Separators and decision trees

Andreas Hoppe

September 16, 2014

Contents

1 Separators

This section is about factor which have different values in a certain time frame. A time frame is a consecutive series of time points, it may or not start at 0h, may or may not end at 14d, and may also consist of only a single time point. A separator is a combination of a factor, a threshold value and a time frame. If the factor's value of a putative mouse above (or below, for separators marked as reverse in the tbale below) the threshold it is predicted that the time after BDL is in the specified time frame. The prediction can be tested on a set of mice with known, and the ratio of correct predictions is called recall.

There are 41 perfect separators of a time frame, i.e. a factor, whose values are on opposites sides of a threshold for mice inside versus outside the time frame. Perfect means, that the time frame is recalled with 100%. 35 of the time frames include the beginning/end of the whole experiment, i.e. values change (mostly up) at a certain time point and remain until 14d.

12 separators predict a single time point. 6 separators of which separate the 0h time point (the control) from the treated mice. The 6 remaining separators time frames are of particular interest, Nr0b2_RNAa is strongly decreases only for all mice in the 6h timepoint, the threshold value leading to a perfect prediction is 162.

The quality of a perfect separator, the separation, is measured by two parameters: (i) the actual gap between the values of the split time frames in relation to the whole factor variation, and (ii) the differences of the averages in relation to the standard deviation inside each of the time frame's data. Separators of a higher separation have a better chance to perform well for mice not included in the training data set.

Cyp24a1_RNAa is also a perfect separator for the 6h time point as Nr0b2_RNAa, but the separation is lower. So for a particular prediction the separators with a high separation are preferred.

Not for all time point there is a single separator (Mmp10_RNAf is a perfect separator for the 18h time point, the only other internal time point with such a separator). However, by the combination of two or three factors, each time point can be predicted, see below.

The factor CTGF_cells appears as a perfect separator for several time frames (0-30h, 0-5d, 0-12h, 0-6h) the first of which is also the topmost separating separator. Thus, it is the best candidate to monitor the disease progress.

Among the RNA, Il28b_RNAz, Col3a1_RNAf, Sparc_RNAf, Il13_RNAz, Pdgfb_RNAf, Tgfb2_RNAf (time frame 0h-2d) as well as Cyp1a2_RNAa (time frame 0h-6h) are factors showing a high separation. With respect to the transcriptional changes there is a large change from the 2d and 5d time point.

factor	time frame start	time frame end	reverse	separation range percentage	separator value by median of gap	minimum of lower value set	maximum of lower value set	minimum of higher value set	maximum of higher value set
CTGF_cells	0h	30h	0	18.6	28	1	20	36	87
S100A4_cells	0h	30h	0	16.3	36	12	32	40	61
Il28b_RNAz	0h	2d	0	15.6	12865	0.002	122.2	25608	$1.6 \cdot 10^5$
CTGF_cells	0h	5d	0	15.1	72.5	1	66	79	87
Bili_blood	0h	2d	0	12	168.4	0.53	144.2	192.5	401.5
Col3a1_RNAf	0h	2d	0	11.2	4.64	0.05	3.38	5.89	22.4
Sparc_RNAf	0h	2d	0	10.4	2.07	0.24	1.73	2.41	6.8
Nr0b2_RNAa	6h	6h	1	9.55	161.9	0	130.6	193.2	655.5
Cyp1a2_RNAa	12h	14d	0	9.34	1.89	0.39	1.76	2.02	3.19
Il13_RNAz	0h	2d	0	6.36	1915	0.001	240.5	3590	52699
Pdgfb_RNAf	0h	2d	0	5.48	1.74	0.37	1.61	1.87	5.13
ALT_blood	0h	0h	0	5.34	63.1	26.5	33.2	93	1146
Tgfb2_RNAf	0h	2d	0	5.11	1.79	0.29	1.54	2.04	10.1
Il17a_RNAz	0h	2d	0	5.01	14034	0.01	0.04	28068	$5.6 \cdot 10^5$
SMA_cells	0h	6h	0	4.92	9	1	6	12	123
Cyp2c37_RNAa	5d	14d	0	4.78	0.79	0.13	0.72	0.86	3.18
CTGF_cells	0h	12h	0	4.65	12	1	10	14	87
CTGF_cells	0h	6h	0	3.49	4.5	1	3	6	87
Cyp24a1_RNAa	6h	6h	1	3.16	67.7	0.02	9.76	125.7	3663
Cd86_RNAz	0h	2d	0	2.7	1.59	0.26	1.48	1.69	7.91
Col1a1_RNAf	0h	2d	0	2.61	5.38	0.07	5.1	5.67	21.6
Ctgf_RNAf	0h	5d	0	2.55	3.46	0.23	3.34	3.57	9.42
S100A4_cells	0h	12h	0	2.04	19.5	12	19	20	61
Bili_blood	0h	0h	0	1.82	5.31	0.53	1.65	8.96	401.5
Cyp2e1_RNAf	5d	14d	0	1.74	0.7	0.26	0.69	0.72	1.99
Hmox1_RNAa	0h	0h	0	1.71	0.57	0.27	0.55	0.59	2.49
SMA_cells	0h	12h	0	1.64	16	1	15	17	123
GLDH_blood	0h	0h	0	1.25	44.3	8.4	20.6	68	3794
Cd14_RNAz	0h	6h	0	0.83	0.39	0.06	0.37	0.42	6.16
SMA_cells	0h	0h	0	0.82	2.5	1	2	3	123
Cxcl2_RNAz	0h	0h	0	0.76	0.13	0.01	0.09	0.18	11.7
Mmp10_RNAf	18h	18h	1	0.59	57.4	0.001	42.8	72	4971
Hgf_RNAz	0h	2d	0	0.55	1.79	0.29	1.77	1.81	7.75
Cdh2_RNAf	6h	12h	0	0.5	0.9	0.51	0.89	0.9	3.2
Cd14_RNAz	0h	0h	0	0.37	0.18	0.06	0.17	0.19	6.16
Il10rb_RNAz	0h	2d	0	0.35	1.37	0.35	1.36	1.38	5
Ccl2_RNAz	0h	6h	0	0.15	0.28	0.05	0.27	0.29	13.3
Timp1_RNAf	0h	0h	0	0.13	0.12	0.06	0.11	0.13	11.7
Il28b_RNAz	6h	2d	0	0.01	9.17	0.002	0.01	18.3	$1.6 \cdot 10^5$
Il2_RNAz	18h	2d	0	0.002	0.12	0.0003	0.001	0.23	14869
Mmp10_RNAf	0h	0h	0	0	0.002	0.001	0.002	0.002	4971

Table 1 – list of perfect separators sorted by the relative gap

Split	question	relative gap (%)
Split between 30h and 2d	e(CTGF_cells)<28 <=> before 2d	18.6047
Split between 12h and 18h	e(CTGF_cells)<12 <=> before 18h	21.0526
Split between 6h and 12h	e(SMA_cells)<9 <=> before 12h	42.8571
Split between 0h and 6h	e(Tnfrsf1a_RNAz)<0.80662 <=> before 6h	37.6697
Split between 18h and 30h	e(Gstm1_RNAa)<1.38529 <=> before 30h	13.8217
Split between 5d and 14d	e(CTGF_cells)<72.5 <=> before 14d	25.4902
Split between 2d and 5d	e(Il28b_RNAz)<21392.5 <=> before 5d	26.2641
24/24 correct, 0/24 wrong predictions; 17 unapplicable.		

Table 2 – decision tree for time points, all mouse data. The last line refers to the check on all 41 mice, whether the decision tree predicts the correct time point. Tests are referred to as unapplicable if the decision tree contains a factor not measured for the tested mouse, in this case, CTGF_cells, which is only measured for three repeats.

2 Decision tree for time points

This section is about predicting the time point a mouse belongs with binary decision trees. A decision tree is a scheme consisting of questions of the form "Is factor X larger than a value y?" When several of these questions have been answered, the automated predictor decides to which time point the mouse belongs.

Only those factors are considered where the values one time frame are in a disjoint interval compared to the values of another time frame. Between the intervals there is a gap, where the size of gap, compared with the range of all values of both intervals — the relative gap — measures the fitness of the factor to tell one time point from another. If several factors are candidates the one with the larger relative gap is perferred. The median of the gap is the suggested splitting point between the two time points.

For the decision tree generated from all mice, see Table 2. For the decision trees of the leave-one-out tests, see Table 3.

Question 1			Question 2			Question 3			Question 4			Question 5			Question 6			Question 7			result
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Col8a1>0.83	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	Il28b>16.6	0h	6h	Nr0b2>97.5	6h	12h	CTGF<12	12h	18h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	Fn1<0.79	0h	6h	Nr0b2>97.5	6h	12h	CTGF<12	12h	18h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	SMA<3.5	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	wrong
CTGF<28	30h	2d	Fn1<0.83	0h	6h	Nr0b2>97.5	6h	12h	CTGF<12	12h	18h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<8.5	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<4	6h	12h	GLDH<268.3	0h	6h	CTGF<12	12h	18h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	ALT<221.6	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<11.5	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<6	6h	12h	Tnfrsf1a<0.81	0h	6h	CTGF<12	12h	18h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	Gstm1<1.29	18h	30h	CTGF<13	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Fn1>1.23	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	wrong
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Cd86<0.61	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	wrong
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Cyp1a2>0.92	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	S100A4<27	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	wrong
CTGF<27.5	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Mki67<0.88	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	ALT<138.1	0h	6h	Nr0b2>97.5	6h	12h	BrdU_NP<0.43	18h	30h	CTGF<13	12h	18h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
S100A4<34.5	30h	2d	CTGF<13	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	SMA<39.5	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	wrong
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Birc5<0.71	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<30.5	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	Gdf2<1.44	2d	5d	CTGF<72.5	5d	14d	wrong
Bili<150.7	2d	5d	CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	correct
S100A4<38	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	BrdU_Stellate<0.97	2d	5d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
Il10rb<2.1	2d	5d	CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	n.appl.
CTGF<66	5d	14d	BrdU_Stellate<1.14	2d	5d	CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	wrong
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Bili<192.6	2d	5d	wrong
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	CTGF<47.5	2d	5d	wrong
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
CTGF<75.5	5d	14d	Il28b<21454	2d	5d	CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	correct
BrdU_Stellate<0.97	2d	5d	CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	n.appl.
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	correct
CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	Il28b<21393	2d	5d	n.appl.
Il28b<18045	2d	5d	CTGF<28	30h	2d	CTGF<12	12h	18h	SMA<9	6h	12h	Tnfrsf1a<0.81	0h	6h	Gstm1<1.39	18h	30h	CTGF<72.5	5d	14d	correct

15/24 correct, 9/24 wrong predictions; 17 unapplicable.

Table 3 – decision trees of leave-one-out tests for time points. The last line refers to the check on all 41 mice, whether the decision tree predicts the correct time point. Tests are referred to as unapplicable if the decision tree contains a factor not measured for the tested mouse, in this case, CTGF_cells, which is only measured for three repeats.

3 Decision trees for phases

Here, only the phases are to be predicted, i.e. 4 classes of mice: control, initial (6-12h), perpetuation (18h-2d), and progression (5-14d). Thus, three questions are sufficient.

3.1 All factors

First, there is no restriction on the selected factors For the decision tree generated from all mice, see Table 4. As you can see, Il28b, Fn1, and CTGF cells are used as predictors. As the first decision is made for Il28b, a factor measured on all repeats, there are only 11 unapplicable mice.

For the decision trees of the leave-one-out tests, see Table 5. The sole wrong prediction is mouse where the Il10rb value is an outlier, resulting Il10rb chosen as the root question.

Split	question	relative gap (%)
Split between perpetuation and progression	e(Il28b_RNAz)<12864.9 <=> before progression	15.6446
Split between control and initial	e(Fn1_RNAf)<0.78008 <=> before initial	17.4244
Split between initial and perpetuation	e(CTGF_cells)<12 <=> before perpetuation	9.7561
30/30 correct, 0/30 wrong predictions; 11 unapplicable.		

Table 4 – decision tree for phases, all mouse data.

Question 1			Question 2			Question 3			result
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Il28b>16.6	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.77	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12855	perpetuation	progression	Il28b>9.17	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<11.5	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<13	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Bili<150.7	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.8	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il10rb<2.1	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	wrong
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Bili<177.1	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
BrdU_Stellate<0.97	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
Il28b<18045	perpetuation	progression	Fn1<0.78	control	initial	CTGF<12	initial	perpetuation	correct
28/29 correct, 1/29 wrong predictions; 12 unapplicable.									

Table 5 – decision trees of leave-one-out tests for phases.

3.2 Excluding all factors with only 3 repeats

Here the cell counts of CTGF, α-SMA, and S100a4, the factors with only 3 repeats, are excluded. For the decision tree generated from all mice, see Table 6. The left out mouse can be ignored, it is a repeat experiment where only few factors are measured, and not the Fluidigm qPCR runs. Not surprisingly, Il28b and Fn1 are again chosen as questions, and CTGF is replaced with Cdh2, encoding Cadherin.

For the decision trees of the leave-one-out tests, see Table 7. One wrong predictions comes from the same mentioned above, the Il10rb is chosen as the root predictor. Another wrong prediction comes from an mouse where Cyp1a2 has an outlier values, and is chosen as a predictor. The third wrong prediction is very close. The spearation value for Cdh2 is set to 0.88 instead 0.9, just enough to let the decision tree fail.

Split	question	relative gap (%)
Split between perpetuation and progression	e(Il28b_RNAz)<12864.9 <=> before progression	15.6446
Split between control and initial	e(Fn1_RNAf)<0.78008 <=> before initial	17.4244
Split between initial and perpetuation	e(Cdh2_RNAf)<0.899067 <=> before perpetuation	0.774677
40/40 correct, 0/40 wrong predictions; 1 unapplicable.		

Table 6 – decision tree for phases excluding all factors with only 3 repeats, all mouse data.

Question 1			Question 2			Question 3			result
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Il28b>16.6	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.77	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12855	perpetuation	progression	Il28b>9.17	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.88	initial	perpetuation	wrong
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cyp1a2>1.17	initial	perpetuation	wrong
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.91	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Bili<150.7	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.8	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il10rb<2.1	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	wrong
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Bili<177.1	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
Il28b<18045	perpetuation	progression	Fn1<0.78	control	initial	Cdh2<0.9	initial	perpetuation	correct
37/40 correct, 3/40 wrong predictions; 1 unapplicable.									

Table 7 – decision trees of leave-one-out tests for phases excluding all factors with only 3 repeats.

3.3 Excluding factors which can not be measured in blood plasma

Additionally to the cell counts of CTGF, α-SMA, and S100a4, further factors are excluded, which presumably can not be found in blood plasma.

For the decision tree generated from all mice, see Table 8. Here, Cdh2 is replaced with Il2, and the interleukin is excreted from cells and also found in blood plasma.

Split	question	relative gap (%)
Split between perpetuation and progression	e(Il28b_RNAz)<12864.9 <=> before progression	15.6446
Split between control and initial	e(Fn1_RNAf)<0.78008 <=> before initial	17.4244
Split between initial and perpetuation	e(Il2_RNAz)>0.115964 <=> before perpetuation	0.447211
40/40 correct, 0/40 wrong predictions; 1 unapplicable.		

Table 8 – decision tree for phases excluding all intracellular factors, all mouse data.

For the decision trees of the leave-one-out tests, see Table 9. Apart from Il10rb, already discussed above, also Bilirubin is used as tree question, and led to a wrong predicton in one case.

Question 1			Question 2			Question 3			result
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Il28b>16.6	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.77	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12855	perpetuation	progression	Il28b>9.17	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.15	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Bili<31.9	initial	perpetuation	wrong
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Bili<150.7	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.8	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il10rb<2.1	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	wrong
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Bili<177.1	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	n.appl.
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<12865	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct
Il28b<18045	perpetuation	progression	Fn1<0.78	control	initial	Il2>0.12	initial	perpetuation	correct

38/40 correct, 2/40 wrong predictions; 1 unapplicable.

Table 9 – decision trees of leave-one-out tests for phases excluding all intracellular factors.