Screening Human Lung Cancer with Predictive Models of Serum Magnetic Resonance Spectroscopy Metabolomics

Short Title: Lung Cancer Screening with Serum Metabolomics Models

Tjada A. Schult[1,2,†], Mara J. Lauer[1,3,†], Yannick Berker[4], Marcella R. Cardoso[1], Lindsey A. Vandergrift[1], Piet Habbel[2], Johannes Nowak[5], Matthias Taupitz[2], Martin Aryee[1], Mari A. Mino-Kenudson[1], David C. Christiani[1,6,*] and Leo L. Cheng[1,*]

1 Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

2 Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

3 University of Würzburg (JMU), Würzburg, Germany

4 German Cancer Research Center (DKFZ), Heidelberg, Germany

5 Radiologische Praxis Gotha (SRH), Gotha, Germany

6 Harvard T.H. Chan School of Public Health, Boston, MA, USA

† Authors with equal contributions

* Correspondence:

David C. Christiani, MD, MPH

665 Huntington Ave. 1-1401, Boston, MA 02115, USA

Christiani.david@mgh.harvard.edu 617-432-3323

or

Leo L. Cheng, PhD

149 13th street, MGH, Charlestown, MA 02129, USA

lcheng@mgh.harvard.edu 617-724-6593

Supplemental Materials

Table S1. Demographic Data for the studied populations.

Training & Validation Cohort

| | FEMALE | | | | | | | | | MALE | | | | | | | | | TOTAL | | |
| | Non-Smoker | | | Former Smoker | | | Smoker | | | Non-Smoker | | | Former Smoker | | | Smoker | | | | | |
Cell Type	N CHI*	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)
Localized Disease (Stage I & IIA)																					
Adeno CA	3	65.5±3.4	3.5±1.1	5	65.9±3.3	3.6±1.4	1	65.4	3.7										9	65.7±3.6	3.6±1.1
Ctrl	2.3±0.6	65.0±4.6		2.0±0.0	65.4±4.0		2	65.0											2.1±0.3	65.2±3.7	
SCC	1	67.5	4.3	1	68.2	3.2										1	70.0	1.6	3	68.6±1.3	3.0±1.4
Ctrl	2	67.0		2	68.0											3	70.0		2.3±0.6	68.3±1.5	
Mix							1	65.7	3										1	65.7	3
Ctrl							2	65.0											2	65	
Advanced Disease (Stage IIB, III, & IV)																					
Adeno CA				3	70.1±3.6	2.0±2.1							2	67.0±1.5	3.2±1.7	1	63.5	3.8	6	68.0±3.6	2.0±1.5
Ctrl				2.3±0.6	69.7±4.0								2.0±0.0	66.5±2.1		2	63.0		2.2±0.4	67.5±3.8	
SCC													2	71.2±2.0	2.0±0.6				2	71.2±2.0	
Ctrl													2.5±0.7	70.5±2.1					2.5±0.7	70.5±2.1	
Mix				1	51.4	2.3							1	55.6	0.7				2	53.5±3.0	1.5±1.1
Ctrl				1	51.0								1	55.0					1.0±0.0	53.0±2.8	
Pathological Stage & Cell Type Unknown**																					
				1	80.7	1.4	1	72.3	0.5										2	76.5±6.0	1.0±0.6
Ctrl				4	80.0		3	72.0											3.5±0.7	76.0±5.7	
Total	4	66.0±3.8	3.7±0.9	11	67.3±7.4	2.8±1.5	3	67.8±3.9	2.4±1.7				5	66.4±6.5	2.2±1.4	2	66.8±4.6	2.7±1.6	25	66.9±5.9	2.8±1.4
	2.3±0.5	65.5±3.8		2.2±0.8	66.8±7.5		2.3±0.6	67.3±4.0					2.0±0.7	65.8±6.5		2.5±0.7	66.5±4.9		2.2±0.6	66.4±5.9	

Testing Cohort

| | FEMALE | | | | | | | | | MALE | | | | | | | | | TOTAL | | |
| | Non-Smoker | | | Former Smoker | | | Smoker | | | Non-Smoker | | | Former Smoker | | | Smoker | | | | | |
Cell Type	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)	N CHI	Age (yrs)	Prior (yrs)
Localized Disease (Stage I & IIA)																					
Adeno CA	5	63.1±12.5	0.4±0.6	6	67.5±10.4	0.9±0.8	4	60.1±6.5	0.6±0.4	2	57.3±9.2	0.4±0.4	2	66.7±1.5	0.8±1.1	1	54.3	<0.1	20	63.2±9.5	0.6±0.6
Ctrl	2.0±1.2	62.4±12.4		2.0±0.9	64.5±8.8		1.3±0.5	58.0±5.1		1.5±0.7	56.5±9.2		2.0±0.0	66.0±1.4		1	53.0		1.8±0.9	61.5±8.7	
SCC	1	65.5	0.1				3	72.0±10.3	0.8±0.6	1	74.7	1.1							5	71.2±6.0	0.7±0.5
Ctrl	2	65.0					3.0±1.0	70.3±8.0		3	74.0								2.8±0.8	70.0±6.5	
Mix				2	67.2±1.9	<0.1													2	67.2±1.9	<0.1
Ctrl				2.0±0.0	65.5±2.1														2.0±0.0	65.5±2.1	
Advanced Disease (Stage IIB, III, & IV)																					
Adeno CA	3	61.1±11.2	1.7±1.2	8	67.9±9.2	2.4±0.9	5	58.0±4.9	0.5±0.3	2	68.4±5.4	0.9±1.1	2	67.5±1.2	1.0±0.4	2	61.8±5.2	0.8±0.1	22	64.2±8.1	0.6±0.5
Ctrl	1.7±1.2	60.7±11.0		2.4±0.9	66.5±9.3		1.2±0.4	56.8±4.5		2.5±0.7	68.0±5.7		2.0±0.0	67.0±1.4		1.5±0.7	60.5±3.5		1.9±0.9	63.1±8.1	
SCC				2	53.6±9.9	0.4±0.5	1	70.9	1	1	51.5	<0.1				1	63.3	<0.1	5	58.6±9.6	0.3±0.5
Ctrl				1.0±1.4	53.0±11.3		2	68.0		1	51.0					2	61.0		1.4±0.9	57.2±9.1	
Total	9	62.7±10.5	0.3±0.5	18	66.1±9.7	0.6±0.6	13	62.9±6.6	0.7±0.4	6	62.9±10.1	0.6±0.7	4	67.1±1.2	0.9±0.7	4	60.3±5.0	0.4±0.5	54	64.0±9.0	0.6±0.5
	1.9±1.1	62.1±10.4		2.1±0.9	64.2±9.1		1.7±0.9	61.2±7.8		2.0±0.9	62.3±10.1		2.0±0.0	66.5±1.3		1.5±0.6	59.8±4.3		1.9±0.9	62.7±9.4	

*CHI = Charlson Health Index of matched Controls.
**For these patients there is an absence of pathology data.
Prior: Years of biobank samples PriorDx. Adeno Ca: Adenocarcinomas. SCC: Squamous cell carcinomas.
Mix: Mixed cell type cancer. Ctrl: Healthy controls.

Table S2. HRMAS MRS spectral regions, their potential contributing metabolites, means and standard deviations of these regions measured from the training cohort, and their overall coefficients in the calculation of the reported canonical score from the included principal components.

Spectral Regions	Possible Metabolites	Mean	SD	Coeff	Coeff*Mave
4.35-4.34	AMP,GDP,GTP,NAD,NADP	0.000697	0.000692	216.682	0.151
4.33-4.27	4,ATP,GDP,GPC,GTP,NADP,Thr	0.005243	0.002107	97.031	0.509
4.26-4.24	F6P,GDP,GTP,NAD,NADP,Thr	0.002019	0.001031	33.539	0.068
4.16-4.08	3,ADP,F6P,GDP,Lac,Pro	0.022348	0.013978	-18.016	-0.403
4.07-4.05	9,ADP,DPG,G6P,m-Ino,Trp	0.002843	0.001567	-87.622	-0.249
4.02-3.99	1,1,5-AG,3PG,AMP,Asn,DPG,G6P,His,IMP,Phe	0.004071	0.003635	-45.822	-0.187
3.98-3.96	1,5-AG,Asn,F6P,G6P,His,MH,Phe,Ser	0.004067	0.002264	5.120	0.021
3.95-3.93	1,5-AG,F6P,G6P,GPC,MH,Ser,Tyr	0.003168	0.002229	-11.615	-0.037
3.92-3.91	Cr,F6P,G6P,GPC,Ser,Tyr	0.015964	0.004352	-6.256	-0.100
3.90-3.89	5,1,5-AG,3PG,Asp,F6P,G6P,Glc,GPC	0.021959	0.005614	-4.435	-0.097
3.88-3.87	3PG,Asp,F6P,G6P,Glc,GPC	0.002068	0.001847	17.412	0.036
3.86-3.82	3,1,5-AG,3PG,ALC,F6P,Glc,GPC,Met,Ser	0.041386	0.010788	-2.765	-0.114
3.81-3.76	3,8,Ala,Arg,Glc,Gln,Glu,GSSG,Lys,Orn,	0.035509	0.007736	-13.041	-0.463
3.75-3.73	3,8,Ala,Arg,G6P,Glc,Gln,Glu,GSSG,Leu,Lys	0.023198	0.006620	11.025	0.256
3.72-3.71	3,G6P,Glc,Leu	0.024683	0.012891	-6.001	-0.148
3.70-3.67	1,5-AG,F6P,G6P,Glc,GPC,Leu,MH	0.011983	0.006237	-24.535	-0.294
3.66-3.64	1,5-AG,F6P,GPC,Ile	0.008435	0.003978	-35.667	-0.301
3.63-3.60	1,5-AG,ALC,F6P,GPC,m-Ino,Val	0.006411	0.001936	-16.570	-0.106
3.59-3.56	1,1,5-AG,ALC,F6P,G6P,GPC,Thr	0.009371	0.004042	11.072	0.104
3.55-3.52	F6P,G6P,Glc,m-Ino	0.021868	0.006409	-7.029	-0.154
3.51-3.49	G6P,Glc,m-Ino,Trp	0.015213	0.006238	0.719	0.011
3.48-3.45	G6P,Glc,Trp	0.042398	0.014055	-1.440	-0.061
3.44-3.43	1,5-AG,Glc,Pro,Tau	0.003928	0.001641	26.229	0.103
3.42-3.40	1,10,11,1,5-AG,Glc,Pro,Tau	0.040611	0.012635	-1.913	-0.078
3.39-3.38	Glc,Pro	0.006539	0.004436	-12.749	-0.083
3.37-3.32	1,5-AG,2,GSSG,Pro	0.005504	0.002623	46.543	0.256
3.27-3.26	1,5-AG,7,G6P,His,m-Ino,Phe,Tau,Trp	0.013070	0.005658	-8.594	-0.112
3.25-3.21	5,7,10,1,5-AG,Arg,Glc,His,m-Ino,Phe,Tau,	0.080753	0.020492	8.392	0.678
3.15-3.14	8,11,His,MH	0.001398	0.001144	-147.481	-0.206
3.12-3.11	8,Phe	0.001238	0.001794	-59.334	-0.073
3.07-2.99	7,9,2-OG,Cr,Lys,MH,Orn,Tyr	0.017550	0.004119	32.713	0.574
2.94-2.92	Asn,GSSG	0.002232	0.001023	169.321	0.378
2.91-2.88	TBD	0.002257	0.001427	155.687	0.351

2.84-2.83	Asn	0.000937	0.000922	225.851	0.212
2.81-2.80	Asp,G3P	0.001618	0.001532	62.998	0.102
2.79-2.71	7,Asp	0.010576	0.003605	54.339	0.575
2.70-2.68	4,7,Asp	0.002685	0.000839	14.861	0.040
2.67-2.66	4,7,Asp,Cit	0.000926	0.000532	-124.268	-0.115
2.48-2.40	10,2-OG,ALC,Gln,GSSG	0.013154	0.003326	-10.582	-0.139
2.39-2.32	4,Gln,Glu,Pro,Suc	0.008611	0.003089	-32.737	-0.282
2.31-2.22	Glu,Met,Pro,Val	0.016407	0.005115	3.945	0.065
2.16-2.11	ALC,Gln,Glu,GSSG,Met	0.015458	0.002328	-28.635	-0.443
2.09-2.07	6,Gln,Glu,Met,Pro	0.009938	0.003045	-29.547	-0.294
2.06-1.99	Glu,Ile,Pro	0.051105	0.013603	5.663	0.289
1.94-1.88	8,Arg,Ile,Lys,Orn,Pro	0.007918	0.002073	62.756	0.497
1.75-1.69	Arg,Leu,Lys,Orn	0.009426	0.002173	28.872	0.272
1.62-1.55	8,Arg	0.009396	0.006181	25.258	0.237
1.54-1.53	8,Lys	0.001267	0.001057	-13.664	-0.017
1.52-1.45	8,Ala,Ile,Lys	0.016822	0.003311	20.462	0.344
1.35-1.34	TBD	0.040360	0.029587	-6.954	-0.281
1.33-1.32	Lac,Thr	0.055574	0.030118	-8.471	-0.471
1.31-1.27	Ile,Lac,Thr	0.084620	0.061008	2.311	0.196
1.06-1.03	Val	0.007862	0.001853	10.980	0.086
1.02-1.01	Val	0.001967	0.001805	48.495	0.095
1.00-0.98	Ile,Val	0.009306	0.002900	20.223	0.188
0.97-0.92	Ile,Leu,Val	0.024732	0.005288	23.169	0.573
0.91-0.84	Ile	0.099356	0.021767	10.720	1.065
				Label	
	Caffeine			1	
	1,7-Dimethyl-xanthine			2	
	Glycerate			3	
	Malate			4	
	Betaine			5	
	Dimethyl-proline			6	
	Carnosine			7	
	Citrulline			8	
	Creatinine			9	
	Carnitine			10	
	Ergothioneine			11	

	Diphospho-glycerate			DPG	
	Fructose-6-phosphate			F6P	
	Glucose-6-phosphate			G6P	
	Glyceraldehyde-3-phosphate			G3P	
	3-Phosphoglycerate			3PG	
	1,5-Anhydroglucitol			1,5-AG	
	Glucose			Glc	
	2-Oxoglutarate			2-OG	
	Succinate			Suc	
	Methyl-histidine			MH	
	Ornithine			Orn	
	Acetyl-carnitine			ALC	
	Glycerophosphocholine			GPC	
	Glutathione disulfide			GSSG	

Coeff: The overall loading factors based on principal component and canonical analyses conducted with the training cohort. Coeff*M_{ave}: Contributing overall loading factors for each spectral region weighted by the mean of the region calculated from the Training cohort. Red colored regions: larger than the median contribution towards NSCLC; Green colored regions: larger than the median contribution towards Healthy controls.

Table S3. **Co-variance analyses of the canonical score against the following potential co-variances.**

Variance	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Days between biobank sample and diagnosis	1	1	0.26042	0.243235	*0.624389*
Age at first serum sample	1	1	0.071566	0.066843	*0.797224*
Gender	1	1	0.706995	0.660339	*0.420917*
Smoker status	2	2	1.416581	0.66155	*0.521224*