bridge INFORMATICS

Marengo

Single Cell RNA Sequencing Analysis in R Workshop: Charting the Path from Raw Sequencing Data to Cell Clusters

Haider Hassan*[1], Shahrzad Ghazisaeidi*[1], Jennifer Martinez[1], Jessica Corrado[1], Jonathan Hsu[2], Dan Ryder[1,3]

*These authors contributed equally;

[1]Bridge Informatics, Inc., 160 Alewife Brook Pkwy, Cambridge, MA 02138; [2]Marengo Therapeutics, Inc., 840 Memorial Dr 4th Floor, Cambridge, MA 02139; [3]Corresponding author and Chief Executive Officer (CEO) of Bridge Informatics, Inc.

Single Cell RNA-Sequencing

- Single-cell RNA sequencing (scRNA-seq) is a technique that allows the profiling of a population of cells at a single cell level.
- scRNA-seq has become integral to precision medicine by providing insights into cellular heterogeneity and transcriptional dynamics within complex biological systems, aiding in the understanding of physiology and pathophysiology.
- The transcriptional profiling of cells also allows the discovery of novel cell types.

Workflow for Sample Preparation for Single Cell Sequencing

- A key aspect of scRNA-seq analysis is to isolate single cells or nuclei from a population of cells[3-6].
- The use of microscopic droplets, in a microfluidic chip, is currently the standard method to isolate single cells for sequencing.
- Libraries are prepared from single-cell suspensions, in which each cell is barcoded for identification.
- Cellular barcodes (short nucleot[...]) are attached to each read that are u[...] thereby, allowing the tracing an[...] cells.
- Unique molecular identifiers (UM[...]) barcodes that are attached to tr[...] amplification, which are then use[...] polymerase chain reaction (PCR[...]) introduced during library prepar[...]
- The 10x Chromium is a famous p[...] cell RNA sequencing that suppor[...] low-depth experiments, and UMI[...] expression level estimates.

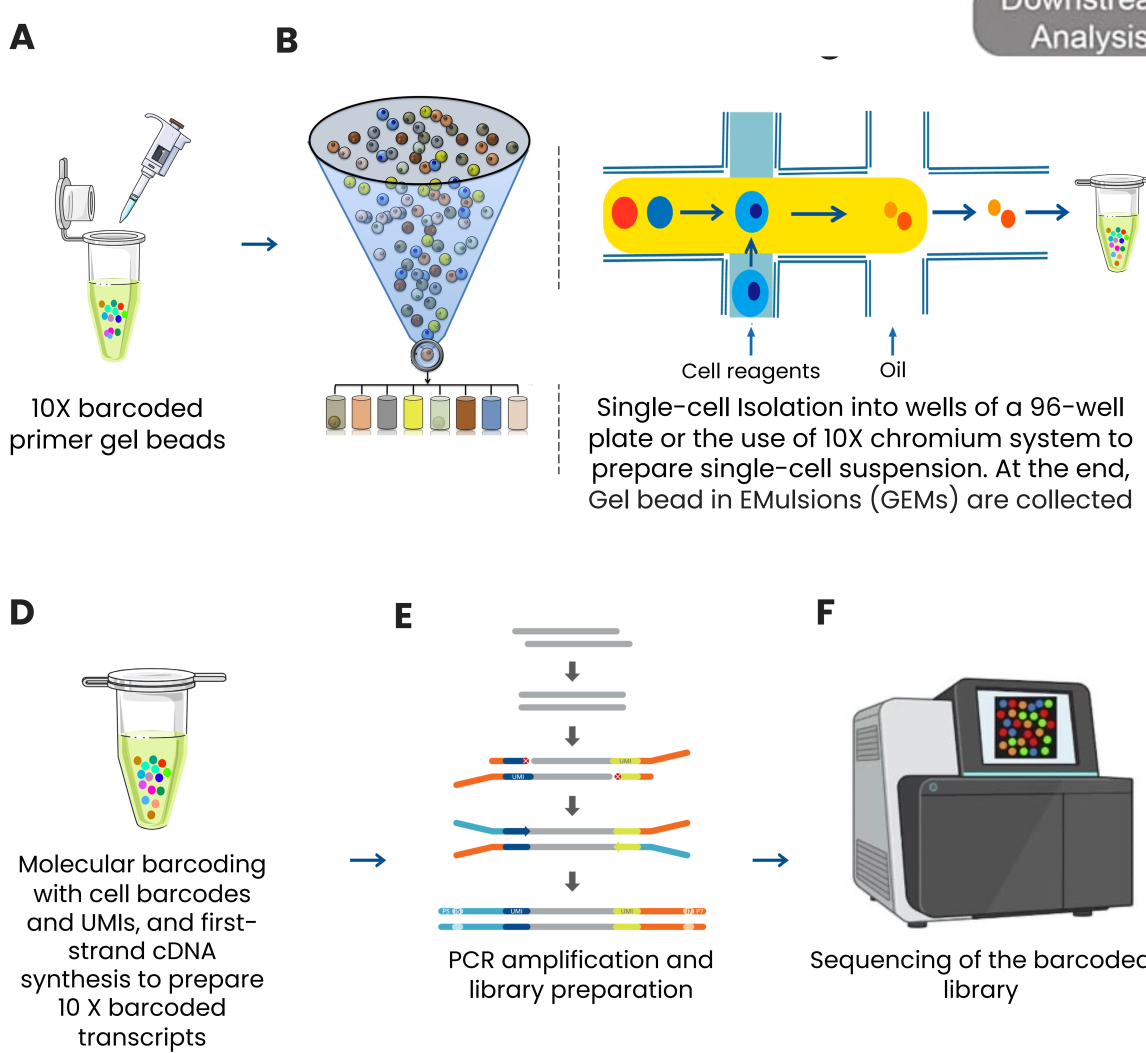

Figure 1. Preparation of samples for single-cell sequencing.
(**A**) Gel beads containing millions of cellular barcodes are generated using the Gel Bead in Emulsion (GEM) technology, with single cell GEM generation occurring in an 8-channel microfluidic chip. Each gel bead is functionalized with barcoded oligonucleotides, comprising sequencing adapters, primers, a 14 bp barcode selected from approximately 750,000 designed sequences to index GEMs, a 10 bp randomer unique molecular identifiers (UMIs), and an anchored 30 bp oligo-dT for priming polyadenylated RNA transcripts for first-strand cDNA synthesis. (**B**) Single cell suspensions are obtained either via serial dilution or through a microfluidic chip, where cells are combined with reagents and gel beads to form GEMs using the 10X Genomics platform. (**C**, **D**) Reverse transcription takes place within GEMs after which cDNAs are pooled. (**E**) Barcoded products are pooled for downstream reactions to create short-read sequencer-compatible libraries, with gel beads containing barcoded oligonucleotides facilitating reverse transcription of polyadenylated RNAs. (**F**) The process concludes with complete deep sequencing of scRNA-Seq libraries, enabling the pooling and sequencing of multiple libraries on a next-generation short-read sequencer like Illumina. Subsequently, analysis pipelines utilize the barcode information to map reads back to their original cells. This picture was adapted from Luigi F. *et al.* (2019)[1] and Shuo *et al.* (2023)[2].

The Six Pillars Of scRNA-seq Computational Workflow

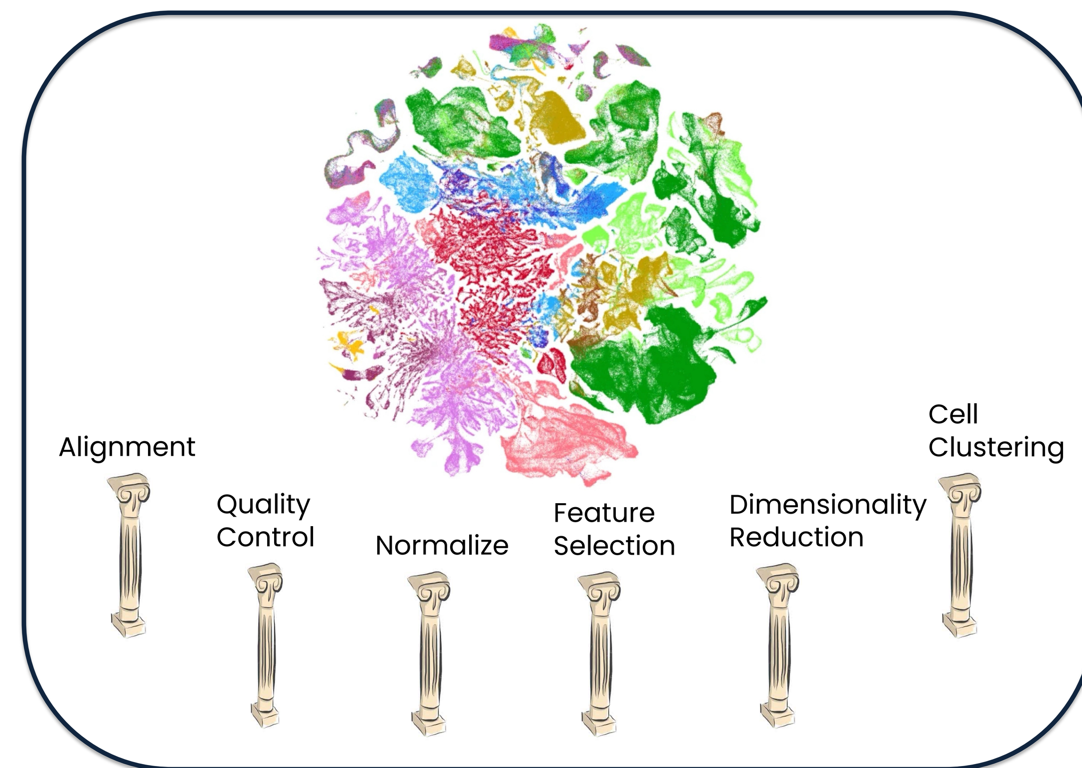

Figure 2. The six pillars of scRNA-seq computational workflow.
Single cell RNA-seq computational analysis consists of six fundamental steps: alignment, quality control, normalization, feature selection, dimensionality reduction, and cell clustering[2]. This image was adapted from Zhizhen *et al.* (2023)[8].

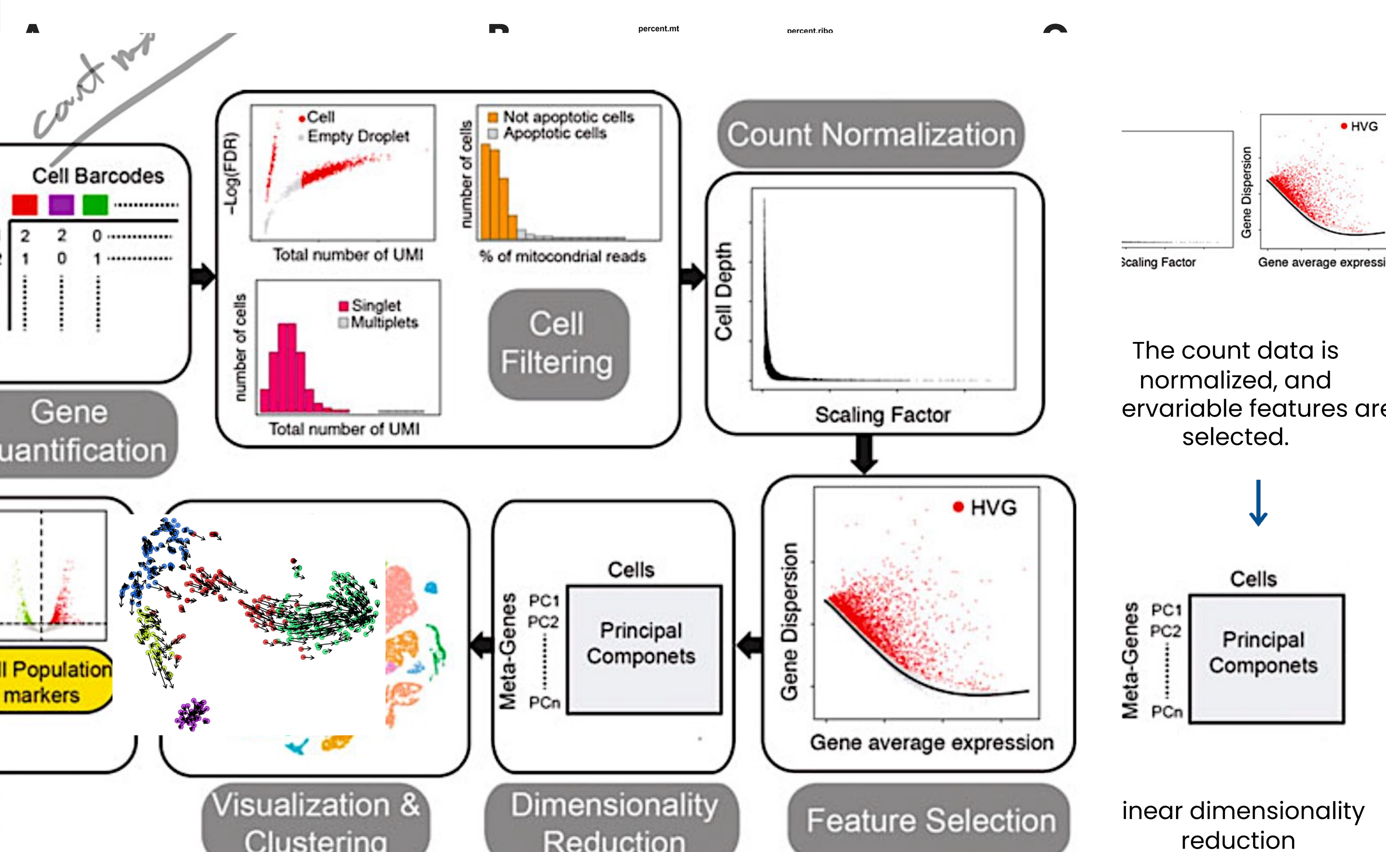

Figure 3. Single Cell RNA sequencing computational pipeline[2,3,5].
(**A**) Raw reads are processed and quantified in the alignment process to generate gene/barcode matrices. (**B**) After alignment, count matrices undergo quality control to filter dead cells, empty droplets, and multiplets. (**C**) Normalization adjusts read counts to account for variability and low signal-to-noise ratio, reducing confounding factors in downstream analysis. Feature selection identifies highly variable features with strong biological signals relative to technical noise, simplifying downstream analysis. (**D**) Linear Dimensional reduction uses principal component analysis on highly variable genes to lower data dimensionality while preserving key properties. (**E**) Non-linear dimensional reduction and clustering utilize a subset of significant principal components to overcome data noisiness. Cells are clustered and visualized based on their PCA scores, using k-means clustering and the Louvain algorithm for community detection to organize cells into groups or clusters. (**F**) Following clustering and annotation, advanced analyses such as RNA velocity and trajectory analysis can be performed. This picture was adapted from Shuo *et al.* (2023)[2], Shaked *et al.* (2021)[3], and Gioele et al. (2018)[4].

Computation Pipelines for scRNA-seq Analysis

- Numerous statistical and computational methods have been developed for scRNA-seq analysis[3,5]. The following are the most widely adapted workflows:
  - **Seurat:** A range of tools for quality control, dimensionality reduction, clustering, and visualization of scRNA-seq datasets.
  - **Monocle:** This library focuses on the inference of dynamic processes such as cell differentiation, trajectory analysis, and pseudo-temporal ordering.
  - **Scanpy:** A scRNA-seq workflow based in Python that allows for analysing large-scale datasets up to 1 million cells and more.
  - **Gf-icf:** A scRNA-seq workflow based on data transformation model termed frequency inverse document frequency (TF-IDF).
- Seurat is the most popular code-based platform for scRNA-seq analysis[3,5,7].

Capturing the Distinct Mechanisms of Action of a Novel Therapeutic

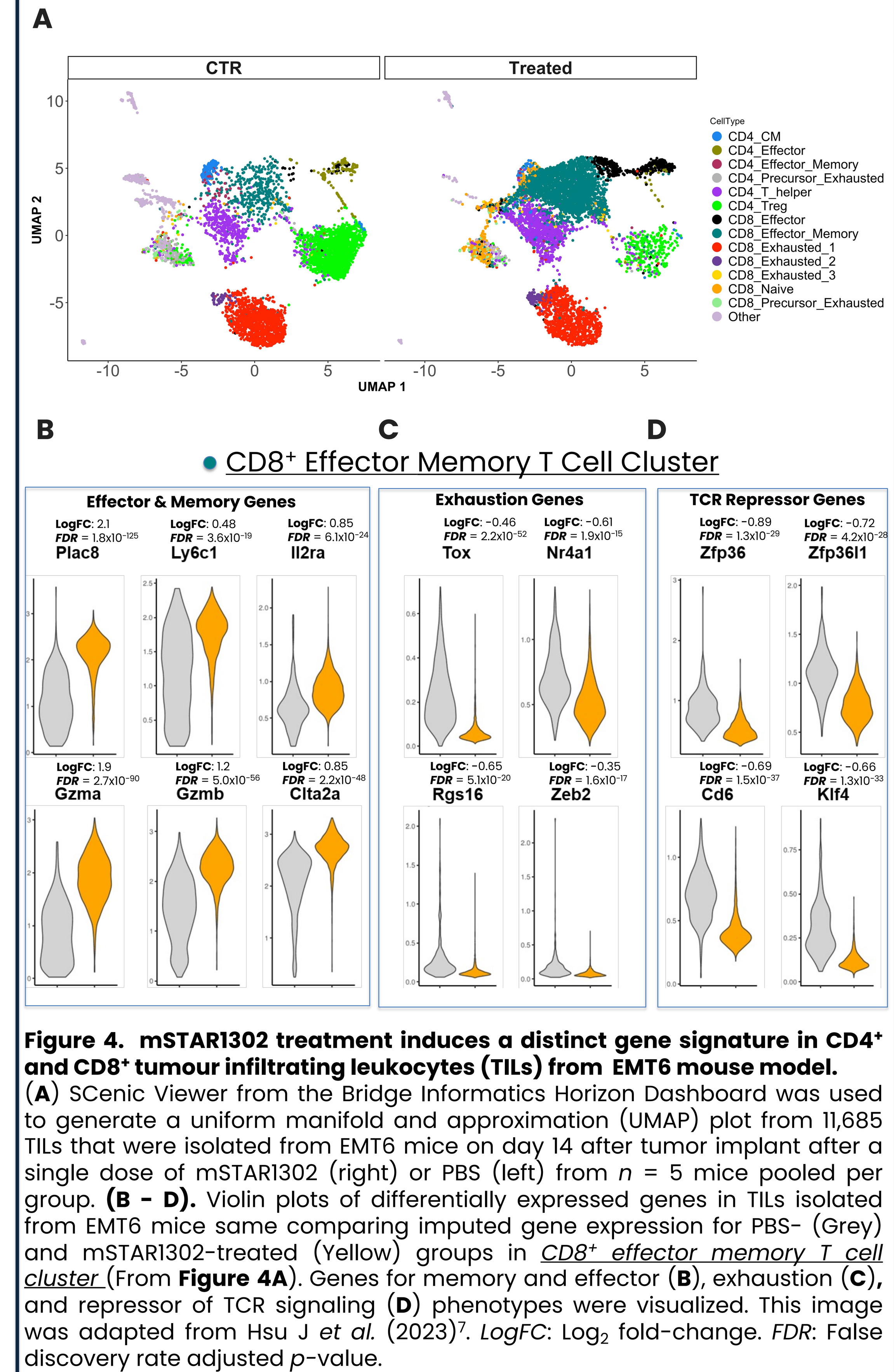

Figure 4. mSTAR1302 treatment induces a distinct gene signature in CD4+ and CD8+ tumour infiltrating leukocytes (TILs) from EMT6 mouse model.
(**A**) SCenic Viewer from the Bridge Informatics Horizon Dashboard was used to generate a uniform manifold and approximation (UMAP) plot from 11,685 TILs that were isolated from EMT6 mice on day 14 after tumor implant after a single dose of mSTAR1302 (right) or PBS (left) from $n = 5$ mice pooled per group. **(B - D).** Violin plots of differentially expressed genes in TILs isolated from EMT6 mice same comparing imputed gene expression for PBS- (Grey) and mSTAR1302-treated (Yellow) groups in *CD8+ effector memory T cell cluster* (From **Figure 4A**). Genes for memory and effector (**B**), exhaustion (**C**), and repressor of TCR signaling (**D**) phenotypes were visualized. This image was adapted from Hsu J *et al.* (2023)[7]. *LogFC*: $\log_2$ fold-change. *FDR*: False discovery rate adjusted *p*-value.

Summary

- scRNA-seq revolutionized precision medicine by comprehensively and unbiasedly capturing gene expression at single-cell resolution.
- Computational analysis of scRNA-seq requires rigorous, specialized statistical efforts in terms of quality control of the cells and annotating cell types based on gene expression.
- scRNA-seq helps investigate cell-type shifts, cell population drifts, and mechanisms of action for drug development.
- By integrating or comparing existing scRNA-seq data sets, we can gain impactful insights into the unique properties of novel therapeutics.

Let's connect!

www.bridgeinformatics.com

/company/bridge-informatics

@Bridge_Info_USA

dan.ryder@bridgeinformatics.com

GitHub: scRNA-seq analysis workshop

Acknowledgments

We want to thank Marengo Therapeutics for allowing us to incorporate their single cell analysis on mSTAR1302[7] in this poster. In particular, we want to thank Jonathan Hsu, Head of Immunology at Marengo Therapeutics, for his collaboration.

References

1. Luigi F *et al.* Cell death and Disease. 2019; 10:27
2. Shuo W *et al.* Int. J. Mol. Sci. 2023; 24(3): 2943
3. Shaked S. *et al.* RNA Bioinformatics, Methods in Molecular Biology, vol. 2284
4. Gioele L.M. *et al.* Nature. 560: 494-498
5. Andrews TS *et al.*. Nat Protoc. 2021; 16(1):1-9
6. Hwang, B., *et al* . Exp Mol Med. 2018; 50(1): 1–14
7. Hsu J. *et al.* *Science Translational Medicine* 2023; 15(724)
8. Zizhen Y. *et al.* Nature. 624: 317-332

Copyright © Property of Bridge Informatics.