

Supporting information for:

Bayesian network inference enables unbiased phenotypic anchoring of transcriptomic responses to cigarette smoke in humans

Danyel G.J. Jennen, Danitsja M. van Leeuwen, Diana M. Hendrickx, Ralph W.H. Gottschalk, Joost*

H.M. van Delft, Jos C.S. Kleinjans

Department of Toxicogenomics, Maastricht University, Universiteitssingel 40, 6229 ER Maastricht, the
Netherlands

Corresponding Author

*(D.G.J.J.) Phone: +31 433883983, E-mail: danyel.jennen@maastrichtuniversity.nl

Table S1. Detailed characteristics of the study population.

Subject ID	DNA adducts per 10 ⁸ nt	Plasma cotinine (ng/ml)	Sex	Birth date	Age @ study	Smoking status	Packyears	Cigarettes per day
PPA0 NS	0.36	4.73	male	3/05/77	27	nonsmoker		
PPD0 NS	0.52	4.92	male	18/04/79	25	nonsmoker		
PPB0 NS	0.93	4.94	female	6/01/82	23	nonsmoker		
PPG0 NS	0.24	4.96	female	26/12/69	35	nonsmoker		
PP5 NS	0.38	5.52	male	8/05/66	39	nonsmoker		
PPH0 NS	0.48	5.62	male	20/11/77	27	nonsmoker		
PP1 NS	0.01	5.69	male	2/01/77	28	nonsmoker		
PPC0 NS	0.67	5.70	female	22/08/80	24	nonsmoker		
PP3 NS	0.12	5.71	female	18/03/80	25	nonsmoker		
PP26 NS	0.53	5.86	male	26/04/66	39	nonsmoker		
PP24 NS	0.54	5.91	male	22/03/74	31	nonsmoker		
PP23 NS	0.53	6.02	male	13/08/79	26	nonsmoker		
PP16 NS	0.87	6.16	female	17/02/82	23	nonsmoker		
PP19 NS	0.41	6.29	male	16/09/80	25	nonsmoker		
PPE0 NS	0.40	6.34	female	29/11/77	27	nonsmoker		
PP7 NS	0.23	6.84	male	12/06/68	37	nonsmoker		
PPL1 S	0.29	40.80	male	23/05/66	38	smoker	12	12
PPB1 S	1.29	60.88	female	6/01/82	23	smoker	4	10
PP14 S	0.28	111.30	female	6/05/83	22	smoker	5	10 to 15
PPM1 S	0.26	114.06	female	24/10/78	26	smoker	1	13
PPA1 S	0.99	126.63	male	3/05/77	27	smoker	2	3
PP11 S	0.53	193.86	male	15/07/78	27	smoker	6	8
PPG1 S	0.79	197.09	female	26/12/69	35	smoker	15	20

PPE1 S	2.68	223.99	female	29/11/77	27	smoker	10	13
PP2 S	0.82	240.91	male	9/07/74	31	smoker	11	12
PP20 S	1.00	249.40	male	17/04/84	21	smoker	6	10
PP15 S	0.70	276.02	female	17/02/84	21	smoker	5	15
PPC1 S	0.97	342.53	female	22/08/80	24	smoker	5	20
PPD1 S	0.57	354.98	male	18/04/79	25	smoker	5	25
PP9 S	0.77	385.52	female	31/03/79	26	smoker	7	13
PP18 S	1.13	473.19	female	12/04/77	28	smoker	13	12
PP4 S	1.25	477.73	female	7/12/83	21	smoker	8	20
PPH1 S	0.83	524.35	male	20/11/77	27	smoker	10	25
PP6 S	0.08	613.10	male	21/05/66	39	smoker	18	25
PP22 S	0.83	629.27	male	3/02/79	26	smoker	10	10
PP21 S	0.44	668.04	male	2/12/80	24	smoker	10	10
PP25 S	0.48	696.89	male	10/01/68	37	smoker	20	20

Table S2. Gene expression values (logFC) and p-values of the differentially expressed genes in the Bayesian network presented in Figure 1B based on a limma moderated t-test (absolute FC>1.2; p-value <0.05) for all individuals (n=37; unpaired analysis) and seven twin pairs (n=14; paired analysis).

Up-regulated		All individuals		Twin pairs	
AgilentID	GeneSymbol	logFC	p-value	logFC	p-value
A_23_P115417	<i>RGLI</i>	0.39227381	0.000808129	0.414571429	0.002410906
A_23_P119916	<i>WNT6</i>	0.279199405	0.004914783	0.126714286	0.251233607
A_23_P168229	<i>TXNDC5</i>	0.508577381	0.004588201	0.719428571	0.005894578
A_23_P258221	<i>ABCC5</i>	0.325276786	0.033874319	0.209	0.157869099
A_23_P31376	<i>LRRN3</i>	0.768130952	0.001079874	0.935142857	0.000823128
A_23_P79518	<i>IL1B</i>	0.4205625	0.00392149	0.364285714	0.030573595
A_24_P123325	<i>TMEM33</i>	0.328699405	0.00186518	0.384428571	0.001001227
A_24_P270814	<i>CRK</i>	0.274845238	0.009854625	0.303571429	0.037723891
A_24_P381604	<i>ITM2B</i>	0.287590857	0.037186765	0.331361857	0.006199428
A_24_P940135	<i>CTBS</i>	0.312693452	0.024337886	0.320571429	0.021984862
A_23_P136964	<i>RPGR</i>	0.399231929	0.007829469	0.244285714	0.10822294
A_23_P168965	<i>C8ORF44</i>	0.279550595	0.011251074	0.211714286	0.040754767
A_23_P53257	<i>AVIL</i>	0.337383929	0.017856128	0.117571429	0.357347059
Down-regulated		All individuals		Twin pairs	
AgilentID	GeneSymbol	logFC	p-value	logFC	p-value
A_23_P116765	<i>LALBA</i>	-0.341467262	0.026630571	-0.317857143	0.039851353
A_23_P144126	<i>FETUB</i>	-0.324129125	0.026782515	-0.370152286	0.00660538
A_23_P147277	<i>TSEN54</i>	-0.289642857	0.000957359	-0.258285714	0.016119455
A_23_P165360	<i>ASB1</i>	-0.295734503	0.00162521	-0.275161857	0.01588848
A_23_P29179	<i>CRYBB3</i>	-0.311184524	0.008083852	-0.209285714	0.090720581

A_23_P313031	<i>LOC100509490</i>	-0.308702381	0.007426683	-0.129	0.280925625
A_23_P345710	<i>XXYLT1</i>	-0.370886905	0.001740739	-0.500304714	0.000938495
A_23_P371885	<i>ARL17A</i>	-0.580032122	0.023127751	-0.294733286	0.316614441
A_23_P393777	<i>PTGDR</i>	-0.439681548	0.000177817	-0.375428571	0.022231381
A_23_P397937	<i>SAMD3</i>	-0.387809524	0.002548021	-0.458714286	0.002149798
A_23_P419714	<i>BTBD11</i>	-0.372279762	0.001837372	-0.382142857	0.001641593
A_23_P78330	<i>CD226</i>	-0.380806929	0.002319418	-0.363304714	0.004037122
A_23_P79134	<i>PLEKHF1</i>	-0.550824405	0.000021141	-0.400142857	0.008364878
A_23_P97889	<i>POLL</i>	-0.310276786	0.001734238	-0.334714286	0.002380174
A_24_P225339	<i>MOB3C</i>	-0.301354365	0.003201662	-0.413523814	0.007398128
A_24_P402222	<i>HLA-DRB3</i>	-0.431244048	0.03894997	-0.157428571	0.540869412
A_24_P66522	<i>AZII</i>	-0.27162664	0.003298114	-0.284780714	0.013684779
A_32_P169754	<i>YBX1</i>	-0.327199405	0.024695353	-0.411285714	0.000861402
A_32_P171124		-0.456479167	0.00054921	-0.661	0.000410713
A_32_P18630	<i>RNF165</i>	-0.432008929	0.000189347	-0.468285714	0.000278816
A_32_P192474	<i>PRRT1</i>	-0.382158333	0.021636629	-0.497885714	0.005647079
A_23_P116091	<i>DPP3</i>	-0.287142857	0.004423671	-0.262142857	0.029302208
A_23_P129738	<i>LLGL2</i>	-0.269744048	0.003647073	-0.237857143	0.06946822
A_23_P144274	<i>JAKMIP1</i>	-0.295416667	0.038497756	-0.188142857	0.154948708
A_23_P150789	<i>PRSS23</i>	-0.528604167	0.003587283	-0.467	0.013548749
A_23_P200001	<i>NEXN</i>	-0.342904762	0.008803982	-0.173	0.224779681
A_23_P212715	<i>CBLB</i>	-0.345559524	0.048647476	-0.295714286	0.010714618
A_23_P354027	<i>KCTD11</i>	-0.315136905	0.006145359	-0.314714286	0.010792703
A_23_P422044	<i>PCDHGA2</i>	-0.351179146	0.005504651	-0.270428571	0.0983871
A_23_P431815	<i>CD7</i>	-0.285452381	0.005750962	-0.261142857	0.029732179
A_23_P6150	<i>FAM182B</i>	-0.268646803	0.003732714	-0.186247614	0.085326362
A_23_P8424	<i>PVRIG</i>	-0.284113095	0.007220516	-0.174857143	0.228831146

A_23_P8913	<i>CA2</i>	-0.540651786	0.007800016	-0.522714286	0.010506026
A_24_P144983	<i>LRRC37BP1</i>	-0.286532738	0.012710858	-0.339	0.006008341
A_24_P205427		-0.445098214	0.000274266	-0.575142857	0.001660414

Table S3. Evaluation of the network performance of the gene-phenotype relationships in the BN in Figure 2A by comparison to the gene-phenotype relationships in the BN for the simulated data with 10% noise level.^a

	total	TP	FP	FN	TN	Sensitivity	Specificity
10%_1	18	10	8	2	1756	0.833333	0.995465
10%_2	12	10	2	2	1762	0.833333	0.998866
10%_3	15	11	4	1	1760	0.916667	0.997732
10%_4	17	7	10	5	1754	0.583333	0.994331
10%_5	15	10	5	2	1759	0.833333	0.997166
10%_6	18	11	7	1	1757	0.916667	0.996032
10%_7	13	9	4	3	1760	0.75	0.997732
10%_8	10	7	3	5	1761	0.583333	0.998299
10%_9	15	8	7	4	1757	0.666667	0.996032
10%_10	17	9	8	3	1756	0.75	0.995465
average						0.766667	0.996712

^aThe total number of possible connections (gene-phenotype) is 1776 with 12 positive connections and 1764 negative connections.

total = total number of gene-phenotype interactions; TP = number of true positives; FP = number of false positives; FN = number of false negatives; TN = number of true negatives.

Table S4. Evaluation of the network performance of the gene-phenotype relationships in the BN in Figure 2A by comparison to the gene-phenotype relationships in the BN for the simulated data with 15% noise level.^a

	total	TP	FP	FN	TN	Sensitivity	Specificity
15%_1	12	7	5	5	1759	0.5833333333	0.997165533
15%_2	11	6	5	6	1759	0.5	0.997165533
15%_3	19	11	8	1	1756	0.916666667	0.995464853
15%_4	16	9	7	3	1757	0.75	0.996031746
15%_5	11	7	4	5	1760	0.5833333333	0.997732426
15%_6	19	9	10	3	1754	0.75	0.994331066
15%_7	16	8	8	4	1756	0.666666667	0.995464853
15%_8	14	10	4	2	1760	0.8333333333	0.997732426
15%_9	19	9	10	3	1754	0.75	0.994331066
15%_10	16	7	9	5	1755	0.5833333333	0.994897959
average						0.691666667	0.996031746

^aThe total number of possible connections (gene-phenotype) is 1776 with 12 positive connections and 1764 negative connections.

total = total number of gene-phenotype interactions; TP = number of true positives; FP = number of false positives; FN = number of false negatives; TN = number of true negatives.

Table S5. Evaluation of the network performance of the gene-phenotype relationships in the BN in Figure 2A by comparison to the gene-phenotype relationships in the BN for the LOOCV data sets.^a

	total	TP	FP	FN	TN	Sensitivity	Specificity
LOOCV1	14	9	5	3	1759	0.75	0.997166
LOOCV2	11	8	3	4	1761	0.666667	0.998299
LOOCV3	17	10	7	2	1757	0.833333	0.996032
LOOCV4	14	9	5	3	1759	0.75	0.997166
LOOCV5	13	8	5	4	1759	0.666667	0.997166
LOOCV6	11	7	4	5	1760	0.583333	0.997732
LOOCV7	12	9	3	3	1761	0.75	0.998299
LOOCV8	14	10	4	2	1760	0.833333	0.997732
LOOCV9	13	9	4	3	1760	0.75	0.997732
LOOCV10	15	9	6	3	1758	0.75	0.996599
LOOCV11	13	10	3	2	1761	0.833333	0.998299
LOOCV12	9	7	2	5	1762	0.583333	0.998866
LOOCV13	10	7	3	5	1761	0.583333	0.998299
LOOCV14	11	9	2	3	1762	0.75	0.998866
LOOCV15	13	11	2	1	1762	0.916667	0.998866
LOOCV16	13	9	4	3	1760	0.75	0.997732
LOOCV17	18	10	8	2	1756	0.833333	0.995465
LOOCV18	15	10	5	2	1759	0.833333	0.997166
LOOCV19	12	10	2	2	1762	0.833333	0.998866
LOOCV20	17	11	6	1	1758	0.916667	0.996599
LOOCV21	10	9	1	3	1763	0.75	0.999433
LOOCV22	11	7	4	5	1760	0.583333	0.997732
LOOCV23	11	8	3	4	1761	0.666667	0.998299

LOOCV24	14	11	3	1	1761	0.916667	0.998299
LOOCV25	11	11	0	1	1764	0.916667	1
LOOCV26	16	10	6	2	1758	0.833333	0.996599
LOOCV27	12	10	2	2	1762	0.833333	0.998866
LOOCV28	11	9	2	3	1762	0.75	0.998866
LOOCV29	16	11	5	1	1759	0.916667	0.997166
LOOCV30	14	7	7	5	1757	0.583333	0.996032
LOOCV31	13	11	2	1	1762	0.916667	0.998866
LOOCV32	12	7	5	5	1759	0.583333	0.997166
LOOCV33	12	9	3	3	1761	0.75	0.998299
LOOCV34	14	8	6	4	1758	0.666667	0.996599
LOOCV35	13	10	3	2	1761	0.833333	0.998299
LOOCV36	12	9	3	3	1761	0.75	0.998299
LOOCV37	16	9	7	3	1757	0.75	0.996032
average						0.761261	0.997392

^aThe total number of possible connections (gene-phenotype) is 1776 with 12 positive connections and 1764 negative connections.

total = total number of gene-phenotype interactions; TP = number of true positives; FP = number of false positives; FN = number of false negatives; TN = number of true negatives.

Figure S1. Volcano plot of all genes showing the negative log₁₀ p-value against the log₂ratios. The red dots indicate the genes with absolute fold change ≥ 1.2 and p-value < 0.05 .

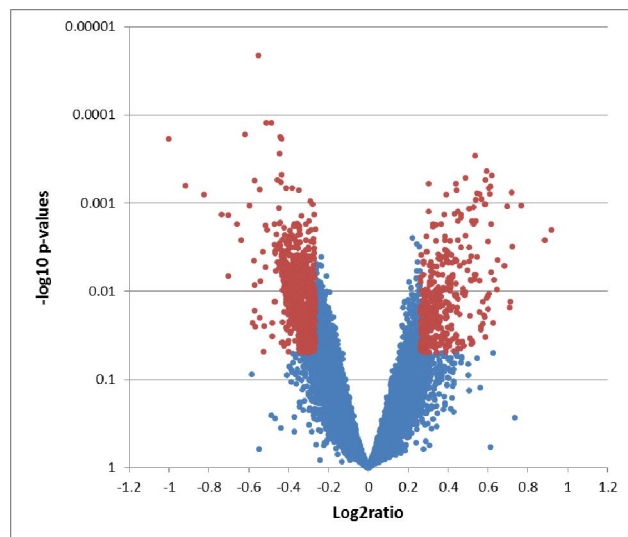


Figure S2. Histogram distribution of log2 ratios of the phenotypic measurements cotinine and DNA-adducts.

