Information-Theoretic Analysis: Cell-Type Markers vs Treatment Response Biomarkers
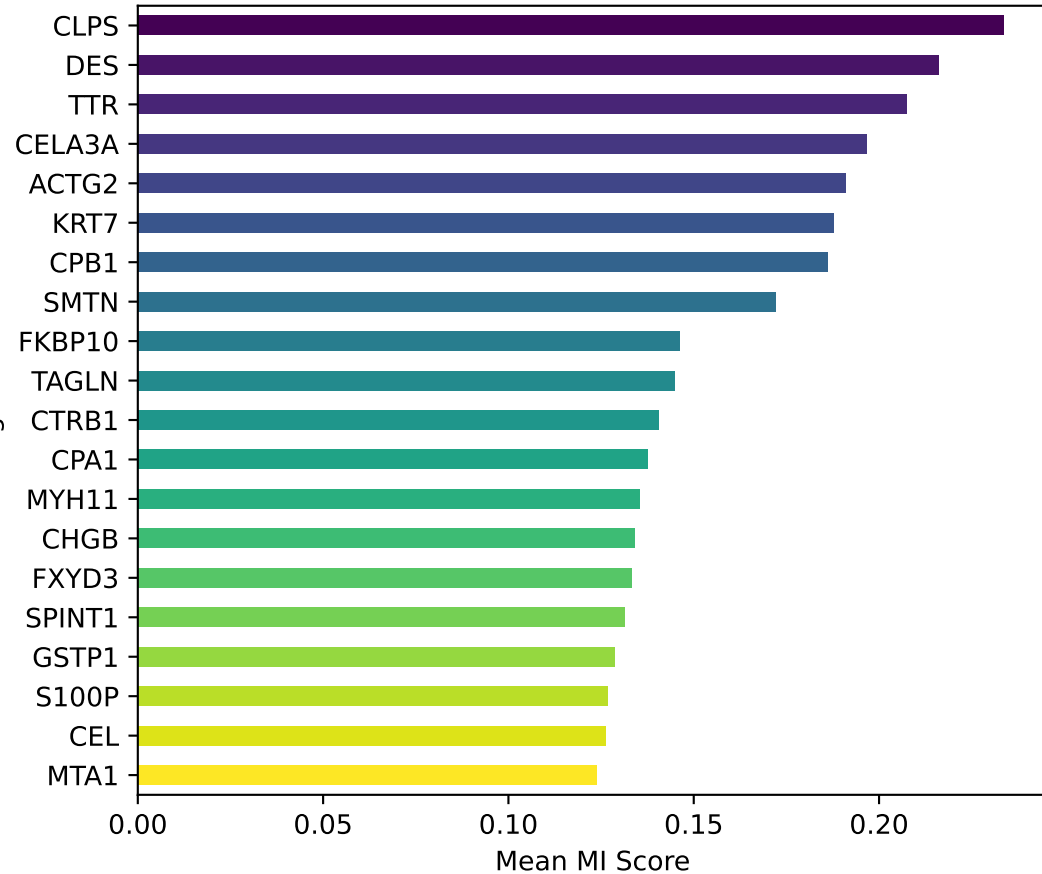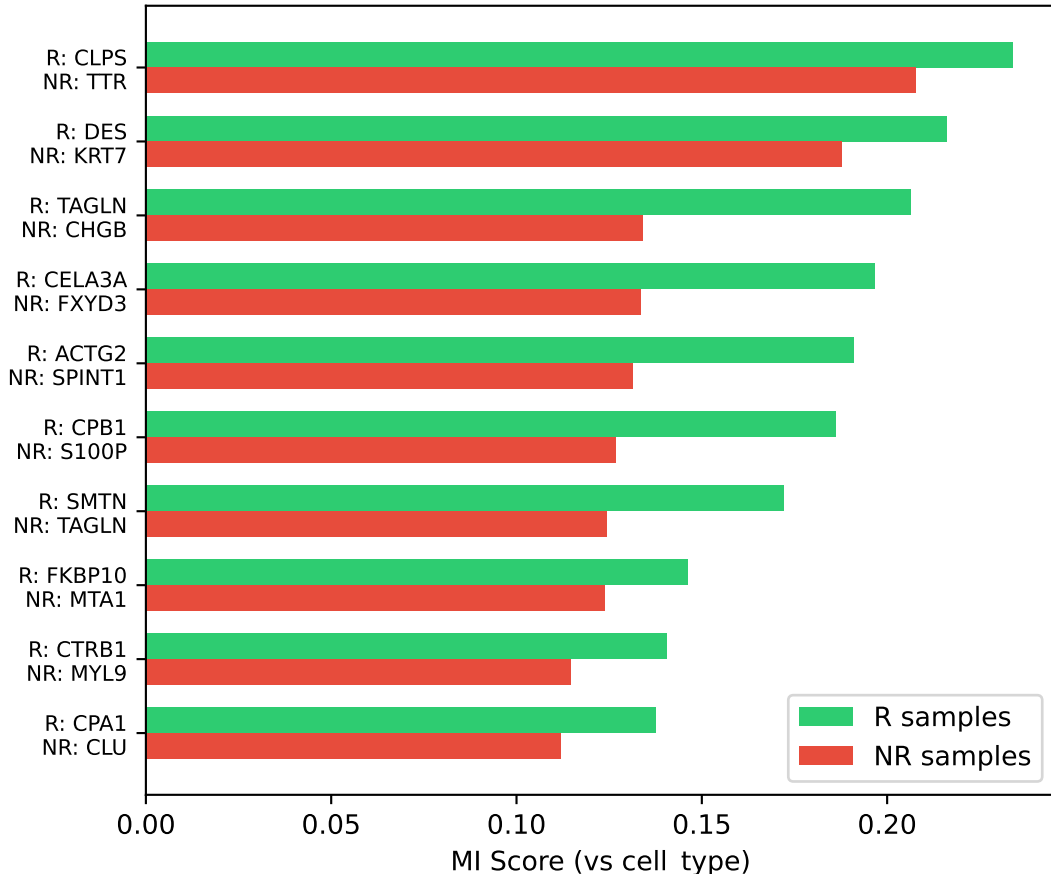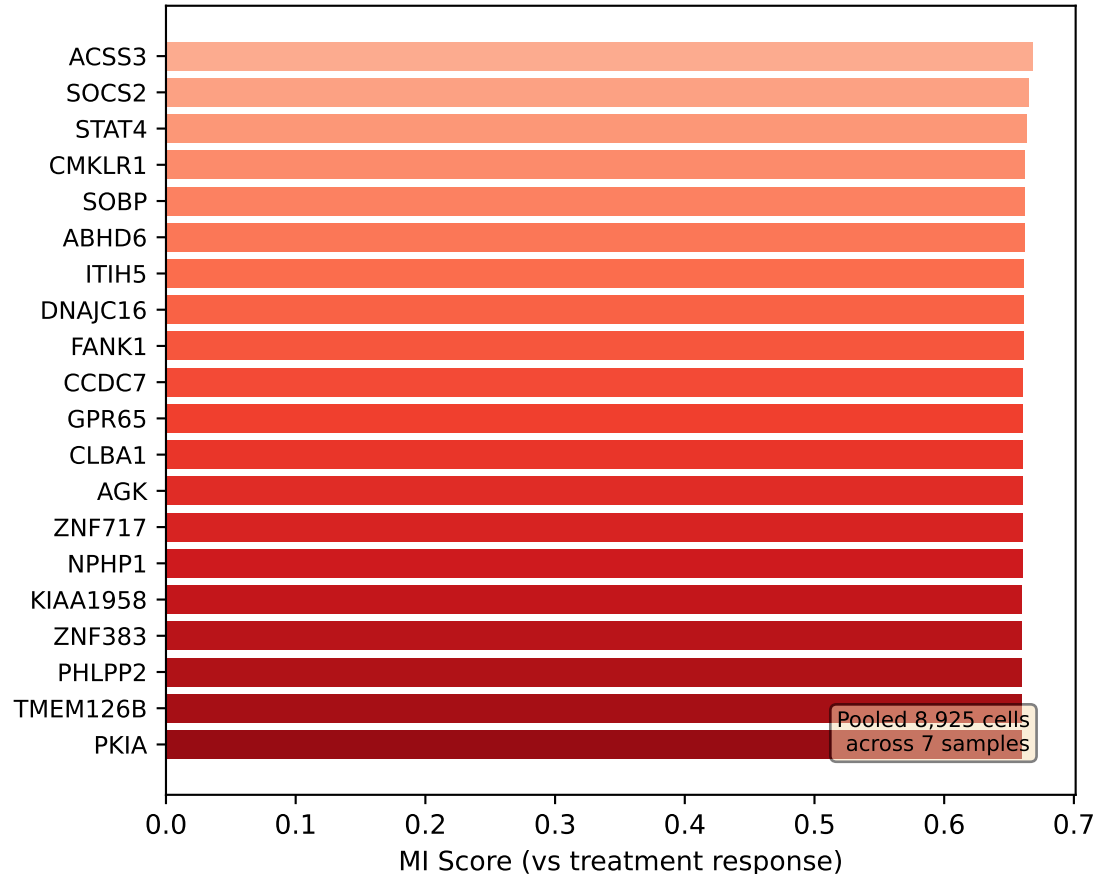
Top 20 Cell-Type Marker Genes (MI vs cell_type, averaged)

Genes (top to bottom): CLPS, DES, TTR, CELA3A, ACTG2, KRT7, CPB1, SMTN, FKBP10, TAGLN, CTRB1, CPA1, MYH11, CHGB, FXYD3, SPINT1, GSTP1, S100P, CEL, MTA1

X-axis: Mean MI Score (0.00 to ~0.23)

Cell-Type Markers Stability Across Response Groups

R sample	NR sample
R: CLPS	NR: TTR
R: DES	NR: KRT7
R: TAGLN	NR: CHGB
R: CELA3A	NR: FXYD3
R: ACTG2	NR: SPINT1
R: CPB1	NR: S100P
R: SMTN	NR: TAGLN
R: FKBP10	NR: MTA1
R: CTRB1	NR: MYL9
R: CPA1	NR: CLU

Legend: R samples (green), NR samples (red)

X-axis: MI Score (vs cell_type)

TRUE Response Biomarkers (MI vs R/NR across all cells)

Genes (top to bottom): ACSS3, SOCS2, STAT4, CMKLR1, SOBP, ABHD6, ITIH5, DNAJC16, FANK1, CCDC7, GPR65, CLBA1, AGK, ZNF717, NPHP1, KIAA1958, ZNF383, PHLPP2, TMEM126B, PKIA

X-axis: MI Score (vs treatment response) (0.0 to 0.7)

Pooled 8,925 cells across 7 samples

Understanding MI Analysis

WHAT IS MUTUAL INFORMATION?

MI measures how much knowing one variable reduces uncertainty about another.

- MI = 0: Variables are independent
- MI > 0: Variables share information
- Higher MI = Stronger association

TWO TYPES OF MI ANALYSIS:

1. MI vs CELL TYPE (Panels 1-2):
   "Which genes best distinguish cell types?"
   → Useful for marker identification
   → Applied per-sample

2. MI vs RESPONSE (Panel 3):
   "Which genes predict R vs NR?"
   → TRUE treatment biomarkers
   → Computed across all samples pooled

KEY DIFFERENCE:

Panel 1-2: Same gene can be "top marker" in both R and NR - it just identifies cell types well regardless of response.

Panel 3: These genes specifically differ between responders and non-responders.

TOP BIOMARKER CANDIDATES:
- ACSS3: MI=0.668
- SOCS2: MI=0.665
- STAT4: MI=0.664
- CMKLR1: MI=0.662
- SOBP: MI=0.662