Supplementary Information: Statistical analysis

In the shorthand formula notation for linear mixed-effects models [1] the models used were:

1. OAscore ~ Age * Genotype for the beta regression with rstanarm function stan_betareg. In this model Age was a continuous variable (unit = 1 month).
2. log OAscore ~ Genotype + (1|Litter) with rstanarm function stan_glmer. In this case the OA scores were far from the boundaries 0 and 6, and their logarithms approximately normally distributed so that we could use the above generalized linear mixed model with a Gaussian or t-distributed noise.
3. log CartilageThickness ~ Age * Genotype + (1|Litter) for use with rstanarm function stan_glmer.
4. ΔCt ~ Primer -1 + RetinoicAcid + Genotype + (1|Animal) with ΔCt the qPCR response values, primer the type of protease primer, RetinoicAcid for treatment with retinoic acid (yes or no), Genotype (wt vs mutant), and Animal for ID of animal.
5. log AggrecanDigestionSignal ~ Genotype + SampleType + (1|Litter) with SampleType being one of Medium or GuHCL and using the rstanarm function stan_glmer.
6. log ZymogramSignal ~ Genotype * RetinoicAcid + (1|Litter) with stan model as given below in section Priors.

Inference of parameter value probabilities was carried out by Bayesian analysis using the probabilistic programming language Stan [2], through the R-package rstanarm [3], version 2.18.2, as interface. Priors proposed by rstanarm functions stan_glmer and stan_betareg, respectively, were used throughout. Priors are listed below.

For each model, Stan generated 4 Markov chains, each of 2000 iterations (1000 for warm-up, 1000 for sampling). Gelman-Rubin $\hat{R}$ values were found in all cases to be less than 1.05, indicating convergence of Markov chains. For all models, posterior predictive checks showed good agreement of measured responses and responses simulated with the respective model. Results of Bayesian analysis are typically reported as means of marginal posteriors of parameters of interest, for instance a mean change of OARSI score per month of ageing. As a measure of uncertainty of these means, we report them together with their respective 95% HDI, i.e. intervals of 95% highest density of marginal posterior probability; in other words, the inferred means are within the 95% HDI with a probability of 0.95.

Priors

For the intercepts and slopes ("coefficients") of the models, the default prior ("specified prior") is given together with the prior after adjusting for the actual data ("adjusted prior"). The adjusted priors are weakly informative, i.e. they are much broader than the distribution of the actual data, but not too broad, so that exploration of nonsensical parameter regions is avoided. Priors for each of the models:

1 OAscore ~ Age * Genotype

Experiment: Ageing of Col2-rtTA-Cre;Ext1$^{e2fl/e2fl}$ mice, scoring at 3, 6, 12 and 18 months

Figure: 1B and D

Intercept (after predictors centered)

~ normal(location = 0, scale = 10)

Coefficients

Specified prior:

~ normal(location = [0,0,0], scale = [2.5,2.5,2.5])

Adjusted prior:

~ normal(location = [0,0,0], scale = [2.50,0.40,0.39])

Auxiliary (phi)

~ exponential(rate = 1)

2 log OAscore ~ Genotype + (1|Litter)

2.1 Experiment: Scoring after surgical induction of OA in Col2-rtTA-Cre;Ext1$^{e2fl/e2fl}$ mice

Figure: 1C

Intercept (after predictors centered)

Specified prior:

~ normal(location = 0, scale = 10)

Adjusted prior:

~ normal(location = 0, scale = 1.7)

Coefficients

Specified prior:

~ normal(location = 0, scale = 2.5)

Adjusted prior:

~ normal(location = 0, scale = 0.44)

Auxiliary (sigma)

Specified prior:

~ exponential(rate = 1)

Adjusted prior:

~ exponential(rate = 5.7)

Covariance

~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)

2.2 Experiment: Scoring after surgical induction of OA in Ndst1$^{+/-}$ mice

Figure: 2A and B

Intercept (after predictors centered)

Specified prior:

~ normal(location = 0, scale = 10)

Adjusted prior:

~ normal(location = 0, scale = 1.9)

Coefficients

Specified prior:

~ normal(location = 0, scale = 2.5)

Adjusted prior:

~ normal(location = 0, scale = 0.48)

Auxiliary (sigma)

Specified prior:

~ exponential(rate = 1)

Adjusted prior:

~ exponential(rate = 5.2)

Covariance

~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)

2.3 Experiment: Scoring after surgical induction of OA in Col2-Cre;Ndst1$^{fl/fl}$ mice

Figure: 2A and B

Intercept (after predictors centered)

Specified prior:

~ normal(location = 0, scale = 10)

Adjusted prior:

~ normal(location = 0, scale = 1.7)

Coefficients

Specified prior:

~ normal(location = 0, scale = 2.5)

Adjusted prior:

~ normal(location = 0, scale = 0.43)

Auxiliary (sigma)

Specified prior:

~ exponential(rate = 1)

Adjusted prior:

~ exponential(rate = 5.9)

Covariance

~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)

3 log CartilageThickness ~ Age * Genotype + (1|Litter)

Experiment: Quantification of articular cartilage thickness in Col2-Cre;Ndst1$^{fl/fl}$ mice, analysis at 1, 3 and 18m

Figure: 2C and D

Intercept (after predictors centered)

Specified prior:

~ normal(location = 0, scale = 10)

Adjusted prior:

~ normal(location = 0, scale = 1.2)

Coefficients

Specified prior:

~ normal(location = [0,0,0,...], scale = [2.5,2.5,2.5,...])

Adjusted prior:

~ normal(location = [0,0,0,...], scale = [0.30,0.30,0.30,...])

Auxiliary (sigma)

Specified prior:

~ exponential(rate = 1)

Adjusted prior:

~ exponential(rate = 8.4)

Covariance

~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)

4 ΔCt ~ Primer -1 + RetinoicAcid + Genotype + (1|Animal)

Experiment: Gene expression analysis by qRT-PCR

Figure: 3A and B

Coefficients

Specified prior:

~ normal(location = [0,0,0,...], scale = [2.5,2.5,2.5,...])

Adjusted prior:

~ normal(location = [0,0,0,...], scale = [4.29,4.29,4.29,...])

Auxiliary (sigma)

Specified prior:

~ exponential(rate = 1)

Adjusted prior:

~ exponential(rate = 0.58)

Covariance

~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)

5 log AggrecanDigestionSignal ~ Genotype + SampleType + (1|Litter)

5.1 Experiment: Detection of NITEGE neo-epitope by Western Blot

Figure: 4A and B

Intercept (after predictors centered)

Specified prior:

~ normal(location = 0, scale = 10)

Adjusted prior:

~ normal(location = 0, scale = 18)

Coefficients

Specified prior:

~ normal(location = [0,0], scale = [2.5,2.5])

Adjusted prior:

~ normal(location = [0,0], scale = [4.53,4.53])

Auxiliary (sigma)

Specified prior:

~ exponential(rate = 1)

Adjusted prior:

~ exponential(rate = 0.55)

Covariance

~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)

5.2 Experiment: Detection of VDIPEN neo-epitopes by Western Blot

Figure: 4A and B

Intercept (after predictors centered)

Specified prior:

~ normal(location = 0, scale = 10)

Adjusted prior:

~ normal(location = 0, scale = 9.3)

Coefficients

Specified prior:

~ normal(location = [0,0], scale = [2.5,2.5])

Adjusted prior:

~ normal(location = [0,0], scale = [2.33,2.33])

Auxiliary (sigma)

Specified prior:

~ exponential(rate = 1)

Adjusted prior:

~ exponential(rate = 1.1)

Covariance

~ decov(reg. = 1, conc. = 1, shape = 1, scale = 1)

6 log ZymogramSignal ~ Genotype*RetinoicAcid + (1|Litter)

Experiment: Analysis of protease activity by gelatine zymography

Figure: 4C and D

For technical reasons, this model could not be implemented in rstanarm. To allow for the modeling of outliers, we had to assume a Student-t distributed noise. We therefore developed the stan program given here:

```
// start of stan code
data {
  int<lower=1> N; // number of measurements
  real y[N]; // log zymogram signal
  int<lower=1> n_litters; // number of litters
  int<lower=1> litter[N]; // litter of each measurement
  int<lower=-1, upper=1> RA[N]; // retinonic acid (RA) treatment
  int<lower=-1, upper=1> gt[N]; // genotype
}

parameters {
  real a_0; // overall intercept
  real a_litter[n_litters]; // litter specific intercept
  real beta_RA; // slope for RA treatment
  real beta_gt; // slope for genotype
  real beta_RA_gt; // slope for interaction of RA and genotype
```

```
  real<lower=0.0> sigma[n_litters]; // standard deviation (sd) of litter signal
  real<lower=0.0> sigma_a; // sd of litter level intercepts
  real<lower=0.0> sigma_l;
  real<lower=0.0> nu;
}

transformed parameters {
  vector[N] mu; // mean y

  for (i in 1:N){ // linear model with interaction
    mu[i] = a_0 +
      a_litter[litter[i]] +
      beta_RA * RA[i] +
      beta_gt * gt[i] +
      beta_RA_gt * RA[i] * gt[i];
  }
}

model {

  // priors
  a_0 ~ normal(0.0, 10.0);
  beta_RA ~ normal(0.0, 2.0);
  beta_gt ~ normal(0.0, 2.0);
  beta_RA_gt ~ normal(0.0, 2.0);
  sigma_a ~ exponential(1);
```

```
  sigma_l ~ exponential(1);
  nu ~ exponential(0.1);

  // hierarchical level of litters
  for (i in 1:n_litters) {
    sigma[i] ~ exponential(sigma_l);
    a_litter[i] ~ normal(0, sigma_a);
  }

  // Student-t noise
  for (i in 1:N) {
    y[i] ~ student_t(nu, mu[i], sigma[litter[i]]);
  }
}

generated quantities { // for PPC and leave-one-ont crossvalidation
  real y_rep[N];
  vector[N] log_lik;

  for (i in 1:N){
    y_rep[i] = student_t_rng(nu, mu[i], sigma[litter[i]]);
    log_lik[i] = student_t_lpdf(y[i] | nu, mu[i], sigma[litter[i]]);
  }
}
// end of stan code
```

References

1. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 2015; 67: 1-48.

2. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software 2017; 76.

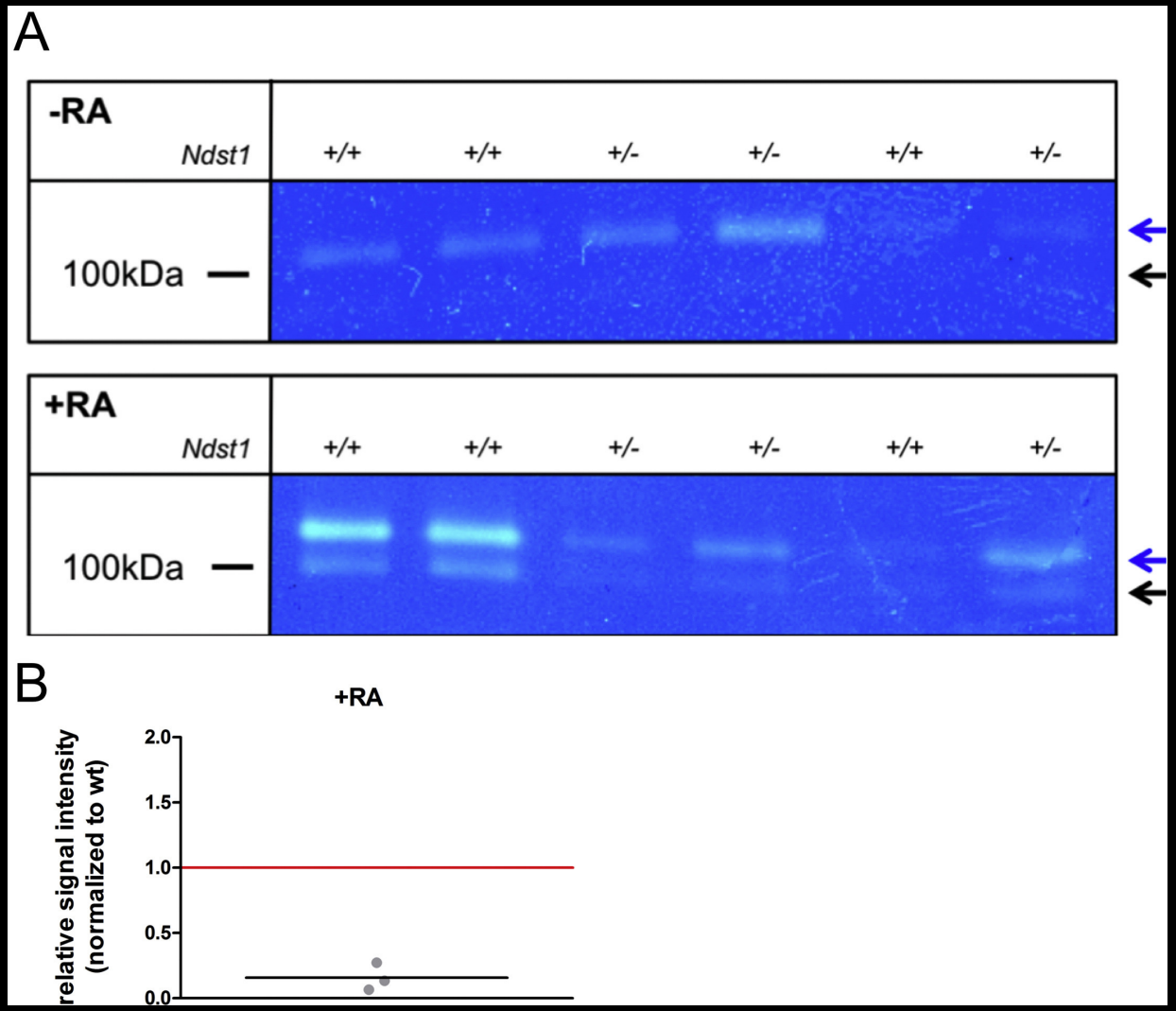
3. Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression modeling via Stan. R package version 2.17.4. 2018.

An altered heparan sulfate structure in the articular cartilage protects against osteoarthritis

A.-C. Severmann, K. Jochmann, K. Feller, V. Bachvarova, V. Piombo, R. Stange, T. Holzer, B. Brachvogel, J. Esko, T. Pap, D. Hoffmann, A. Vortkamp

Osteoarthritis and Cartilage
Volume 28 Issue 7 Pages 977-987 (July 2020)
DOI: 10.1016/j.joca.2020.04.002

Supplementary Fig. 1



Osteoarthritis and Cartilage 2020 28977-987DOI: (10.1016/j.joca.2020.04.002)

Supplementary Table 3

	Figure	Parameter	Effect size	95% HDI	
Ageing *Col2-Cre;Ext1$^{e2fl/e2fl}$*	1B	Genotype	Mean OA-score 0.49 (M/WT)	-0.27	1.25
		Age	Mean OA-score 0.13 (M/WT)	0.08	0.18
		Genotype and Age	Mean OA-score -0.09 (M/WT)	-0.16	-0.03
ACLT surgery *Col2-Cre;Ext1$^{e2fl/e2fl}$*	1C	Genotype	Median factor 0.83 (M/WT)	0.72	0.96
ACLT surgery *Ndst1*	2B	Genotype	Median factor 0.82 (M/WT)	0.74	0.90
ACLT surgery *Col2-Cre;Ndst1$^{fl/fl}$*	2B	Genotype	Median factor 0.87 (M/WT)	0.76	1.00
Cartilage thickness *Col2-Cre;Ndst1$^{fl/fl}$*	2D	Genotype	Mean factor 1.24 (M/WT)	1.15	1.33
Expression *Ndst1*	3A	Genotype	Mean effect		
	Adamts4		0.39	-1.09	2.18
	Adamts5		0.05	-1.14	1.28
	Mmp2		-0.01	-1.24	1.24
	Mmp3		0.02	-1.23	1.36
	Mmp9		0.21	-0.97	1.61
	Mmp13		-0.11	-1.54	1.23
Expression *Ndst1*	3B	Genotype	Mean effect		
	Timp1		0.04	-0.78	1.04
	Timp2		0.11	-0.75	1.12
	Timp3		0.11	-0.62	1.28
Aggrecan degradation +RA	4A NITEGE	Genotype	Mean increase 4.24 (WT/M)	1.05	18.55
		Sample type	Mean increase 2.14 (WT/M)	0.52	8.75
	VDIPEN	Genotype	Mean increase 1.54 (WT/M)	1.00	2.34
		Sample type	Mean increase 1.63 (WT/M)	1.09	2.45
Zymography -RA	4B	Genotype	Median factor 1.09 (M/WT)	0.90	1.34
+RA		Genotype	Median factor 0.77 (M/WT)	0.60	0.96
Zymography +RA	Sup. 2	Genotype	Mean factor 0.69 (M/WT)	0.44	1.14

Supplementary Table 4

Figure			
	Ext1$^{e2fl/e2fl}$	*Col2-rtTA-Cre;Ext1*$^{e2fl/e2fl}$	
1B/D	3M: 11 6M: 6 12M: 4 18M: 6	3M: 11 6M: 6 12M: 2 18M: 8	Individuals
	Ext1$^{e2fl/e2fl}$	*Col2-rtTA-Cre;Ext1*$^{e2fl/e2fl}$	
1C	Sham: 8 ACLT: 11	Sham: 7 ACLT: 13	Individuals, from 4 litters
	Ndst1$^{+/+}$	*Ndst1*$^{+/-}$	
2B	Sham: 6 ACLT: 12	Sham: 5 ACLT: 9	Individuals, from 3 litters
	Ndst1$^{fl/fl}$	*Col2-Cre;Ndst1*$^{fl/fl}$	
	Sham: 6 ACLT: 12	Sham: 6 ACLT: 9	Individuals, from 3 litters
	Ndst1$^{fl/fl}$	*Col2-Cre;Ndst1*$^{fl/fl}$	
2D	 1m: 6 3m: 6 18m: 7	 1m: 5 3m: 7 18m: 6	Individuals, from 1m: 2 litters 3m: 3 litters 18m: 4 litters
	Ndst1$^{+/+}$	*Ndst1*$^{+/-}$	
3A, B Litter 1: Litter 2: Litter 3: Litter 4:	 1 pool (2 mice) 1 pool (4 mice) 1 pool (3 mice) 1 pool (3 mice)	 1 pool (2 mice) 1 pool (3 mice) 1 pool (3 mice) 1 pool (4 mice)	Used samples: Adamts4, 5: litter 1-4 Mmp2: litter 1-3 Mmp3, 9, 13: litter 1-4 Timp 1, 2, 3: litter 1-3
	Ndst1$^{+/+}$	*Ndst1*$^{+/-}$	
4B Litter 1: Litter 2: Litter 3:	 1 pool (7 mice) 2 pools (4 mice each) 2 pools (3 mice each)	 1 pool (3 mice) 1 pool (2 mice) 1 pool (2 mice)	Used samples NITEGE: litter 1-3 VDIPEN: litter 1-3
	Ndst1$^{+/+}$	*Ndst1*$^{+/-}$	
4D	17	17	Individuals (from 6 litters)
	Ndst1$^{+/+}$	*Ndst1*$^{+/-}$	
S2	3	3	Individuals (from 1 litter)

Supplementary Table 2

	Primer 1	Primer 2
Adamts4	gagtcccatttcccgcagaa	ataaccgtcagcaggtagcg
Adamts5	gggcacaggctactatgtgg	gccgtcacatccagttctca
Mmp2	cgcgtaaagtatgggaacgc	ggtaaacaaggcttcatgggg
Mmp3	gtctccctgcaaccgtgaag	acccttgagtcaacacctgg
Mmp9	ccagccgacttttgtggtct	tggcctttagtgtctggctg
Mmp13	gccattaccagtctccgagg	ggtcacgggatggatgttca
Timp1	cttcttggttccctggcgta	ggacctgatccgtccacaaa
Timp2	gctggacgttggaggaaaga	gacagcgagtgatcttgcac
Timp3	ccgacatcgtgatccggg	cacgtggggcatcttactga

Supplementary Table 1

	Primer 1	Primer 2	Primer 3	Annealing Temperature [°C]
Ext1$^{e2fl/e2fl}$	gagtccatcctgctctgcat	ttgttgcatgggaaagacaa		62
Col2-rtTA-Cre	gagtgatgaggttcgcaaga	ctacaccagagacgg		55
R26R-LacZ	aagaccgcgaagagtttgtc	aaagtcgctctgagttgttat	ggagcgggagaaatggatatg	58
Ndst1 *Ndst1*$^{fl/fl}$	catcctctgaggtgaccgc	ccagggcgtcagggcctcctg	cccagatggcgagactgagg	60
Col2-Cre	gagtgatgaggttcgcaaga	ctacaccagagacgg		55