DLBCL Prognostic Gene Signatures

Genome-Wide Survival Analysis by Cell of Origin

HMRN/Lacy Cohort Analysis

2026-01-19

Contents

1 Executive Summary

This analysis identifies **prognostic gene signatures** in Diffuse Large B-Cell Lymphoma (DLBCL) through unbiased genome-wide survival analysis across Cell of Origin (COO) subtypes.

1.1 Key Findings

- **Global Analysis (n=1,303)**: 6,748 probes significantly associated with survival (FDR < 0.05)
- **COO-Specific Signatures**: Largely non-overlapping, indicating distinct biology
- **IPI Independence**: 2,839 genes remain significant after adjusting for International Prognostic Index
- **Protective Pathways**: T-cell receptor signaling, immune activation
- **Adverse Pathways**: Cell cycle, DNA repair, E2F targets

2 Methodology

2.1 Study Cohort

Table 1: COO Distribution in HMRN/Lacy Cohort

COO Subtype	N	Percentage
ABC	345	26.5
GCB	517	39.7
MHG	164	12.6
UNC	277	21.3

Data Source: HMRN/Lacy Cohort (GSE181063)

- **Platform**: Illumina HumanHT-12 V4 BeadChip (GPL14951)
- **Probes**: 29,372 expression probes
- **Samples**: 1303 with survival data
- **Endpoint**: Overall Survival (OS)

2.2 Statistical Analysis

2.2.1 Univariate Survival Analysis

For each probe i and sample j:

$$h(t|x_{ij}) = h_0(t) \cdot \exp(\beta_i \cdot x_{ij})$$

Where:

- $h(t)$ = hazard at time t
- $h_0(t)$ = baseline hazard
- x_{ij} = standardized expression (z-score)
- β_i = log hazard ratio

Multiple Testing Correction: Benjamini-Hochberg FDR

2.2.2 IPI-Adjusted Analysis (Multivariate)

$$h(t|x, IPI) = h_0(t) \cdot \exp(\beta_{gene} \cdot x + \beta_{IPI} \cdot IPI)$$

Genes with $p_{adjusted} < 0.05$ are considered **IPI-independent**.

3 Results by COO Subset

3.1 Global Analysis (All COO)

Table 2: Global Survival Analysis Summary (n=1,303)

Metric	Value
Total probes tested	29,349
Significant ($p < 0.05$)	10,402
Significant (FDR < 0.05)	6,748
Protective (HR < 1, FDR < 0.05)	2,713
Adverse (HR > 1, FDR < 0.05)	4,035

3.1.1 Kaplan-Meier: Global Signature Score

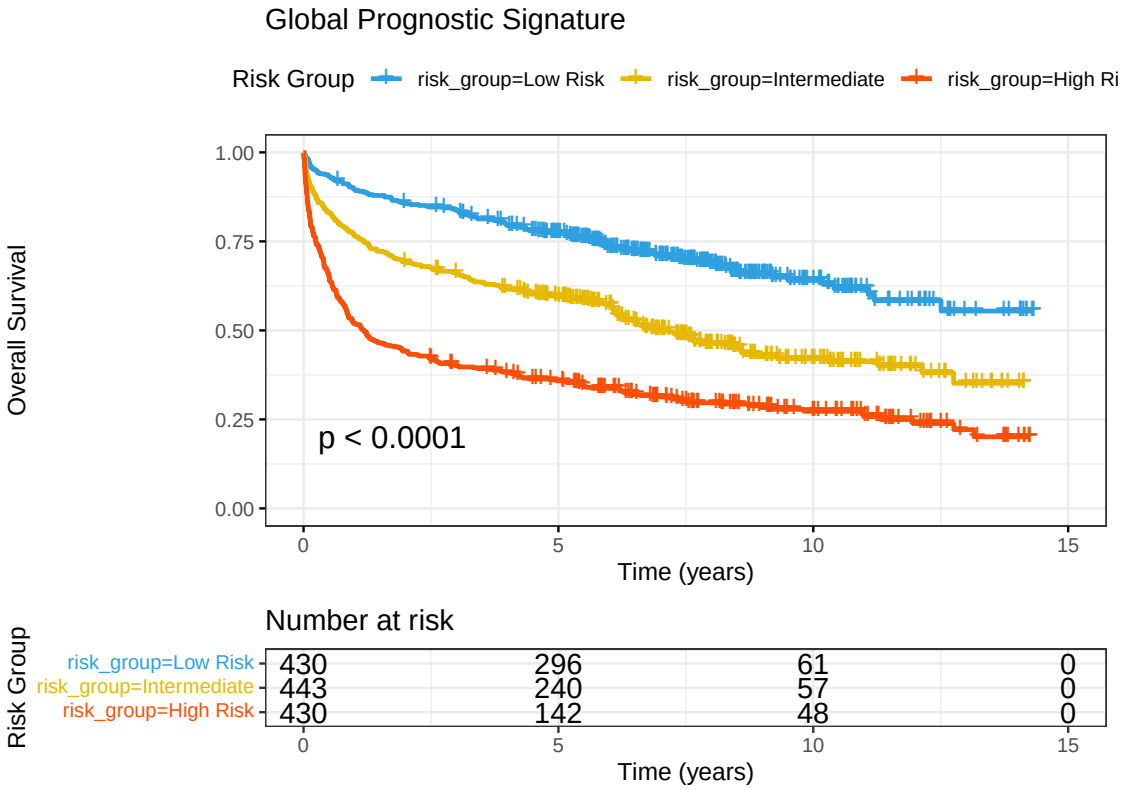

Figure 1: Survival by Global Prognostic Signature Score

3.1.2 Top Protective Genes (Global)

Table 3: Top 20 Protective Genes (HR < 1) - Global Analysis

Gene	Probe	HR	95% CI	P-value	FDR
SIRPG	ILMN_2383058	0.796	(0.758-0.836)	1.44e-19	4.23e-15
SIRPG	ILMN_1771801	0.794	(0.754-0.835)	4.19e-19	6.15e-15
MMP9	ILMN_1796316	0.761	(0.713-0.813)	3.56e-16	2.61e-12
CD3D	ILMN_2261416	0.825	(0.788-0.864)	5.45e-16	2.70e-12
CD3E	ILMN_1739794	0.776	(0.73-0.825)	5.53e-16	2.70e-12
CD1E	ILMN_2335754	0.736	(0.684-0.793)	7.83e-16	3.25e-12
BCL11B	ILMN_1665761	0.813	(0.773-0.855)	9.40e-16	3.25e-12
PRKCH	ILMN_1780898	0.701	(0.643-0.765)	9.97e-16	3.25e-12
LST1	ILMN_2345353	0.699	(0.64-0.764)	2.07e-15	5.97e-12
CD6	ILMN_1746565	0.822	(0.783-0.863)	2.40e-15	5.97e-12
LCP2	ILMN_1658962	0.693	(0.632-0.758)	2.44e-15	5.97e-12
LST1	ILMN_1688373	0.638	(0.571-0.713)	2.69e-15	6.07e-12
TCF7	ILMN_1676470	0.806	(0.763-0.85)	3.06e-15	6.40e-12
LCK	ILMN_2377109	0.767	(0.717-0.819)	3.27e-15	6.40e-12
SH2D1A	ILMN_1705892	0.761	(0.71-0.817)	2.17e-14	3.99e-11
JAML	ILMN_1778723	0.779	(0.73-0.831)	4.58e-14	7.47e-11
FNDC1	ILMN_1734653	0.829	(0.789-0.871)	1.30e-13	1.91e-10
PDLIM3	ILMN_2230025	0.791	(0.743-0.842)	1.42e-13	1.95e-10
CSF2RA	ILMN_2376455	0.777	(0.726-0.83)	1.58e-13	1.98e-10
ABLIM1	ILMN_1785424	0.770	(0.718-0.825)	1.62e-13	1.98e-10

3.1.3 Top Adverse Genes (Global)

Table 4: Top 20 Adverse Genes (HR > 1) - Global Analysis

Gene	Probe	HR	95% CI	P-value	FDR
POLR3G	ILMN_2070349	1.329	(1.242-1.421)	1.09e-16	1.07e-12
EXO1	ILMN_1719089	1.366	(1.26-1.482)	4.25e-14	7.34e-11
C19orf48	ILMN_1681124	1.287	(1.205-1.374)	5.82e-14	8.99e-11
UBE2Q2P2	ILMN_3235825	1.333	(1.235-1.439)	1.46e-13	1.95e-10
EXO1	ILMN_2351916	1.327	(1.228-1.435)	1.07e-12	1.01e-09
POLR1G	ILMN_1747870	1.421	(1.288-1.567)	2.11e-12	1.63e-09
EME1	ILMN_1750102	1.433	(1.296-1.584)	2.21e-12	1.67e-09
UBE2V2	ILMN_2076567	1.590	(1.396-1.811)	3.10e-12	2.22e-09
TMEM97	ILMN_1710962	1.525	(1.353-1.72)	5.49e-12	3.50e-09
PCLAF	ILMN_1732150	1.288	(1.198-1.384)	5.91e-12	3.69e-09
SNORA71C	ILMN_3240216	1.464	(1.313-1.633)	7.34e-12	4.34e-09
MACIR	ILMN_1677292	1.427	(1.289-1.58)	7.39e-12	4.34e-09
GTPBP3	ILMN_1777156	1.582	(1.384-1.807)	1.56e-11	7.89e-09
DNA2	ILMN_2282959	1.401	(1.269-1.547)	2.74e-11	1.30e-08
SLC19A1	ILMN_1698996	1.282	(1.192-1.38)	2.91e-11	1.34e-08
E2F7	ILMN_1798210	1.269	(1.182-1.363)	6.29e-11	2.51e-08
KIF23	ILMN_1716553	1.338	(1.225-1.462)	9.01e-11	3.33e-08

FAM72B	ILMN_3243986	1.356	(1.236-1.486)	9.09e-11	3.33e-08
EHHADH	ILMN_1701507	1.253	(1.17-1.341)	9.82e-11	3.47e-08
ECE2	ILMN_1762883	1.515	(1.335-1.72)	1.23e-10	4.10e-08

3.2 GCB Subset Analysis (n=517)

Table 5: GCB Subset Survival Analysis Summary

Metric	Value
Total probes tested	29,357
Significant (FDR < 0.05)	909
Protective	392
Adverse	517

3.2.1 Kaplan-Meier: GCB Signature

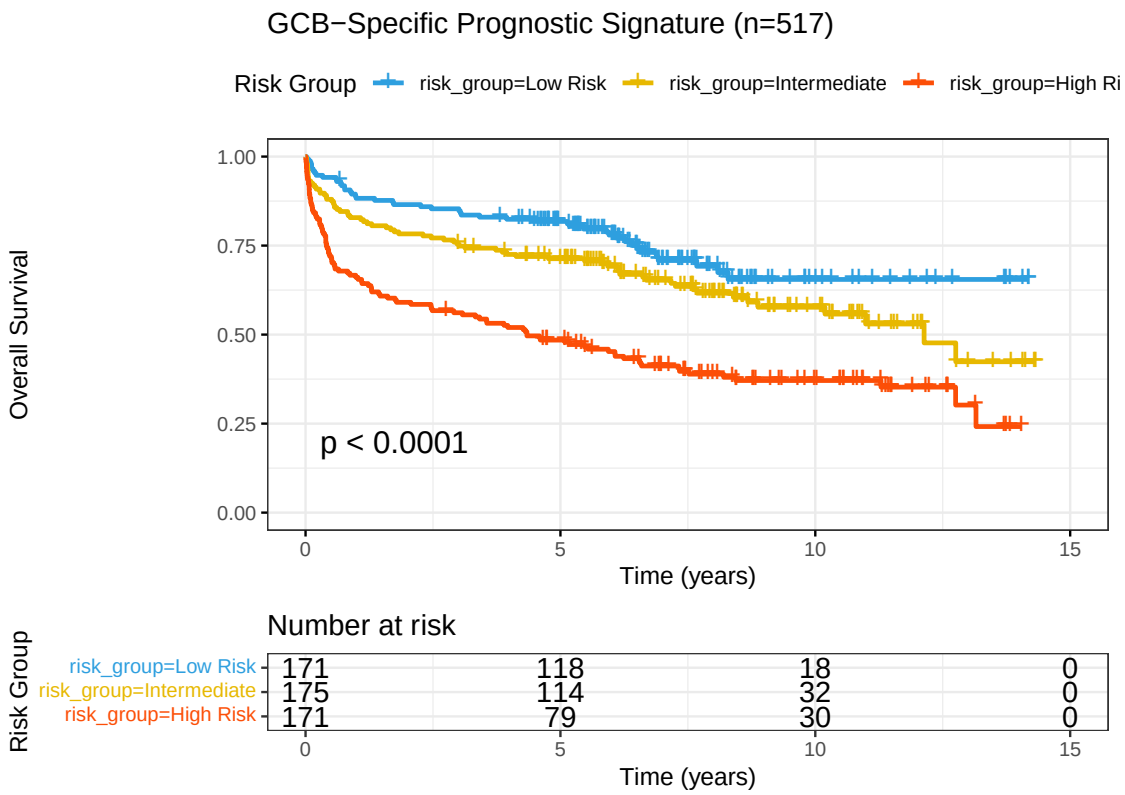

Figure 2: Survival by GCB-Specific Prognostic Signature

3.2.2 Top GCB-Specific Genes

Table 6: Top 15 GCB-Specific Prognostic Genes

Gene	HR	Direction	P-value
CD3D	0.569	Protective	6.38e-08
TIGIT	0.676	Protective	1.01e-07
QPCTL	1.353	Adverse	3.57e-07
PRND	0.808	Protective	4.40e-07
CD3E	0.734	Protective	4.74e-07
CD1E	0.718	Protective	5.13e-07
PHRF1	1.477	Adverse	5.59e-07
ZNF792	0.741	Protective	5.85e-07
FHIT	0.727	Protective	6.11e-07
PKMYT1	1.273	Adverse	9.51e-07
NME2	1.473	Adverse	9.84e-07
KIF23	1.543	Adverse	1.00e-06
SIRPG	0.789	Protective	1.39e-06
MAL	0.793	Protective	1.40e-06
CD3D	0.779	Protective	1.45e-06

3.3 ABC Subset Analysis (n=345)

Table 7: ABC Subset Survival Analysis Summary

Metric	Value
Total probes tested	29,361
Significant (FDR < 0.05)	967
Protective	374
Adverse	593

3.3.1 Kaplan-Meier: ABC Signature

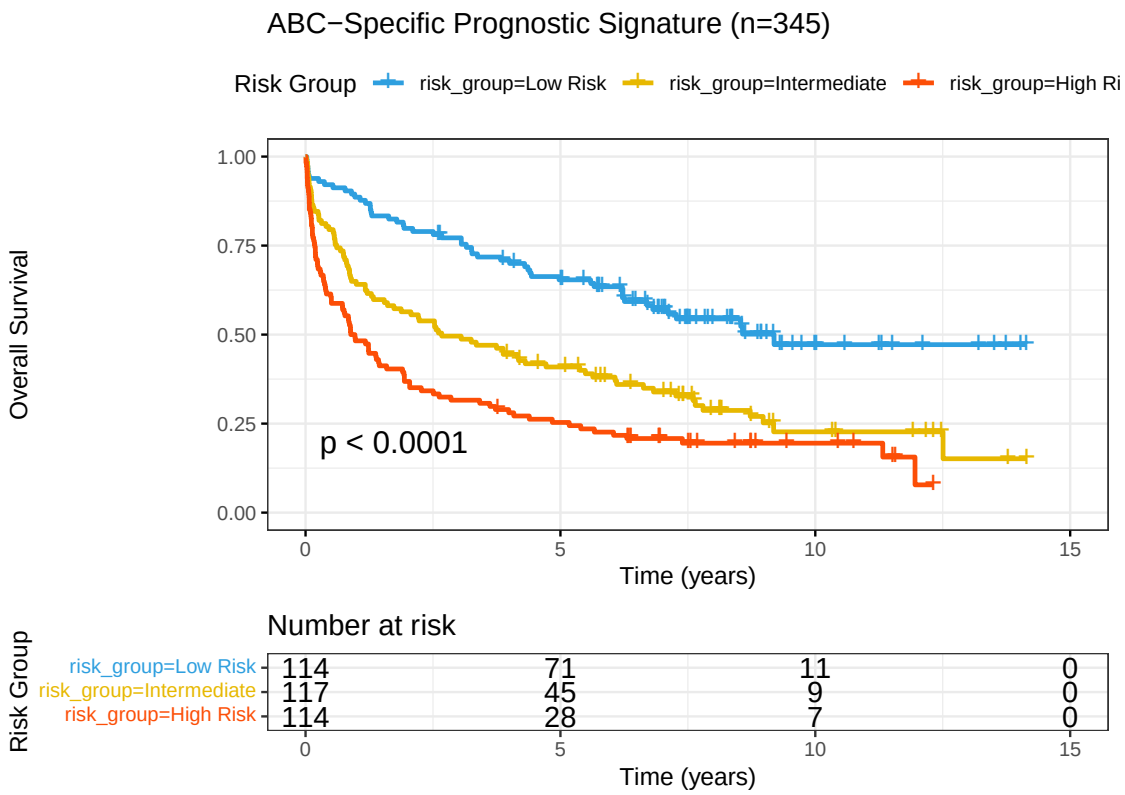

Figure 3: Survival by ABC-Specific Prognostic Signature

3.4 MHG Subset Analysis (n=164)

Table 8: MHG Subset Survival Analysis Summary

Metric	Value
Total probes tested	29,361
Significant (FDR < 0.05)	21
Significant ($p < 0.05$)	2,792

Note: MHG subset shows limited significant genes due to smaller sample size and inherently poor prognosis.

3.5 UNC Subset Analysis (n=277)

Table 9: UNC Subset Survival Analysis Summary

Metric	Value
Total probes tested	29,362
Significant (FDR < 0.05)	373
Protective	195
Adverse	178

4 Pathway Enrichment Analysis (GSEA)

4.1 Methods

Gene Set Enrichment Analysis was performed using **enrichR** with the following databases:

- KEGG 2021 Human
- GO Biological Process 2023
- Reactome 2022
- MSigDB Hallmark 2020
- BioPlanet 2019

4.2 Global Protective Pathways

Table 10: Top Protective Pathways (Good Prognosis)

Database	Pathway	Overlap	Adj. P-value
KEGG	T cell receptor signaling pathway	16/104	4.06e-12
MSigDB Hallmark	Allograft Rejection	19/200	2.37e-11
KEGG	Primary immunodeficiency	9/38	1.84e-08
KEGG	Hematopoietic cell lineage	12/99	4.43e-08
KEGG	Th1 and Th2 cell differentiation	10/92	2.56e-06
MSigDB Hallmark	IL-2/STAT5 Signaling	13/199	4.95e-06
KEGG	Cell adhesion molecules	10/148	1.57e-04
KEGG	PD-L1 expression and PD-1 checkpoint pathway in cancer	8/89	1.57e-04
KEGG	Th17 cell differentiation	8/107	5.26e-04
MSigDB Hallmark	Epithelial Mesenchymal Transition	10/200	8.03e-04
KEGG	Yersinia infection	8/137	2.67e-03
KEGG	Cytokine-cytokine receptor interaction	11/295	7.40e-03
MSigDB Hallmark	Complement	8/200	1.38e-02
MSigDB Hallmark	Coagulation	6/138	2.80e-02
MSigDB Hallmark	Apical Junction	7/200	3.15e-02

Key Protective Pathways:

- **Allograft Rejection** (p = 2.4e-11): T-cell mediated immunity
- **IL-2/STAT5 Signaling** (p = 4.9e-06): T-cell activation
- **T-cell Receptor Signaling**: Immune surveillance

4.3 Global Adverse Pathways

Table 11: Top Adverse Pathways (Poor Prognosis)

Database	Pathway	Overlap	Adj. P-value
MSigDB Hallmark	E2F Targets	20/200	8.73e-12
MSigDB Hallmark	G2-M Checkpoint	15/200	2.46e-07
MSigDB Hallmark	Myc Targets V1	14/200	1.19e-06
KEGG	Cell cycle	10/124	2.90e-04
KEGG	DNA replication	6/36	2.90e-04
KEGG	Mismatch repair	5/23	3.29e-04
MSigDB Hallmark	Myc Targets V2	6/58	4.98e-04
KEGG	RNA transport	10/186	2.92e-03
KEGG	Homologous recombination	5/41	3.71e-03
KEGG	One carbon pool by folate	3/20	4.30e-02
KEGG	Nucleotide excision repair	4/47	5.21e-02
KEGG	Vitamin digestion and absorption	3/24	5.54e-02
MSigDB Hallmark	mTORC1 Signaling	7/200	6.14e-02
MSigDB Hallmark	UV Response Up	6/158	6.14e-02
KEGG	Pyrimidine metabolism	4/56	7.69e-02

Key Adverse Pathways:

- **E2F Targets**: Cell cycle progression
- **G2M Checkpoint**: Proliferation
- **DNA Repair / Homologous Recombination**: Genomic instability
- **MYC Targets**: Oncogenic signaling

4.4 Pathway Summary by COO Subset

Table 12: Pathway Enrichment Summary by COO Subset

Subset	Top Protective	Top Adverse
Global	Allograft Rejection, IL-2/STAT5	E2F Targets, Cell Cycle
GCB	T helper molecules, Apoptosis	E2F/MYC Targets
ABC	RNA metabolism	E2F/MYC Targets, DNA Repair
MHG	Limited enrichment	Cell cycle (nominal)
UNC	TCR signaling, CD8+ T cells	E2F Targets

5 IPI-Independent Analysis (GCB Subset)

5.1 Rationale

The **International Prognostic Index (IPI)** is the standard clinical risk score for DLBCL. We tested whether gene expression signatures provide prognostic information **independent of IPI**.

5.2 Methodology

Analysis Design:

1. **Cohort**: 317 GCB samples with complete IPI data
2. **Model Comparison**:
   - **Univariate**: Gene expression only
   - **Multivariate**: Gene expression + IPI score
3. **Significance**: Genes with adjusted $p < 0.05$ in multivariate model

5.3 IPI-Independence Results

Table 13: IPI Independence Analysis Summary

Metric	Value
Samples with IPI data	317
Univariate significant ($p < 0.05$)	3,037
IPI-adjusted significant ($p < 0.05$)	2,839
Remain significant after IPI adjustment	2,311
Lose significance after IPI adjustment	726
Gain significance after IPI adjustment	528

5.4 Kaplan-Meier: IPI-Independent Signature

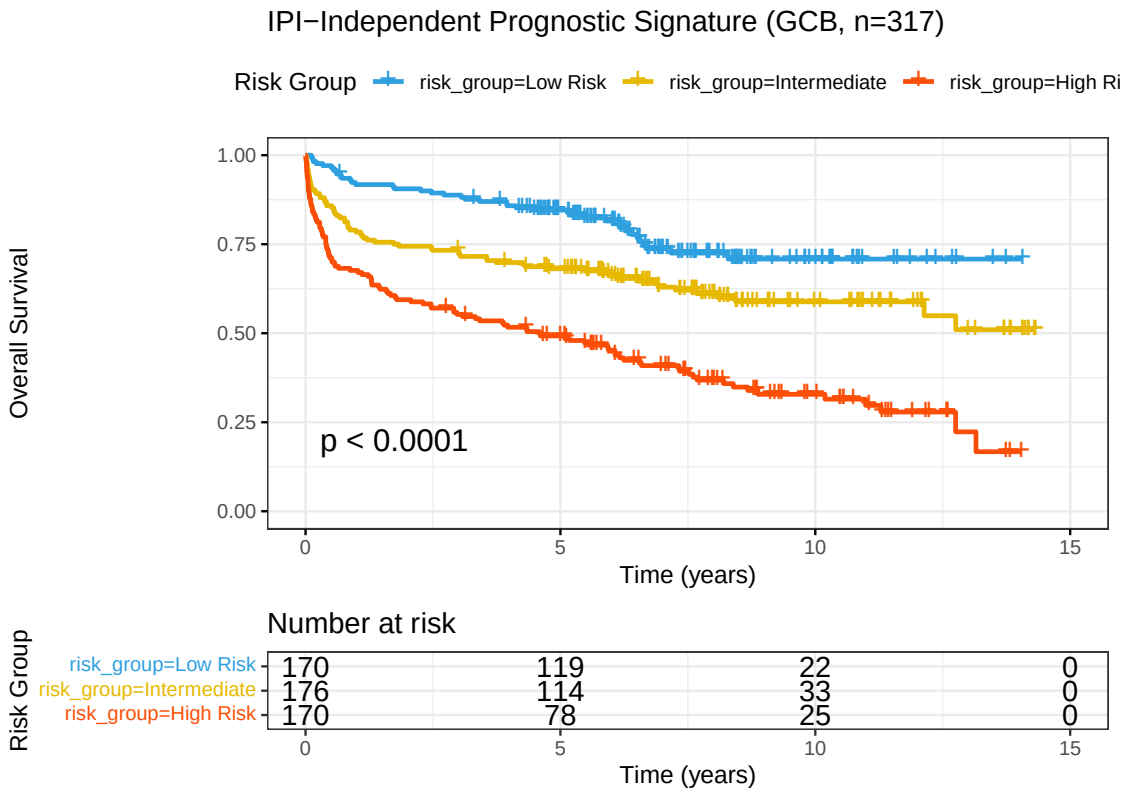

Figure 4: IPI-Independent Signature in GCB Samples with IPI Data

5.5 Top IPI-Independent Genes

5.5.1 Protective (IPI-Independent)

Table 14: Top IPI-Independent Protective Genes

Gene	Univariate HR	Adjusted HR	Uni P	Adj P	IPI P
ZZZ3	0.569	0.552	4.87e-07	2.22e-07	1.46e-07
PRND	0.777	0.773	1.43e-06	9.93e-07	1.78e-07
TRMO	0.682	0.666	2.60e-05	1.01e-05	1.08e-07
LGALS8	0.725	0.710	5.48e-05	2.41e-05	1.43e-07
ZNF792	0.760	0.741	9.03e-05	3.21e-05	1.08e-07
MAL	0.761	0.796	2.24e-06	7.61e-05	3.65e-05
ZNF680	0.733	0.716	1.90e-04	7.80e-05	1.17e-07
CD1E	0.711	0.733	1.65e-05	9.44e-05	4.78e-06
ISCA1	0.747	0.713	6.15e-04	1.17e-04	4.18e-08
ARHGEF6	0.662	0.651	1.51e-04	1.21e-04	2.23e-07
MYNN	0.696	0.670	4.30e-04	1.24e-04	8.98e-08
BBOF1	0.720	0.689	5.33e-04	1.27e-04	5.79e-08
PALD1	0.655	0.638	3.46e-04	1.41e-04	9.22e-08
YIPF3	0.536	0.519	2.47e-04	1.64e-04	1.94e-07
TNFRSF17	0.812	0.816	1.46e-04	1.86e-04	4.06e-07

5.5.2 Adverse (IPI-Independent)

Table 15: Top IPI-Independent Adverse Genes

Gene	Univariate HR	Adjusted HR	Uni P	Adj P	IPI P
PPCDC	1.439	1.434	1.99e-07	3.05e-07	5.56e-07
CSTPP1	1.697	1.674	4.97e-07	1.73e-06	1.29e-06
NOMO1	1.915	2.110	4.07e-05	3.88e-06	1.58e-08
OR51D1	1.750	1.803	2.65e-05	9.82e-06	1.18e-07
ING5	1.672	1.696	1.24e-05	1.07e-05	2.47e-07
HIC1	1.307	1.336	4.10e-05	1.31e-05	6.39e-08
NSD2	1.465	1.458	1.19e-05	2.24e-05	8.20e-07
HDGFL2	1.667	1.704	3.49e-05	2.39e-05	1.73e-07
ARHGAP11B	2.068	2.006	6.68e-06	2.76e-05	3.36e-06
DGKZ	1.707	1.690	1.66e-05	3.13e-05	6.05e-07
SGTA	1.510	1.479	9.17e-06	3.29e-05	2.71e-06
ECE1	1.442	1.452	5.28e-05	4.69e-05	2.40e-07
OTUD7A	2.248	2.182	1.63e-05	4.72e-05	6.36e-07
ERCC1	1.336	1.334	3.75e-05	4.98e-05	4.31e-07
ACOT7	1.442	1.495	1.96e-04	6.51e-05	7.70e-08

5.6 IPI-Independent Pathway Analysis

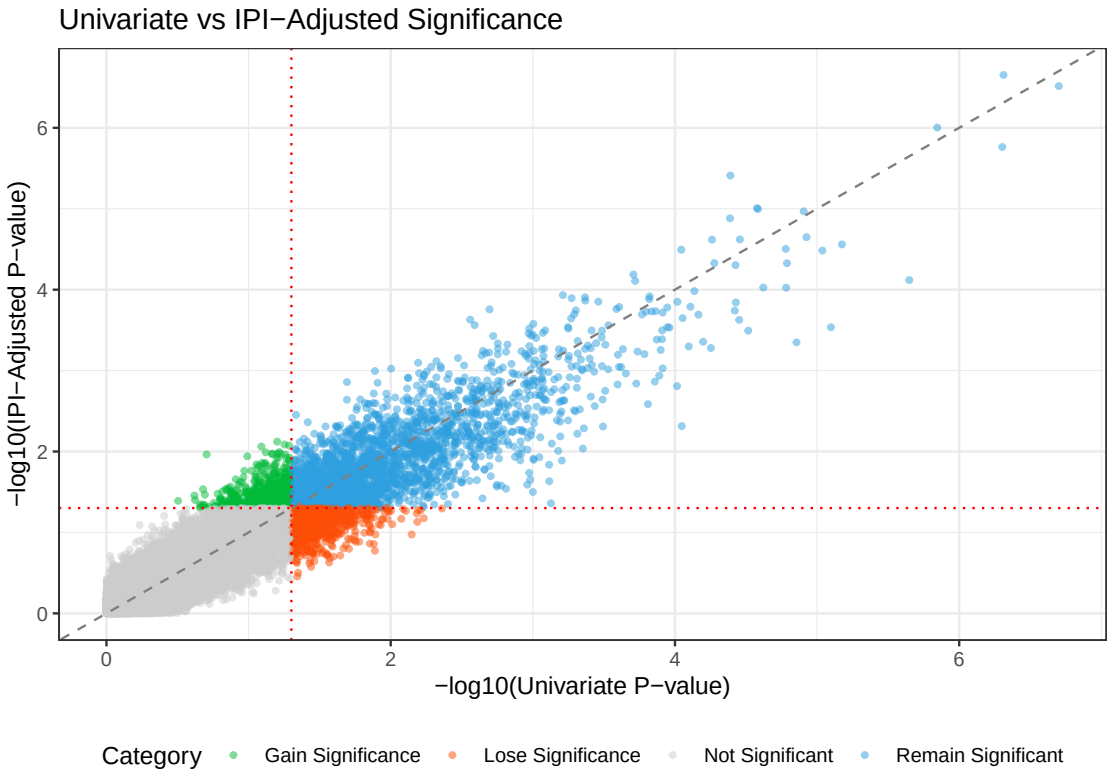

Figure 5: Comparison of Univariate vs IPI-Adjusted Significance

6 Summary and Conclusions

6.1 Key Findings

1. **Strong Prognostic Signatures Exist**
   - Global: 6,748 significant probes (FDR < 0.05)
   - Most genes show IPI-independent prognostic value
2. **COO-Specific Biology**
   - Minimal overlap between GCB, ABC, MHG signatures
   - Each subset has distinct prognostic determinants
3. **Protective Pathways (Good Prognosis)**
   - T-cell receptor signaling
   - Immune activation (IL-2/STAT5)
   - Allograft rejection pathways
4. **Adverse Pathways (Poor Prognosis)**
   - Cell cycle / proliferation
   - E2F and MYC targets
   - DNA repair mechanisms
5. **IPI Independence**
   - 93% of significant genes remain significant after IPI adjustment
   - Gene expression provides additional prognostic information beyond clinical factors

6.2 Clinical Implications

- Gene expression signatures could **refine risk stratification** beyond IPI
- **T-cell infiltration/function** appears protective across subtypes
- **Proliferation signatures** consistently predict poor outcome
- COO-specific signatures may enable **tailored prognostication**

7 Appendix: Subset Comparison

Table 16: Summary Comparison Across COO Subsets

Subset	N	FDR < 0.05	Top Protective Gene	Top Adverse Gene
Global	1303	6748	SIRPG	POLR3G
GCB	517	909	CD3D	QPCTL
ABC	345	967	TCF7	ENPP5
MHG	164	21	FAM167A	UBE2J2
UNC	277	373	PCED1B	CDKN2A

Analysis performed using R with survival, survminer, dplyr, and enrichR packages.

Data source: HMRN/Lacy Cohort (GSE181063)