SIGPOLYTOPE USER MANUAL

1. WHAT IS SIGPOLYTOPE?

SigPolytope is an R package designed for the geometric representation and comparison of molecular signatures in a multidimensional latent space. Rather than treating signatures as flat feature vectors or opaque, black-box embeddings, SigPolytope models each molecular signature as a polytope, whose geometry encodes biologically meaningful information.

The geometric structure of each polytope reflects explicitly defined biological dimensions, such as correlation patterns, tumor–normal behavior, survival associations, and immune or microenvironmental features. In this framework, geometry becomes a means of interpretation rather than mere visualization.

The core idea underlying SigPolytope is straightforward: a molecular signature should be understood as a geometric object, not simply as a list of genes.

2. CORE CONCEPTS

2.1. Latent dimensions

In SigPolytope, a latent dimension is an explicitly defined, biologically interpretable variable that contributes to the geometric representation of a molecular signature. Latent dimensions are always numeric and may represent either continuous values (such as effect sizes or scores) or directional signals (such as risk versus protection).

Examples of latent dimensions include correlation strength, survival risk direction, immune phenotypes, and tumor–normal regulatory behavior. Importantly, SigPolytope does not infer or learn latent dimensions automatically. Instead, all dimensions are constructed upstream, based on prior biological analyses, ensuring full interpretability of the resulting geometry.

2.2. Signature versus regulator

Each circuitry analyzed in SigPolytope consists of two complementary entities: a signature and a regulator. The signature represents the molecular program of interest,

such as a gene set or multi-omic signature, while the regulator corresponds to the interacting or regulatory component associated with that signature.
Both entities are embedded within the same latent space and are described using identical dimensions. This shared representation allows for direct geometric comparison between the signature and its regulator, enabling quantitative assessment of convergence or divergence between the two.

2.3. Dimensionality (P)

SigPolytope does not assume a fixed number of latent dimensions. Instead, the dimensionality of the latent space is fully determined by the user and adapts automatically throughout the analysis. A space defined by five dimensions yields a five-dimensional polytope, while a space defined by eighteen dimensions produces an eighteen-dimensional polytope.

In general, a latent space with *P* dimensions results in a polytope with *2P* canonical vertices, corresponding to the positive and negative directions of each dimension. This flexibility allows SigPolytope to be applied seamlessly to both exploratory analyses with few dimensions and paper-aligned analyses with high-dimensional biological representations.

3. DATA REQUIREMENTS

To use SigPolytope, a minimal set of structured data is required. Each observation must include a unique circuitry identifier (Circuitries_id) and a collection of numeric features describing both the molecular signature and its corresponding regulator. These features define the latent dimensions used for geometric construction.

Crucially, the signature and regulator must be described using matching dimensions, meaning that both entities share the same set of latent variables. This alignment ensures that both are embedded within the same latent space and can be compared geometrically.

Conceptually, a minimal input may resemble the following structure, where each row corresponds to a single circuitry and paired columns describe the signature and regulator:

Circuitries_id | rho_sig | rho_int | OS_dir_sig | OS_dir_int

4. TYPICAL WORKFLOW

The SigPolytope workflow follows a clear sequence of steps that transform biologically derived features into geometric objects suitable for comparison and visualization.

4.1. Step 1 — Prepare latent dimensions

The first step consists of selecting which biological dimensions will define the latent space. These dimensions must be biologically meaningful and numerically encoded. For example, a five-dimensional latent space may be defined as:

```
cols <- c("rho", "TN_dir", "OS_dir", "OS_strength", "Immune_dir")
```

This selection explicitly defines the geometry of the space in which signatures and regulators will be embedded.

4.2. Step 2 — Build feature matrices

Next, two feature matrices are constructed: one for the molecular signatures and one for the regulators. Both matrices must have identical column names, corresponding to the chosen latent dimensions, while rows represent individual circuitries.

In this structure, rows correspond to circuitries and columns correspond to latent dimensions. Conceptually, the matrices can be represented as:

```
X_sig <- as.matrix(signature_features)
X_int <- as.matrix(regulator_features)
```

Ensuring identical column names between X_sig and X_int is essential for downstream geometric alignment.

4.3. Step 3 — Build geometry objects

Once the feature matrices are prepared, they are converted into geometry objects using the core SigPolytope constructors:

```
sig_geom <- build_signature_geometry(
  data = sig_df,
  cols = cols,
  id_col = "Circuitries_id"
)
reg_geom <- build_regulator_geometry(
  data = int_df,
  cols = cols,
  id_col = "Circuitries_id"
)
```

These geometry objects store the latent features, associated metadata, and the labels of each latent dimension, forming the foundation for all subsequent analyses.

4.4. Step 4 — Compute convergence

Geometric similarity between the signature and regulator is quantified using the convergence function:

```
conv <- compute_circuitry_convergence(
  sig_geom,
  reg_geom,
  n_components = 3
)
```

This step evaluates how closely the geometric representations of the signature and regulator align within the latent space.

4.5. Step 5 — Build and visualize polytopes

Finally, the geometric objects are assembled into a tensor, projected into three-dimensional space for visualization, and used to construct the polytope representation of a circuitry:

```
tensor <- list(
  features_sig = sig_geom$features,
  features_int = reg_geom$features,
  meta = sig_geom$meta
)
embedding <- compute_pca_embedding_from_tensor(tensor)
poly <- build_circuitry_polytope(tensor, embedding, index = 1)
plot_circuitry_3d(poly)
```

It is important to emphasize that principal component analysis (PCA) is used exclusively for visualization purposes. All geometric computations, including polytope construction and convergence metrics, are performed in the original P-dimensional latent space.

5. UNDERSTANDING THE GEOMETRY

5.1. Polytopes

In SigPolytope, each circuitry is represented as a polytope constructed in a *P*-dimensional latent space. For a space defined by *P* latent dimensions, the resulting polytope contains *2P* canonical vertices, corresponding to the positive and negative directions of each dimension.

These vertices encode the extremal configurations of the signature along each biological axis. The convex hull formed by these vertices captures the overall organizational footprint of the circuitry, providing a compact geometric summary of how different biological dimensions jointly structure the signature.

5.2. Barycenters

The barycenter of a polytope represents its geometric “center of mass” within the latent space. Conceptually, it reflects the overall balance among all latent dimensions, integrating directional and magnitude information into a single geometric point.

Barycenters provide a compact summary of a molecular signature or regulator, enabling straightforward quantitative comparisons. In SigPolytope, distances between the barycenters of signatures and their corresponding regulators are used to quantify geometric convergence or divergence between the two entities.

6. EXPLORATORY VERSUS PAPER-ALIGNED USAGE

SigPolytope supports two complementary modes of analysis: exploratory usage and paper-aligned usage. In exploratory mode, users typically work with a small number of latent dimensions, usually between three and six. This mode emphasizes intuitive visualization and is well suited for hypothesis generation and exploratory data analysis.

In contrast, paper-aligned usage relies on the full set of biologically defined dimensions, such as an eighteen-dimensional latent space. This mode prioritizes reproducibility and methodological rigor, making it appropriate for formal analyses reported in methods sections of scientific manuscripts.

Importantly, the same core functions and workflow are used in both modes. The distinction lies solely in the number and nature of the latent dimensions selected by the user.

7. WHAT SIGPOLYTOPE IS NOT

SigPolytope is not a clustering tool, a machine learning model, a feature learning algorithm, or a survival analysis method. It does not perform statistical inference on raw data nor does it learn representations automatically.

Instead, SigPolytope assumes that such analyses have been conducted upstream. Its purpose is to provide a geometric framework for interpreting and comparing the resulting biologically informed features.

8. COMMON PITFALLS

Several common issues may arise when using SigPolytope. These include mismatched column names between signature and regulator feature matrices, the use of non-numeric latent dimensions, and the misinterpretation of PCA axes as biologically meaningful dimensions.

Another frequent pitfall is mixing exploratory and paper-aligned interpretations within the same analysis. Clear separation of these modes is essential to ensure both interpretability and reproducibility.

9. DESIGN PHILOSOPHY

SigPolytope is built around three guiding principles. First, explicit biological definitions are favored over latent inference, ensuring that all dimensions are interpretable. Second, geometry is treated as a form of interpretation rather than mere visual decoration. Third, dimensionality is understood as a deliberate analytical choice rather than a fixed constraint imposed by the method.

10. CITATION

When using SigPolytope in scientific work, users should cite the corresponding manuscript and software release. The appropriate citation information can be obtained directly within R using:

```
citation("SigPolytope")
```