**Original datasets**

- Literature curated nucleolar association dataset ordered by PubMed ID reference
  - PubMed ID = 17470000

**Intersection**

- SILAC-derived dataset defining relative nucleolar, nucleoplasmic and cytoplasmic abundance ratios

**Training / Testing sets**

1. **Training set and test set 1**
   - 30 nucleolar-enriched
   - 22 nucleolar-nucleoplasmic
   - 24 nucleolar-cytoplasmic
   - +200 randomly selected non-nucleolar proteins

2. **Test set 2 (independent literature-based test set)**
   - 15 nucleolar-enriched
   - 7 nucleolar-nucleoplasmic
   - 16 nucleolar-cytoplasmic
   - +100 randomly selected non-nucleolar proteins

3. **Test set 3 (SILAC test set)**
   - 52 nucleolar-enriched
   - 24 nucleolar-nucleoplasmic
   - 63 nucleolar-cytoplasmic
   - 344 non-nucleolar proteins

* Dataset used to determine the abundance ratio thresholds defining the nucleolar association classes for the SILAC dataset