Sensitivity analysis based on the Random Forest machine learning algorithm identifies candidate genes for regulation of innate and adaptive immune response of chicken

Aneta Polewko-Klim [*] , Wojciech Lesiński [*] , Agnieszka Golińska [*],
Krzysztof Mnich [†], Maria Siwek [‡], Witold R Rudnicki [*] [†] [§]

October 24, 2019

ADDITIONAL TABLES

This supplementary tables contain more detailed description of results.

Tables 1, 2, 3 and 4 present detailed results of the cross-validated feature selection procedure. All SNPs that were ever indicated as relevant for any of the traits in the cross-validated feature selection for KLH7, LPS, LTA and KLH0, are presented.

The additional Table 5 contains the summary of performance of the best Random Forest models for 3 phenotypic traits (KLH7, LPS and LTA) measured in 1000 repeats of 3-fold cross-validation.

The detailed results for various combinations of genes used for building predictive models for KLH7 phenotypic trait are given in Tables 6, 7 and 8.

[*]Institute of Informatics, University of Białystok
[†]Computational Centre, University of Bialystok
[‡]Animal Biotechnology Department, University of Technology and Life Sciences
[§]Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw

Table 1: *SNPs that were selected in multiple runs of feature selection procedure for KLH7.*

100%	rs10731333	297	rs14068006	297	rs14072521	297
	rs16102750	297	rs15035880	297	rs15039342	297
	rs15714774	297	rs15725673	297	rs16001483	297
$> 83.3\%$	rs15035854	296	rs14105858	295	rs15965697	292
	rs29005402	291	rs16690726	285	rs15008890	277
	rs14075158	269	rs13508431	265	rs15827424	261
	rs15006760	257	rs15005804	251		
$> 50\%$	rs15810344	241	rs14692425	238	rs15826598	228
	rs14071669	227	rs15826603	223	rs14070244	203
	rs16651464	201	rs14777688	171	rs15820319	161
	rs15947324	151				
$> 10\%$	rs15714740	119	rs13660984	92	rs15943419	92
	rs14110474	77	rs14110445	71	rs15039217	68
	rs13532733	49	rs15723904	46	rs15943775	44
	rs14110239	35				
$< 10\%$	rs14660338	28	rs15821339	28	rs16776013	24
	rs15001183	20	rs14066122	18	rs14071662	9
	rs14075602	7	rs15946187	7	rs15820855	6
	rs15820342	5	rs13507614	3	rs15820324	3
	rs14659246	2	rs15005761	2	rs16651459	2

Table 2: *SNPs that were selected in multiple runs of feature selection procedure for LPS*

100%	rs14074824	297	rs14105858	297	rs14110474	297
	rs15035880	297	rs15039217	297	rs15040786	297
	rs15714774	297	rs15826598	297	rs15965697	297
	rs15968294	297	rs16102750	297	rs16666588	297
$> 85\%$	rs15810344	291	rs10729486	290	rs14075115	290
	rs15012782	283	rs16776013	278	rs16001483	274
	rs15820319	268	rs14777688	264		
$> 50\%$	rs15005804	251	rs13508431	249	rs15827424	244
	rs13507637	228	rs10731333	222	rs13530680	194
	rs14660338	188				
$> 10\%$	rs15946187	146	rs13741184	134	rs15965754	125
	rs14071669	117	rs15826603	109	rs15820342	105
	rs16651257	101	rs14670031	85	rs15820338	81
	rs15732513	77	rs10724273	71	rs16653010	58
	rs14070244	55	rs15005784	37	rs16653011	34
$< 10\%$	rs14659200	25	rs16651592	22	rs15725673	18
	rs16653032	18	rs15726281	17	rs14075155	14
	rs15726279	14	rs15734956	14	rs10731518	8
	rs16122989	8	rs14071662	6	rs15008925	6
	rs15821511	6	rs16651523	6	rs13531794	3
	rs14659025	3	rs14669992	3	rs15035854	3
	rs15943407	3				

Table 3: *SNPs that were selected in multiple runs of feature selection procedure for LTA*

100%	rs10730793	297	rs14070244	297	rs14072943	297
	rs14110239	297	rs15035880	297	rs15039217	297
	rs15714774	297	rs15725673	297	rs15732513	297
	rs15820338	297	rs15821339	297	rs15826598	297
	rs15943775	297	rs15965697	297	rs16102750	297
$> 85\%$	rs15035854	296	rs29005402	296	rs13532735	291
	rs15012782	268	rs14110474	264		
$> 50\%$	rs15810344	249	rs14072521	243	rs13660984	232
	rs15040786	213	rs15005761	205	rs15005804	199
	rs14075141	198	rs15726281	192	rs15946187	168
	rs14105858	151				
$> 10\%$	rs16776013	143	rs14071662	99	rs13530680	98
	rs15008890	92	rs15006760	89	rs16651464	66
	rs14659246	61	rs16666588	59	rs13531794	49
	rs15943921	44	rs16651257	42	rs16690726	38
	rs15005784	29				
$< 10\%$	rs14659025	20	rs15035851	10	rs15826603	9
	rs15820326	7	rs15039342	6	rs14777688	4
	rs13508431	3	rs14072536	3	rs14659217	3
	rs14659238	3	rs14692409	3	rs15005789	3
	rs15943777	3	rs14071669	2	rs15820855	1
	rs16001483	1				

Table 4: *SNPs that were selected in multiple runs of feature selection procedure for KLH0*

100%	rs10724273	297	rs14072536	297	rs14670031	297
	rs15012782	297	rs15035880	297	rs15039217	297
	rs15040786	297	rs15714774	297	rs15943921	297
	rs15965697	297	rs16102750	297		
$> 85\%$	rs13508431	295	rs15827424	291	rs15826603	286
	rs15005789	277	rs13741184	262	rs14105858	260
	rs13507613	259	rs15734956	255	rs15726281	251
	rs16666588	251				
$> 50\%$	rs16651257	235	rs14669992	216	rs10727433	205
	rs16098888	205	rs13507614	195	rs15006760	180
$> 10\%$	rs15723904	146	rs14075143	144	rs14070244	139
	rs15943775	126	rs15723913	125	rs15725673	120
	rs13530680	96	rs13507637	85	rs16651464	83
	rs15734948	75	rs14110445	66	rs16653011	65
	rs16653010	55	rs13660984	49	rs15035854	42
	rs15946187	33	rs15810344	30		
$< 10\%$	rs15820319	28	rs15826598	28	rs14110496	25
	rs15039467	25	rs13531794	22	rs14066122	21
	rs15723899	20	rs15001183	18	rs14659200	13
	rs14074824	12	rs15820338	11	rs14075115	9
	rs15039474	4	rs15968294	3	rs15008890	2

Table 5: *Summary of performance of the best Random Forest models for 3 phenotypic traits measured in 1000 repeats of 3-fold cross-validation.*

Trait	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. dev.
KLH7	5.78%	13.74%	15.03%	14.94%	16.44%	21.81%	2.13%
LPS	-7.93%	3.01%	4.52%	4.38%	6.08%	10.91%	2.44%
LTA	-5.14%	6.24%	8.03%	7.84%	9.55%	15.61%	2.50%

Table 6: *Performance of Random Forest models for KLH7, built on genes from group I plus a reduced subset of group II. The best combination is indicated by bold text.*

Genes	1st Qu.	Median	Mean	3rd Qu.	Std. dev.
gr. I	10.78%	12.44%	12.24%	13.88%	2.31%
gr. I+II	12.86%	14.18%	14.07%	15.53%	2.05%
gr. I+PRCKB	12.40%	13.76%	13.74%	15.31%	2.22%
gr. I+IL9R	12.16%	13.66%	13.56%	15.17%	2.27%
gr. I+MAP2K3	11.41%	13.00%	12.85%	14.43%	2.30%
gr. I+CARD11	10.72%	12.25%	12.08%	13.73%	2.28%
gr. I+ST6GAL1	10.49%	12.16%	11.99%	13.61%	2.37%
gr. I+PTGER4	10.41%	11.97%	11.77%	13.29%	2.29%
gr. I+GPC1	10.01%	11.50%	11.32%	12.81%	2.25%
gr. I+ILR9+PRCKB	**13.74%**	**15.03%**	**14.94%**	**16.44%**	**2.13%**
gr. I+ILR9+MAP2K3	12.49%	14.04%	13.91%	15.49%	2.26%
gr. I+CARD11+ILR9	12.47%	13.83%	13.65%	15.10%	2.15%
gr. I+PRCKB+ST6GAL	12.20%	13.58%	13.53%	15.01%	2.09%
gr. I+MAP2K3+PRCKB	11.81%	13.34%	13.24%	14.69%	2.19%
gr. I+PRCKB+PTGER4	11.85%	13.27%	13.19%	14.72%	2.15%
gr. I+ILR9+ST6GAL	11.67%	13.16%	13.02%	14.51%	2.22%
gr. I+ILR9+PTGER4	11.60%	13.08%	12.95%	14.56%	2.18%
gr. I+MAP2K3+PTGER4	11.41%	12.99%	12.84%	14.51%	2.33%
gr. I+CARD11+PRCKB	11.49%	12.99%	12.85%	14.37%	2.20%
gr. I+MAP2K3+ST6GAL1	11.45%	12.88%	12.81%	14.26%	2.15%
gr. I+CARD11+PTGER4	11.27%	12.78%	12.64%	14.15%	2.16%
gr. I+ILR9+GPC1	11.07%	12.58%	12.44%	14.03%	2.16%
gr. I+CARD11+ST6GAL1	11.26%	12.56%	12.40%	13.82%	2.12%
gr. I+GPC1+PRCKB	10.83%	12.41%	12.18%	13.81%	2.32%
gr. I+MAP2K3+GPC1	10.96%	12.40%	12.25%	13.77%	2.28%
gr. I+CARD11+GPC1	10.60%	12.35%	12.07%	13.69%	2.23%
gr. I+CARD11+MAP2K3	10.32%	11.94%	11.83%	13.38%	2.27%
gr. I+ST6GAL1+PTGER4	10.14%	11.78%	11.60%	13.10%	2.30%
gr. I+GPC1+PTGER4	9.95%	11.44%	11.38%	13.01%	2.34%
gr. I+GPC1+ST6GAL	9.62%	11.03%	10.93%	12.45%	2.15%

Table 7: *Performance of Random Forest models for KLH7, built using genes from group I*

Genes	1st Qu.	Median	Mean	3rd Qu.	Std. dev.
gr. I	10.78%	12.44%	12.24%	13.88%	2.31%
gr. I−UNC13D	7.91%	9.50%	9.41%	11.11%	2.42%
gr. I−CRLF3	4.82%	6.64%	6.50%	8.29%	2.52%
gr. I−MAPK8IP3	6.32%	7.58%	7.42%	8.84%	1.92%

Table 8: *Performance of Random Forest models for KLH7, built using various combinations of groups of genes*

Group	1st Qu.	Median	Mean	3rd Qu.	Std. dev.
gr. I+II+III	10.12%	11.31%	11.20%	12.44%	1.81%
gr. I	10.78%	12.44%	12.24%	13.88%	2.31%
gr. I+II	12.86%	14.18%	14.07%	15.53%	2.05%
gr. I+III	7.82%	9.22%	9.13%	10.66%	2.15%
gr. II	-1.19%	0.73%	0.61%	2.63%	2.81%
gr. III	-2.23%	-0.57%	-0.72%	1.05%	2.48%
gr. II+III	3.10%	4.48%	4.37%	5.90%	2.13%