 **Objective**:

* Develop a large, annotated transcriptomic atlas of nervous system tumors and non-tumor entities to enable comprehensive gene expression analyses.

 **Data Collection**:

* 5,402 neoplastic and 1,973 non-neoplastic samples collected from public sources.
* Used Applied Biosystems GeneChip for data uniformity, ensuring consistency in processing.

 **Methodology**:

* Raw data were reprocessed, normalized, and harmonized to create a cohesive dataset.
* Machine learning tools (FIt-SNE, DBSCAN, OPTICS) were used to identify clusters by diagnosis.
* Unclassified samples were reclassified using machine learning classifiers (random forest, gradient boosting, proximity classifiers).

 **Key Findings**:

* Clustering by diagnosis was achieved, with clusters primarily diagnosis-driven.
* DNA methylation’s diagnostic uniqueness extends to transcriptomic data across nervous system neoplasms.
* The dataset includes rare tumors, spans all ages, and integrates samples worldwide, supporting broad comparative analyses.

 **Applications**:

* Enables comparative gene expression analysis among nervous system neoplasms.
* Supports diagnostic refinement, especially in cases like pilocytic astrocytoma and ganglioglioma.
* Dataset is useful for exploring biological relationships between tumors and healthy tissues.

 **Limitations**:

* Some diagnostic inconsistencies remain due to surrogate variables like geographic origin.
* High classifier accuracy may reflect potential overfitting; not yet suited for clinical application.

 **Conclusion**:

* A comprehensive, high-dimensional dataset of nervous system tumors and non-tumor tissues was created.
* The atlas supports future research and holds potential for rare disease data integration and exploration.