate QC visualizations
- Provide recommendations

Example 9.2: Reproducibility Check

> "I want to ensure my analysis is reproducible. Can you document the complete workflow that was executed, including: all MCP servers used, tool versions, parameters passed to each tool, reference genome version, and computational environment details? Export this as a workflow specification that could be re-run via Seqera Platform. Also, generate checksums for all intermediate and final output files."

MCP Servers Used: mcp-seqera, all servers (logging)
Expected Actions:

- Document complete workflow

- Record all parameters and versions
- Generate workflow specification
- Create reproducibility report
- Calculate file checksums

Summary Table: Prompts by Pipeline Stage

Stage	Focus Area	Example Count	Primary Servers
Stage 1	QC & Ingestion	2	mcp-FGbio, mcp-spatialtools
Stage 2	Segmentation	2	mcp-openImageData, mcp-spatialtools, mcp-deepcell
Stage 3	Alignment	2	mcp-FGbio, mcp-spatialtools, mcp-seqera
Stage 4	Quantification	2	mcp-spatialtools, mcp-deepcell, mcp-FGbio
Stage 5	Analysis	2	mcp-huggingFace, mcp-tcga, mcp-mockEpic, mcp-seqera
Cross-Stage	Workflows	2	All servers
Advanced	Complex Analysis	2	Multiple servers
Exploratory	Investigation	2	Multiple servers
QC & Validation	Quality Assurance	2	All servers
TOTAL		**18 Examples**	

Prompt Design Best Practices

✅ Good Prompt Characteristics:

1. **Clear objective** - State what you want to accomplish
2. **Specific parameters** - Mention file names, thresholds, reference versions
3. **Expected output** - Describe what results you need
4. **Context** - Provide sample type, organism, experimental details
5. **Decision criteria** - Explain how to interpret results

❌ Avoid:

- Vague requests without parameters
- Assuming default values without stating them
- Omitting critical context (organism, sample type)
- Multiple unrelated questions in one prompt

Integration Patterns

Pattern 1: Sequential Pipeline

Prompt → Server 1 → Result → Prompt → Server 2 → Result → ...

Each stage builds on previous results with Claude maintaining context.

Pattern 2: Parallel Execution

Prompt → [Server 1, Server 2, Server 3] → Results merged → Final output

Multiple servers work simultaneously, Claude coordinates.

Pattern 3: Iterative Refinement

Prompt → Server → Result → Evaluate → Adjust parameters → Re-run

Claude helps optimize parameters based on intermediate results.

Appendix: Quick Reference

Common Parameter Ranges

Parameter	Typical Range	Purpose
Q-score threshold	20-30	Quality filtering
Min UMIs per spot	100-1000	Spot filtering
Min genes per spot	200-500	Spot filtering
Min spots per gene	3-10	Gene filtering
STAR threads	4-32	Alignment speed
Cell segmentation threshold	0.5-0.9	Sensitivity/specificity

File Format Reference

Format	Purpose	Example
FASTQ	Raw sequencing reads	sample_R1.fastq.gz
BAM	Aligned reads	sample_aligned.bam
H5AD	Expression matrix (AnnData)	sample_matrix.h5ad
TIFF	Histology images	tissue_section_HE.tif
MAF	Mutation data	mutations.maf

Document Version: 1.0
Last Updated: October 24, 2025
Feedback: Please report issues or suggest additional example prompts