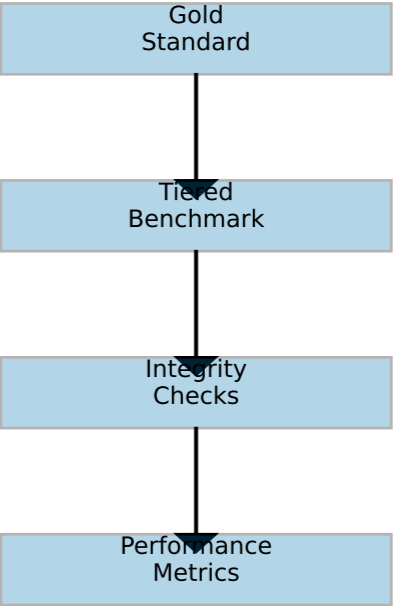# Figure 1: Task Taxonomy and Benchmark Integrity Framework

A. Task Taxonomy

```
Task A: GWAS Credible Set → Causal Gene
  Input: LD-expanded SNP set from GWAS locus
  Output: Ranked list of candidate genes
  Methods: Distance, ABC, PoPS, eQTL, Hi-C

Task B: Enhancer → Gene
  Input: Known regulatory element (CRISPRi-validated)
  Output: Target gene(s)
  Methods: ABC, Hi-C, correlation

Task C: Variant → Regulatory Element
  Input: Individual variant
  Output: Regulatory activity prediction
  Methods: ChromHMM, DeepSEA, enformer
```

B. Validation Pipeline

```
Gold
Standard
   │
   ▼
Tiered
Benchmark
   │
   ▼
Integrity
Checks
   │
   ▼
Performance
Metrics
```

C. Benchmark Integrity Checklist

- ✓ Task-appropriate
- ✓ No training leakage
- ✗ Source diversity
- ✓ Temporal ordering
- ✓ No circular validation
- ✓ Size adequacy

D. Benchmark Statistics

```
Task A (GWAS → Gene):
  Loci: 14,016
  Genes: 4,892
  Tier-0 (G2P): 1,248 pairs
  Tier-1 (ClinVar): 2,764 pairs

Task B (Enhancer → Gene):
  Enhancers: 19,825
  Genes: 3,141
  CRISPRi-validated: 15,692
  MPRA-validated: 4,133

Evidence Sources:
  G2P/OMIM: 38%
  ClinVar: 29%
  CRISPR: 21%
  MPRA: 12%
```