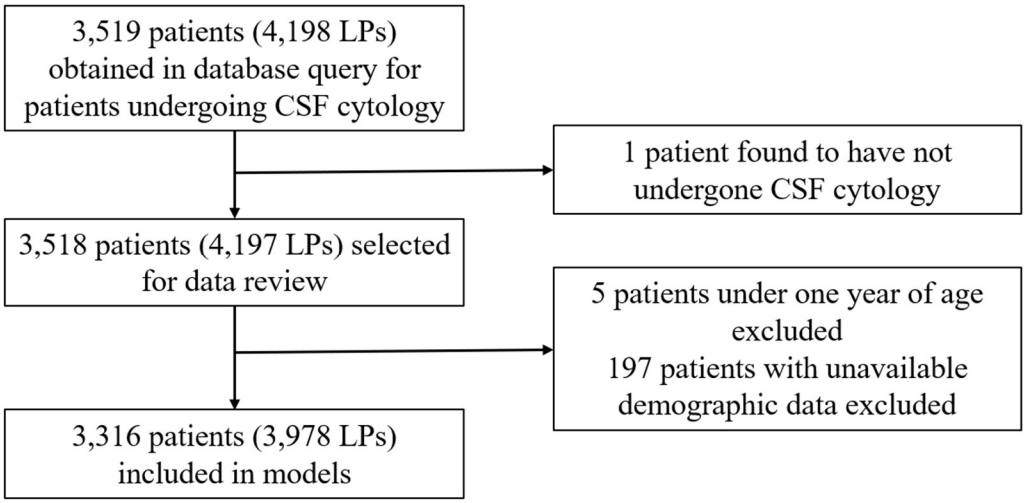
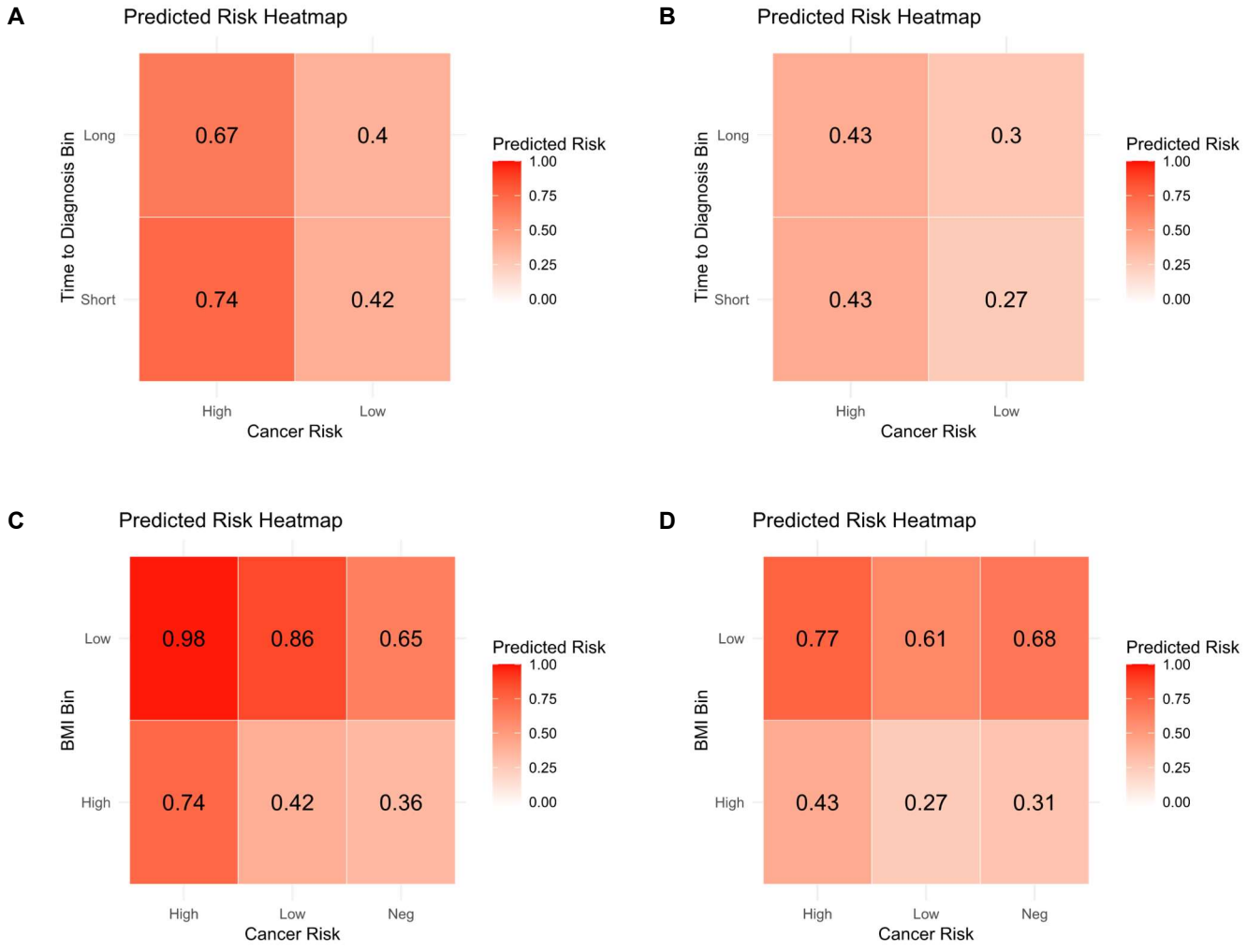
Supplementary Files

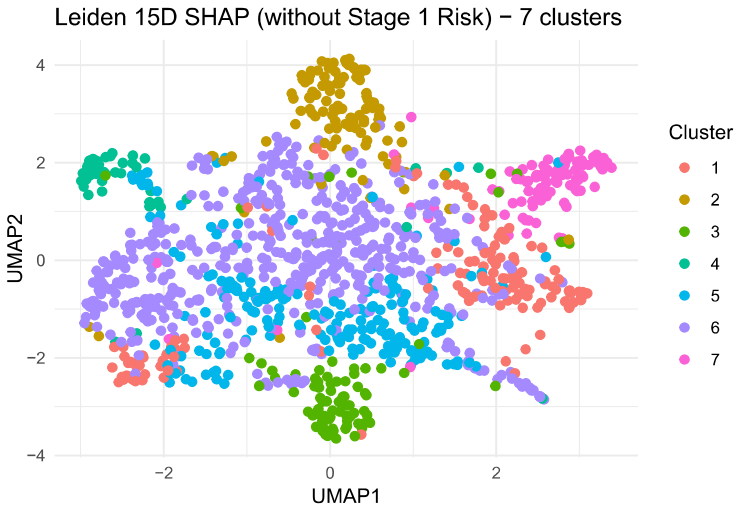
Supplementary Figure 1. Subject inclusion flowchart.



Supplementary Figure 2. Two-dimensional heatmaps for first stage RF model risk score prediction. Plots represent the RF model risk score for old patients (age ≥ 55 years; A, C) and young patients (age < 55 years, B, D). Variables not included in the axes for each plot were defaulted to the lowest-risk result for each variable.



Supplementary Figure 3. Leiden clustering for Stage 2 model, without Stage 1 Risk. Clusters are displayed over UMAP plot for visualization.



Supplementary Table 1. Full patient and clinical characteristics. Numerical values are provided as mean +/- standard deviation or median (interquartile range). Categorical variables are provided as N (%).

	CSF Negative for LMM (n = 3786)	**CSF Positive for LMM (n = 192)**	**p-value**
Demographics			
Age (mean +/- SD)	55.0 +/- 17.6	58.4 +/- 16.3	**0.005**
BMI (mean +/- SD)	28.4 +/- 7.29	26.0 +/- 6.49	**<0.001**
Sex (n, % female)	2059 (51.5)	89 (44.5)	0.0613
Health History			
Smoking (n, %)			
No	1896 (50.1)	95 (49.5)	0.069
Former	1379 (36.4)	81 (42.2)	
Yes	511 (13.5)	16 (8.33)	
Cancer Type (n, %)			
None	2012 (53.1)	18 (9.37)	**<0.001**
Low Risk	658 (17.4)	47 (24.5)	
High Risk	1116 (29.5)	127 (66.1)	
Duration of cancer history*	9.30 (0.21, 64.3)	14.2 (0.99, 49.6)	0.185
Asthma	455 (12.0)	16 (8.33)	0.153
CAD	434 (11.5)	23 (12.0)	0.918
COPD	765 (20.2)	31 (16.1)	0.201
Heart Failure	476 (12.6)	21 (10.9)	0.578
HTN	1969 (52.0)	96 (50.0)	0.639
IBD	155 (4.09)	18 (9.37)	**<0.001**
Multiple Sclerosis	481 (12.7)	7 (3.64)	**<0.001**
Hx of transplant	310 (8.19)	24 (12.5)	**0.049**
NSAIDs	617 (16.3)	34 (17.7)	0.678
Steroids	524 (13.8)	23 (12.0)	0.533
Chemotherapy	434 (11.5)	53 (27.6)	**<0.001**

CSF Analysis			
RBC	2 (0, 36)	8 (8, 87)	**<0.001**
TNC	2 (1, 6)	11 (3, 48)	**<0.001**
TNC corrected	1 (1, 4)	10 (3, 48)	**<0.001**
Lymphocytes %	78 (58, 89)	58 (25, 80.5)	**<0.001**
Protein	29 (42, 65)	40 (71, 167)	**< 0.001**
Glucose	64 (57, 77)	38 (57, 71)	**< 0.001**
Labs			
Neutrophil count	5.05 (3.36, 7.62)	5.64 (3.50, 8.65)	0.141
Lymphocyte count	1.41 (0.86, 2.06)	0.94 (0.56, 1.48)	**< 0.001**
Hemoglobin	12.5 (10.5, 13.9)	11.9 (10.2, 13.4)	**0.021**
Platelets	230 (177, 287)	201 (136.5, 266)	**<0.001**
Total Protein	6.8 (6.1, 7.3)	6.4 (5.9, 7.0)	**<0.001**
AST	22 (17, 31)	21 (16, 29)	0.125
ALT	20 (14, 35)	25 (14, 39)	0.117
BUN	15.0 (11.0, 20.0)	16.0 (11.0, 19.5)	0.964
Creatinine	0.68 (0.84, 1.05)	0.60 (0.74, 0.94)	**<0.001**
Chloride	102.0 (100.0, 105.0)	101.0 (98.5, 105.0)	**0.0014**

* Only for patients with a positive history of cancer prior to their LP.

Supplementary Table 2. Random forest model performance in select subgroups.

Subgroup	Sensitivity	Specificity	Accuracy
Cancer History			
Lymphoma	80.0	68.3	0.696
Breast	75.0	75.6	0.756
Brain	80.0	62.0	0.632
Skin	1.00	53.7	0.641
Negative History	25.0	95.1	0.943
Cancer Risk			
High Risk	85.4	67.8	0.696
Low Risk	94.1	68.1	0.707
Age			
Young	80.0	87.9	0.877
Old	80.4	77.8	0.780

Supplementary Table 3. Leiden clustering of Stage 1 and Stage 2 of the RF model. Cells represent the percent (Stage 1, except median risk score) or median (Stage 2) value for each feature per cluster. For binary variables in Stage 1, the percent of one class is shown; variables with > 2 classes have all classes shown.

Stage 1 Feature	Cluster 1 (n=189)	Cluster 2 (n=185)	Cluster 3 (n=49)	Cluster 4 (n=193)	Cluster 5 (n=79)	Cluster 6 (n=158)	Cluster 7 (n=142)
Median risk score	0.131	0.967	0.854	0.418	0.148	0.341	0.110
Positive for LMM	5.3	13.0	12.2	2.1	2.5	1.3	1.4
Age ≥ 56.5	51.9	61.1	75.5	59.1	41.8	44.9	31.7
BMI ≥ 30	37.6	28.6	20.4	40.9	35.4	43.7	41.5
Hypertension	46.6	49.2	69.4	62.2	55.7	44.3	40.8

COPD	13.2	10.8	69.4	51.8	5.06	7.59	17.6
MS	15.3	6.48	6.12	10.4	11.4	9.49	31.7
Current smoker	18.0	13.0	20.4	16.6	15.2	13.3	22.5
Former smoker	39.7	35.7	49.0	37.8	22.8	27.2	31.7
Never smoker	42.3	51.4	30.6	45.6	62.0	59.5	75.8
Chemotherapy	10.6	9.73	2.04	11.4	6.33	3.80	6.34
High-risk cancer history	22.2	41.1	59.2	21.2	21.5	21.5	26.0
Low-risk cancer history	9.52	18.9	18.4	24.9	6.33	10.1	11.3
Cancer history ≥ 4.5 mo	19.6	33.5	67.3	26.9	8.86	13.3	19.7
Cancer history < 4.5 mo	12.7	26.5	10.2	21.2	19.0	18.3	19.7
No cancer history	68.3	40.0	22.4	53.9	72.2	68.3	62.7

Stage 2 Feature	**Cluster 1 (n=642)**	**Cluster 2 (n=121)**	**Cluster 3 (n=62)**	**Cluster 4 (n=69)**	**Cluster 5 (n=232)**	**Cluster 6 (n=93)**
Median risk score	0.0085	0.0092	0.2363	0.0210	0.0228	0.0086
Positive for LMM	3.9	2.5	35.5	4.3	4.3	3.2
Stage 1 Risk	0.192	0.564	0.807	0.597	0.664	0.341
CSF RBC	2.0	3.0	13.5	1.0	1.5	3.0
CSF TNC	1.0	1.0	39.5	1.0	1.0	2.0
CSF Lymph Percentage	78.0	67.0	52.5	74.0	84.0	85.0
CSF Protein	42	45	127	43	35	48
CSF Glucose	63	71	43	66	59	64
Lymphocyte	1.56	0.88	0.80	1.34	1.33	1.59
Neutrophil	4.90	6.55	7.26	6.58	3.69	4.57
Hemoglobin	12.9	9.7	12.0	11.8	12.1	12.8
Platelet count	235	177	213	410	229	254
Total protein	6.8	6.0	6.3	6.9	6.9	8.1
ALT	21	19	19	30	16	23
AST	21	29	19.5	27	19	24
BUN	15	36	16	15	13	14
Creatinine	0.81	2.15	0.75	0.75	0.80	0.87
Chloride	102	102	102	102	103	101

Supplementary Table 4. Exploratory analysis of alternate definitions of LMM positivity. "Imaging" model defined positive cases by MRI of brain and spine. "Combined" model defined positive cases as those meeting either "Imaging" criteria or the CSF criteria used in the main model. Models were trained on the same pre-specified Discovery dataset; metrics represent model performance on the same held-out Validation dataset.

Model Info	# Predictors	AUC	Sensitivity	Specificity	Accuracy
Imaging					
Stage 1 Model	9	0.751	83.1	63.1	0.650
Stage 2 Model	16	0.741	80.3	59.6	0.621
Combined					
Stage 1 Model	9	0.728	75.8	63.6	0.648
Stage 2 Model	16	0.743	73.2	65.3	0.662

Supplementary Table 5. Predictor variables included in two-stage random forest model. BMI = body mass index, COPD = chronic obstructive pulmonary disease, MS = multiple sclerosis, LP = lumbar puncture, RBC = red blood cell, TNC = total nucleated cell.

Stage 1	
Predictor	**Format**
Age	Binned: High ≥ 55 years, Low < 55 years
BMI	Binned: High ≥ 30 kg/m^2, Low < 30 kg/m^2
Prior history of malignancy	Binned: Long ≥ 4.5 months, Short < 4.5 months, Negative (if no known cancer diagnosis prior to LP)
Malignancy risk level	Binned: High, Low, N/A (if no known cancer diagnosis prior to LP).
Smoking history	Current, Former, Never
Prior receipt of chemotherapy	Yes/No
Prior medical history of hypertension	Yes/No
Prior medical history of COPD	Yes/No
Prior medical history of MS	Yes/No
Stage 2	
Predictor	**Format**
CSF RBC count	Cells per μL
CSF TNC count (corrected for RBC count)	Cells per μL (correction: CSF TNC – (CSF RBC/500)
CSF lymphocyte percentage	Percentage
CSF protein	mg/dL
CSF glucose	mg/dL
Hemoglobin	g/dL
Platelet count	K/μL
Neutrophil count	K/μL
Lymphocyte count	K/μL
Blood urea nitrogen	mg/dL
Creatinine	mg/dL
Chloride	mmol/L
Total protein	g/dL
ALT	U/L
AST	U/L

Supplementary Table 6. Classification of prior malignancy risk. Assignment based on current literature estimates of LMM prevalence within each malignancy type; prevalence >5% constituted high risk. Several malignancies are so uncommonly associated with LMM that no incidence estimates were identified in the literature; these are represented as “Rare”. Unknown cancers were defaulted to Low Risk.

Malignancy Type	Incidence of LMM	References
High Risk		
Breast	5-11.4%	Lamba (2022), Mittica (2015), Bhambhvani (2020)
Lung	9-25%	Bhambhvani (2020), Nayar (2017)
Melanoma	6-75%	Bhambhvani (2020), Nayar (2017), Lamba (2021)
Medulloblastoma (children only)	20-30%	Russo (2022)
Primary CNS glioma	4.7-16.2%	Anderson (2019), Park (2023)
Leukemia	1-19%	Bhambhvani (2020), Rozovski (2015)
Lymphoma	5-30%	Bhambhvani (2020), Nayar (2017)
Low Risk		
Bone	1.8%	Gensler (2025)
Ependymoma	2.2%	Benesch (2019)
Eye	Rare	Fedorenko (2016)
Meningioma	0.9%	Park (2015)
Pituitary tumor	Rare	Lehman (2003)
Carcinoid	Rare	Drappatz (2007)
Gastrointestinal (all)	0.61%	Baccili (2024)
Genitourinary (all)	0.03%	Yust-Katz (2013)
Gynecologic (all)	Rare	Yano (2020)
Neuroendocrine	Rare	Yang (2017)
Sarcoma	Rare	Zhang (2024)
Thyroid	Rare	Drappatz (2007
Head and neck	Rare	Lamba (2021)
Heart	Rare	Marrone (2025)
Histiocytosis	Rare	Abrassart (2025)
Myeloma	1.2%	Katoudritou (2024)
Mesothelioma	Rare	Jiang (2019), Petrovic (2004)