**Coding Description for Bioinformatics Role: TRUST4 Pipeline Implementation**

### **Objective**: Implement the TRUST4 (TCR and BCR Repertoire Utilities for Solid Tissue) pipeline to analyze immune repertoires in cancer samples. This exercise aims to assess coding skills, problem-solving abilities, and understanding of high level immunogenomics concepts.

Task Overview:

1. Implement the TRUST4 pipeline (<https://github.com/liulab-dfci/TRUST4>)
2. Identify and select 2-3 cancer samples with available immunophenotyping data (hot or cold tumor)
3. Run the pipeline on the selected samples
4. Report your code and high-level results

Detailed Instructions:

1. TRUST4 Pipeline Implementation:
   * Familiarize yourself with the TRUST4 tool and its requirements
   * Set up the necessary environment and dependencies (e.g., using conda or docker)
   * Implement a workflow to run TRUST4 on multiple samples
2. Sample Selection:
   * Choose 2-3 cancer samples with available RNA-seq data and known immunophenotyping status (hot or cold tumor)
   * Samples can be obtained from public repositories like The Cancer Genome Atlas (TCGA) or Gene Expression Omnibus (GEO)
3. Pipeline Execution:
   * Run the TRUST4 pipeline on your selected samples
   * Ensure proper input formatting and parameter settings
   * Document any modifications or additional steps you implement
4. Deliverables:  
    Code:
   * Provide well-commented, readable code for your implementation
   * Include any scripts for data preprocessing, pipeline execution, and result analysis
   * Use version control (e.g., git) and share a link to your repository
5. Challenges and Solutions:
   * Describe any specific challenges encountered during the implementation
   * Explain how you addressed these challenges
6. Results Summary:
   * Provide a very high-level explanation of the output
   * Discuss whether the results are consistent across samples and with expected outcomes
   * Include basic visualizations if applicable (e.g., clonotype diversity, V/J gene usage)
7. Quality Control:
   * Explain how you assess the quality of the results and process
   * Describe any QC metrics or checks you implemented
   * Discuss how you would validate the reliability of the output

Additional Guidelines:

* Use best practices for reproducible research (e.g., environment management, documentation)
* You may use additional tools or packages as needed, but justify their use

Relevant Packages and Resources:

* TRUST4: <https://github.com/liulab-dfci/TRUST4>
* Bioconductor (for potential downstream analysis): <https://www.bioconductor.org/>
* Pandas (for data manipulation): <https://pandas.pydata.org/>
* Matplotlib or Seaborn (for visualizations): <https://matplotlib.org/> or <https://seaborn.pydata.org/>
* Snakemake or Nextflow (for workflow management, if desired): <https://snakemake.readthedocs.io/> or <https://www.nextflow.io/>

Submission:  
Please submit your code, documentation, and results within [specify timeframe, e.g., 5 days]. Feel free to reach out if you have any questions or need clarification on any aspect of the assignment.