Supplemental Material

Tables S1-S5 (pages 10-14), then Figures S1-S3 (pages 15-24), follow section F

A: CNV assay description and details

For all three samples, an attempt was made to type every individual twice for each assay. In the Michigan sample, a single fluorescent HEX dye was used for the HSPD21 assay, and duplication was achieved by typing each sample on at least two different 384-well typing plates, which helped to capture normal variation in the results of both the PCR amplification and the capillary run. For the Erlangen sample, duplication was achieved by mixing a small amount of product from two different PCR reactions—one labeled with a FAM dye and one with a HEX dye. This mix was digested with *Hae*III before electrophoresis. PCRs for the three CNV assays used for the Nijmegen sample were also performed with two different dye colors and mixed before capillary electrophoresis (FAM and NED dyes for HSPD21, FAM and HEX dyes for PRT107A, and HEX and NED dyes for the indel marker), and the duplexed products for these three markers were multiplexed in a single capillary. Duplication for the Nijmegen and Erlangen samples will hence control for normal variation in PCR amplification, but not for possible variation in capillary runs. On the other hand, using two different PRT estimates for the Nijmegen sample adds the ability to control for variation among different PRT assays.

An uncalibrated raw measure of copy number was obtained for the PRT assays by taking twice the ratio of the variably copy number DEFB4 peak to the diploid paralog peak. An unweighted mean of the duplicate PRT measures for each individual was then computed for the Erlangen and Michigan samples and for the PRT107A assay of the Nijmegen sample, but a weighted mean (weights in 2:1 ratio for FAM and NED estimates) was used for the HSPD21 assay in the Nijmegen sample because of problems with

the NED dye peaks. Peak height was used for computing ratios in the Michigan and Nijmegen samples, and peak area for the Erlangen sample. For the Nijmegen and Erlangen samples, the mean PRT ratio estimates of copy number were calibrated within the experiment using reference standards of known copy number. No such calibration was made for the Michigan data. After calibration, a further linear adjustment was applied to the PRT107A assay of the Nijmegen sample, since the calibrated estimates for this assay consistently underestimated copy number compared to the HSPD21 assay (see also Aldhous et al., 2010). For the Nijmegen sample, heights for the three potential allele peaks of the indel variant were averaged for the two fluorescent dye colors. No calibration was performed.

The triplex CNV assay used for the Nijmegen cohort combines the results of the unrounded PRT107A and HSPD21 estimates of copy number with the integer estimate of copy number most compatible with the ratio of peaks of the indel marker. For the two PRTs, the likelihood of each possible copy number was assessed using a Gaussian model. The probability of each possible copy number from the indel ratio test was assessed via an empirically validated likelihood analysis. A likelihood ratio analysis was then used to obtain a single integer estimate of copy number that maximizes the joint probability of the observed data (Aldhous et al. 2010). However, analysis with CNVtools requires unrounded non-integer estimates of copy number, and a probability-weighted copy number value was computed to provide an unrounded triplex assay estimate of copy number for model fitting and analysis.

The analysis of the Nijmegen sample considers, in addition to measures of copy number from the triplex assay and the two individual PRTs, a composite measure for the two PRT assays. Originally the intent was to analyze both the unweighted mean and the first principal component of the two different PRT estimates. In practice, these two composite measures are nearly identical when applied to the Nijmegen sample, whether principal components analysis (PCA) starts with the covariance or the correlation matrix for the two PRT variables (r^2 = 0.9999 in both cases). Since the mean can include any sample where at least one of the two PRT assays was successful, but PCA is by necessity restricted to samples where both PRTs worked, the unweighted mean was used as a composite measure.

B: Data quality, censoring and heterogeneity

For the Michigan and Erlangen samples, which both used only a single PRT assay, a filter was applied for censoring that rejected samples in which two measurements differed by more than 15% pf the mean. Because most Michigan samples that failed this filter were typed a third time, for these samples a threshold of a 7.5% coefficient of variation was used for censoring, which is equivalent to a 15% pairwise difference when there are only two estimates. For the Nijmegen samples, the 15% rule was applied to analyses of the individual HSPD21 and PRT107A assays. For analysis of copy number estimates based on the mean of the two PRTs, or those from the triplex assay, any individual who failed the 15% rule for either of the PRT assays was excluded from the censored analysis. Only 3.9% of the Michigan sample, and between 4.6-5.5% of the Erlangen sample qualified for censoring. Censoring affected a much larger

portion of the Nijmegen sample, removing 34.3-34.4% of the HSPD21 estimates, 22.6-23.0% of the PRT107A estimates, and 42.8-43.0% of the PRT mean and triplex assay estimates.

Supplementary Table 3 shows case and control sample sizes for the various single cohort analyses that are based on all possible combinations of analysis stage and collection center. This table also shows sample sizes for analyses based on different CNV assays for the Nijmegen collection and analyses of uncensored vs. censored data for all samples. Note the suffixes, which are used in combination with the collection center name in many of the results files to distinguish the various individual analyses—e.g., Nijmegen2_2c denotes an analysis that uses censored PRT107A assay data for 277 discovery (old) individuals in the Nijmegen sample.

The correlation between ratio estimates determined by peak area and peak height was very high for the Michigan sample (r^2 = 0.980), and also for the PRT107A-FAM , PRT107A-HEX, and HSPD21-FAM assays of the Nijmegen sample (r^2 = 0.980, 0.957, 0.992, respectively). The correlation between peak area-based versus peak height-based ratios was much lower in the Nijmegen sample for the HSPD21-NED assay (r^2= 0.732), reflecting distortion of peak area values by a nearby NED “dye peak”.

C: Raw CNV data analysis

The raw unrounded copy number data, both censored and uncensored, for each combination of cohort and CNV assay are compiled in a full data set available from JA (john.armour@nottingham.ac.uk) on request. Outlier estimates that were excluded from later analysis (see next section) are highlighted in yellow in that data set.

Histograms of these raw data are shown in Supplementary Figure 1. The dot density histograms give a fine-grained frequency display of the unrounded estimates of defensin copy number. The full range of estimates is shown in all cases. The different defensin copy numbers in the recruited individuals yield a distribution of unrounded estimates characterized by a mixture of overlapping bell-shaped curves or peaks. The quality of the estimates can be evaluated based on several criteria—peak symmetry, the degree of separation of the copy number peaks, the peak to valley height ratio, and the coincidence of peak means with integer copy number values. Based on these criteria, among the PRT assays the Michigan sample clearly has the best quality data, with distinct, symmetric, well-separated peaks for copy numbers of two to seven that rise far above the valleys between peaks, and peak means at or very close to integer values. The Erlangen data are next in quality, and the Nijmegen PRT estimates exhibit the poorest quality. The HSPD21 results are clearly superior to those of the PRT107A assay in the Nijmegen sample, so much so that the distribution of mean PRT estimates is not as good as the HSPD21 estimates alone. However, even the relatively poor clustering of PRT107A estimates around integer copy numbers is adequate for proper model fitting and association analysis (see sections D and E). The dot histograms clearly show the advantage of the triplex assay over a single PRT, as it has by far the tightest clustering of unrounded estimates around integer values among any of the samples.

The observed copy number estimates for the full uncensored dataset range from approximately 1.5 to 11 for the Michigan sample, 1 to 11 for the Erlangen sample, and 1.5 to 9 for the Nijmegen sample. A very large and nearly identical fraction of the unrounded estimates for all three cohorts fall within bounds of 1.5–7.5 copies (99.2%, 99.2%, 99.1% for Michigan, Erlangen, Nijmegen, respectively). Outlier estimates, defined as being at least ~0.5 copy number different from any other estimate in the cohort, are few—none in the Michigan cohort, two (1.083 and 11.055) in the Erlangen cohort, and one (8.86) in the Nijmegen cohort.

The histograms also show that censoring has little impact on the quality of the Michigan data, but offers a slight improvement in tightness of clustering for the Erlangen data, and a noticeable improvement for the Nijmegen data. This is probably just a simple consequence of the proportion of estimates subject to censoring in each cohort. Another informative way to display the raw unrounded copy number data is to superimpose kernel density curves for the distributions of estimates in cases and in controls (see Supplementary Figure 2). For each plot, differences in the sample sizes of cases and controls is normalized by adjusting the curves so that both have a total area of 1.0 density-copy number units. A nonparametric kernel density estimator is used, which does not impose any parametric functional form upon the density distribution. A relatively low value (0.15 on a 0-1 scale) for the tension parameter is used when plotting these curves, which smoothes out minor irregularities, but retains important changes in the density distributions.

These curves allow a visual assessment of potential differential biases in the distribution of copy number estimates between cases and controls, which can lead to spurious association test results if left uncorrected. For the Michigan sample, there appears to be no bias whatsoever in the copy number peak means. Variances of the peaks may be slightly greater in cases, though this is only apparent for peaks of copy number five and greater, and then primarily only at the base of the peak. The Erlangen sample, on the other hand, shows a very strong differential bias in both peak means and peak variances for the discovery (old), replication (new), and combined (all) data. Peaks for cases are substantially shorter and wider than those for controls and have means between 0.1 and 0.2 copies greater than those of controls. The censored Erlangen estimates show a similar pattern in terms of differential bias. Curves for the Nijmegen samples are more difficult to interpret because most of them display greater irregularity on account of the relatively small sample size and/or the relatively poor quality of clustering. However, little differential bias in peak means is evident, while there does appear to be noticeably greater peak variances for controls with the uncensored PRT measures. Censoring in this case appears to have a dramatic effect, mostly eliminating the differences in peak variances. The strong differential bias of the Erlangen data is likely a consequence of the Erlangen cases and controls being processed on different typing plates, whereas the Nijmegen and Michigan case and control samples were interspersed on each typing plate.

Because the case and control density curves bound equal areas, a visual assessment of association between disease and copy number is also possible. For the Michigan sample it is difficult to discern any

trend between increasing copy number and the ratio of case and control peak areas. The Erlangen curves for the discovery and combined datasets do appear to show a trend of increasing case:control peak area with increasing copy number, but this is confounded with differential biases in peak means and variances. The Nijmegen samples, especially after censoring, show a clear trend of an increasing ratio of case to control peak areas with increasing copy number.

D: Fitting distributions using CNVtools

CNVtools (Barnes et al., 2008) implements a likelihood ratio test that integrates classification of the raw copy number estimates into integer classes on the one hand, and association testing on the other, into a single procedure that compares mixture model fits under nested hypotheses. This method can handle noisy CNV data and account for systematic shifts in peak means or variances due to case or control membership. Also, because it combines copy number inference and association testing into a single statistical procedure, uncertainty implicit in the mixture model is correctly propagated into the association test (this includes not only the probabilities for each copy number class, but the uncertainties in the probabilities themselves). There are many different parameters of the mixture model fitting that can be varied, yielding many possible combinations that can variably affect the quality of fit. We therefore first refined the fit without the additional complexity of simultaneous association testing.

The choice of model parameters for distribution fitting within CNVtools was subjected to extensive evaluation for the (largest) Michigan sample before extending to the Erlangen and Nijmegen samples. In all cases, to obtain convergence it was necessary to restrict the modeling of either the component means or variances or both to a linear rather than a completely free model. The optimal number of components also varied, ranging from 6 to 9 for the uncensored data and 5 to 9 for the censored data.

The largest (Michigan) data set was used to exhaustively compare the degree of fit provided by CNVtools (version 1.44.0) for dozens of combinations of model parameters. These parameters included the number of fitted components (component = copy number), the minimum allowable frequency for any component, the use of Gaussian versus T distributions for the components, the modeling (free, linear, or constant) for the component means and variances, incorporation of case-control batch effects (differential biases) for the component means and variances, the utility of starting values (random, integer, or empirical) for the component means, the exclusion versus inclusion of extreme outlier measurements, and the maximum number of iterations of the fitting algorithm. Selection of the best fitting model was based on minimization of a combination of the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). These two criteria try to simultaneously maximize goodness of fit (measured by the log-likelihood statistic) while minimizing over-fitting caused by the use of too many parameters. From this comparison it was clear that a minimum component frequency of 2, Gaussian rather than T distributions, at least 30,000 iterations of the fitting algorithm, and the use of 10 different sets of starting values for the algorithm (consisting of 3 starts with integer values, 3 starts with

values derived empirically from kernel density curves, and 4 starts with randomly derived values), all resulted in consistently better fits to the raw data.

Later exploratory analysis of the Erlangen and Nijmegen data also showed the need to exclude extreme outlier observations. We then fixed this subset of parameters to their optimal values, but still varied the number of components, the modeling of component means and variances, and the use of case-control batch effects for further model fitting analysis. The need for incorporating case and control batch effects depended upon the degree of differential biases in the sample set and CNV assay used. Note the use of a "batch" term in the modeling of the moments whenever the p-value for the test of differential case-control bias is less than 0.10 (a low stringency p-value threshold was selected to ensure bias correction even in borderline situations). Tests for case-control bias were constructed as likelihood ratio tests that used the log-likelihoods for models fit with and without a case-control batch term. These tests for bias formally verified what was seen from the kernel density curves of cases and controls—the discovery and replication samples of the uncensored Erlangen cohort both show severe differential bias in peak means, which was not seen for the uncensored Michigan or the Nijmegen samples. All uncensored datasets have at least a modestly significant bias in peak variances, which rises to a more serious level for the original Erlangen cohort and the PRT107A assay of the Nijmegen sample. Censoring had little impact on the degree of bias for the Michigan and Erlangen data, but for the Nijmegen data it ameliorated variance bias for the three PRT measures while increasing variance bias for the triplex estimates.

Overlays of the fitted component peaks to the underlying histograms of CNV signal for cases and controls show that the selected best fit model was indeed a good fit in all cases (not shown). The tables of model fitting parameters list a more quantitative measure of degree of fit, a cluster quality score, which CNVtools calculates by comparing the locations of the peak means with the standard error for each pair of adjacent peaks (Supplementary Tables 1 and 2). This score measures how well the component clusters are separated; a quality score greater than 4 is usually good enough for association studies. Interestingly, this quantitative measure recapitulates the order of quality deduced from the dot histograms (Supplementary Figure 1): Nijmegen triplex > Michigan > Erlangen > Nijmegen PRT. Note also that the Nijmegen PRT107A assay gives the poorest cluster quality, with values under the acceptable threshold of 4, although the fitted model was still good enough to converge and allow further association testing. Bar graphs of most likely copy numbers for uncensored data are given in the manuscript as Figure 1, and for censored data as Supplementary Figure 3.

The full set of fitted CNV data is tabulated in an Excel file available from JA (john.armour@nottingham.ac.uk) on request.. For each analysis, this file lists the subject ID, affection status (1=control, 2 = case), posterior probabilities of each fitted component (P1, P2, P3, etc.), and the most likely component number (cn), which is based on Bayesian maximum *a posteriori* inference. Note that component number usually does not correspond directly with copy number—for most datasets the first component is usually a copy number of two, and for some datasets higher component numbers that are sequential represent copy numbers that may differ by more than one copy (i.e., components 8 and 9 for the Michigan data probably represent copy numbers of 9 and 11). These CNV data were

derived under the null hypothesis of no association, but the posterior probabilities are very close in value to those derived under both the null and alternative hypotheses when simultaneously running the fitting and association procedures (see next section).

E: Details of Association testing

Formal testing of association was achieved by using CNVtools to simultaneously applying mixture model fitting and association testing to the raw copy number data. For the Nijmegen and Erlangen samples, association was tested for 24 combinations of parameters: with or without a case-control batch term for means (2) x linear or free modeling of means (2) x with or without a batch term for variances (2) x constant, linear, or free modeling of variances (3). Association testing of the Michigan sample used 48 combinations of parameters: 7, 8, or 9 component classes (3) x with or without case-control batch term for means (2) x linear or free mean modeling (2) x with or without case-control batch term for variances (2) x linear or free variance modeling (2). For every cohort analysis the best fitting model and those models with nearly as good a fit, as assessed by the Akaike and Bayesian information criteria, all had nearly identical ORs, p-values, and confidence intervals. Hence the association results are robust to reasonable variations in the parameters selected for model fitting.

Detailed output for the best fitting models (not shown) was used to construct a likelihood ratio test for association under the null hypothesis of no association (H0) and the alternative hypothesis of association (H1). This detailed output included values for ns (the number of batches, which is always 2 for cases plus controls), nc (the number of copy number classes in the fitted mixture model), nind (the number of individuals), lnL (the log-likelihood of the observed data under the hypothesis), alpha (a matrix of mixture model proportions for each combination of copy number class and case-control batch, control = [1] and case = [2]), mean (a matrix of mixture model means), var (a matrix of mixture model variances), and pdc (a matrix of p(disease|c); i.e., the probability of disease given that component number). Posterior probabilities under both H0 and H1 were usually very similar to the posteriors computed under H0 with no association testing. This model output also included the status of the fitting algorithm (C = converged, M = maximum iterations reached without convergence, P = posterior density problem, and F = fit failed); C is the only acceptable status and was seen for all the best fitting models adopted.

A likelihood ratio association test was then constructed using the log-likelihoods output under the null hypothesis of no association (H0) and the alternative hypothesis of association (H1). Odds ratios were computed using the pdc values, and 95% Wald-type confidence intervals derived using the normally distributed likelihood ratio test statistic. These association test results for all single cohort analyses are summarized in Table 1 and Supplementary Table 4.

Each single cohort analysis generated a pair of output files, with results for two different assumed disease models—a linear model where the effect on the log-odds of disease is proportional to the number of copies, and an allelic model where the odds are not constrained by a linear trend but may

freely vary among copy numbers. The allelic model, which has more degrees of freedom than the linear trend model, will always fit the data better but will have less power to detect association. A likelihood ratio test can be used to assess whether the extra fit afforded by the allelic model is statistically significant—if not, the linear model is chosen instead. The association results for both the linear trend and allelic models are given in Table 1 and Supplementary Table 4. In all cases, the extra fit provided by the allelic model was not significant, so the simpler and more powerful linear trend model is the better choice.

F: Clinical details of new Michigan and Erlangen samples

Characteristics of Michigan cohort

Recruitment criteria

Cases and controls were recruited from southeastern Michigan and were of self-reported European Caucasian ancestry. Individuals were defined as affected if chronic plaque or guttate psoriasis lesions covered more than 1% of the total body surface area or if at least two skin, scalp, nail or joint lesions were clinically diagnostic of psoriasis. Controls were required to be unrelated to each other or to any case, and to be free of a family history of psoriasis.

Demographic and phenotypic characteristics of Michigan psoriasis cases (n = 2616):

- Gender (n = 2616 with data): 1382 females (52.8%) and 1234 males (47.2%)
- Type of psoriasis lesions (n = 1570 with data): 1476 (94.0%) with chronic plaque lesions only, 38 (2.4%) with both chronic plaque and guttate lesions, and 56 (3.6%) with guttate lesions only. These lesional subphenotypes were not explicitly recorded in the earliest years of patient recruitment (though recruitment has always been restricted to patients with chronic plaque and/or guttate lesions), which explains the relatively small proportion of the sample with recorded data.
- Type I & type II psoriasis, defined as age at onset ≤40 & >40 years of age (n = 2603 with data): 2023 (77.7%) with type I or early-onset psoriasis and 580 (22.3%) with type II or late-onset psoriasis
- Age at onset (n = 2603 with data): median is 24 years; mean ± sd = 28.0 ± 16.8 years
- Age at exam (n = 2607 with data): median is 51 years; mean ± sd = 49.5 ± 16.5 years
- Psoriatic arthritis (n = 2190 with data): 518 (23.7%) with psoriatic arthritis and 1672 (76.3%) with a definite absence of psoriatic arthritis

- Fingernail symptoms such as pitting or lifting (n = 2514 with data): 1532 (60.9%) with nail symptoms and 982 (39.1%) with no nail symptoms

Characteristics of new Erlangen samples

Recruitment criteria

Cases and controls were recruited at nine different centers in Germany (3 rehabilitation hospitals, 5 university hospitals, 1 private practice). Individuals were defined as affected if they had chronic plaque psoriasis or psoriatic arthritis. Controls were required to be free of a family history of psoriasis.

Demographic and phenotypic characteristics of independent Erlangen psoriatic cases (n = 1397):

- Gender (n = 1395 with data): 535 females (38.4%) and 860 males (61.6%)

- Type I & type II psoriasis, defined as age at onset $\leqslant$40 & >40 years of age (n = 1337 with data): 1186 (88.7%) with type I or early-onset psoriasis and 151 (11.3%) with type II or late-onset psoriasis

- Age at onset (n = 1315 with data): mean ± sd = 25.6 ± 13.0 years

- Age at recruitment (n = 1375 with data): mean ± sd = 50.0 ± 13.0 years

- Psoriatic arthritis (n = 1397 with data): 646 (46.2%) with psoriatic arthritis and 751 (53.8%) with a definite absence of psoriatic arthritis at the time of recruitment

The 621 individuals of the independent German control cohort had no psoriasis vulgaris at the time of recruitment, when the average age was 30.5 ± 9.3 years. All of them were healthy blood donors recruited in Northern Germany, and 358 individuals (58%) were male.

Table S1: Model fitting for single cohort models (uncensored)

cohort	subset	assay	no. of case/control	outliers dropped	nc[1]	p-value for test of differential case-control bias		modelling of moments		cluster quality
						means	variances	means	variances	
Michigan	all	HSPD21	2616 / 2526	0	9	0.20	0.0037	~strata(cn)	~batch*cn	5.940
Erlangen	all	HSPD5	1696 / 910	2	8	4.1×10^{-15}	1.4×10^{-5}	~strata(batch,cn)	~batch*cn	4.404
	new	HSPD5	1396 / 621	1	8	2.0×10^{-6}	0.013	~strata(batch,cn)	~batch*cn	4.380
	old	HSPD5	300 / 289	1	6	1.3×10^{-11}	1.8×10^{-9}	~batch*cn	~strata(batch,cn)	4.273
Nijmegen	all	HSPD21	187 / 238	1	6	0.37	0.0059	~cn	~batch*cn	4.270
		PRT107A	152 / 229	2	6	0.87	0.019	~cn	~batch*cn	3.547
		PRT mean	188 / 246	1	6	0.63	1.4×10^{-5}	~cn	~batch*cn	4.120
		triplex	188 / 246	1	6	0.52	0.0035	~cn	~batch	21.840
	old	HSPD21	170 / 230	1	6	0.49	0.0078	~cn	~batch*cn	4.254
		PRT107A	137 / 219	2	6	0.79	0.015	~cn	~batch	3.609
		PRT mean	171 / 235	1	6	0.93	1.6×10^{-5}	~cn	~batch*cn	4.152
		triplex	171 / 235	1	6	0.40	0.13	~strata(cn)	~1	25.283

[1]number of fitted components (copy numbers)

Table S2: Model fitting for single cohort models (censored)

cohort	subset	assay	no. of case/control	outliers dropped	nc[1]	p-value for test of differential case-control bias		modelling of moments		cluster quality
						means	variances	means	variances	
Michigan	all	HSPD21	2553 / 2444	0	9	0.45	0.0025	~strata(cn)	~batch*cn	5.984
Erlangen	all	HSPD5	1594 / 875	2	8	1.8×10^{-14}	1.3×10^{-5}	~strata(batch,cn)	~batch*cn	4.480
	new	HSPD5	1314 / 593	1	8	1.5×10^{-6}	0.018	~strata(batch,cn)	~batch*cn	4.443
	old	HSPD5	280 / 282	1	6	6.9×10^{-11}	5.2×10^{-7}	~batch*cn	~batch*cn	4.622
Nijmegen	all	HSPD21	158 / 121	1	6	0.68	0.86	~cn	~ cn	5.388
		PRT107A	126 / 167	2	6	0.33	0.063	~cn	~batch*cn	3.838
		PRT mean	137 / 109	2	5	0.76	0.059	~cn	~batch*cn	5.228
		triplex	137 / 109	2	5	0.34	5.9×10^{-6}	~cn	~batch*cn	29.101
	old	HSPD21	146 / 117	1	6	0.62	0.93	~cn	~cn	5.361
		PRT107A	112 / 163	2	6	0.64	0.044	~cn	~batch	3.580
		PRT mean	126 / 105	2	5	0.36	0.10	~strata(cn)	~1	5.181
		triplex	126 / 105	2	5	0.31	1.0×10^{-5}	~cn	~batch*cn	29.049

[1]number of fitted components (copy numbers)

Table S3: Single cohort analyses

			uncensored		censored	
cohort	**subset**	**assay**	**na / nu / n[1]**	**suffix**	**na / nu / n[1]**	**suffix**
Michigan	all	HSPD21	2616 / 2526 / 5142	1	2553 / 2444 / 4997	1c
Erlangen	all	HSPD5	1698 / 910 / 2608	1	1596 / 875 / 2471	1c
	new	HSPD5	1397 / 621 / 2018	2	1315 / 593 / 1908	2c
	old	HSPD5	301 / 289 / 590	3	281 / 282 / 563	3c
Nijmegen	all	HSPD21	188 / 238 / 426	1_1	159 / 121 / 280	1_1c
		PRT107A	154 / 229 / 383	1_2	128 / 167 / 295	1_2c
		mean of 2 PRTs	189 / 246 / 435	1_3	139 / 109 / 248	1_3c
		triplex	189 / 246 / 435	1_4	139 / 109 / 248	1_4c
	old	HSPD21	171 / 230 / 401	2_1	147 / 117 / 264	2_1c
		PRT107A	139 / 219 / 358	2_2	114 / 163 / 277	2_2c
		mean of 2 PRTs	172 / 235 / 407	2_3	128 / 105 / 233	2_3c
		triplex	172 / 235 / 407	2_4	128 / 105 / 233	2_4c

[1]na / nu / n = no. affected, no. unaffected, total number

Table S4: Single cohort association tests (censored)

cohort	subset	assay	no. of case/control	linear trend model			allelic model p-value	p-value for test of fit of allelic vs. linear trend model
				OR	95% CI	p-value		
Michigan	all	HSPD21	2553 / 2444	1.059	1.006–1.115	0.029	0.50	0.92
Erlangen	all	HSPD5	1594 / 875	1.095	1.013–1.183	0.019	0.47	0.97
	new	HSPD5	1314 / 593	1.058	0.967–1.156	0.22	0.81	0.90
	old	HSPD5	280 / 282	1.204	1.021–1.419	0.027	0.25	0.78
Nijmegen	all	HSPD21	128 / 167	1.276	0.985–1.654	0.065	0.60	0.99
		PRT107A	139 / 109	1.224	0.960–1.560	0.10	0.62	0.93
		PRT mean	139 / 109	1.413	1.071–1.864	0.015	0.27	0.98
		triplex	147 / 117	1.397	1.062–1.836	0.017	0.30	0.99
	old	HSPD21	114 / 163	1.265	0.969–1.651	0.085	0.65	0.99
		PRT107A	128 / 105	1.228	0.944–1.597	0.13	0.51	0.75
		PRT mean	128 / 105	1.374	1.034–1.826	0.029	0.25	0.89
		triplex	1596 / 875	1.374	1.036–1.824	0.028	0.20	0.78

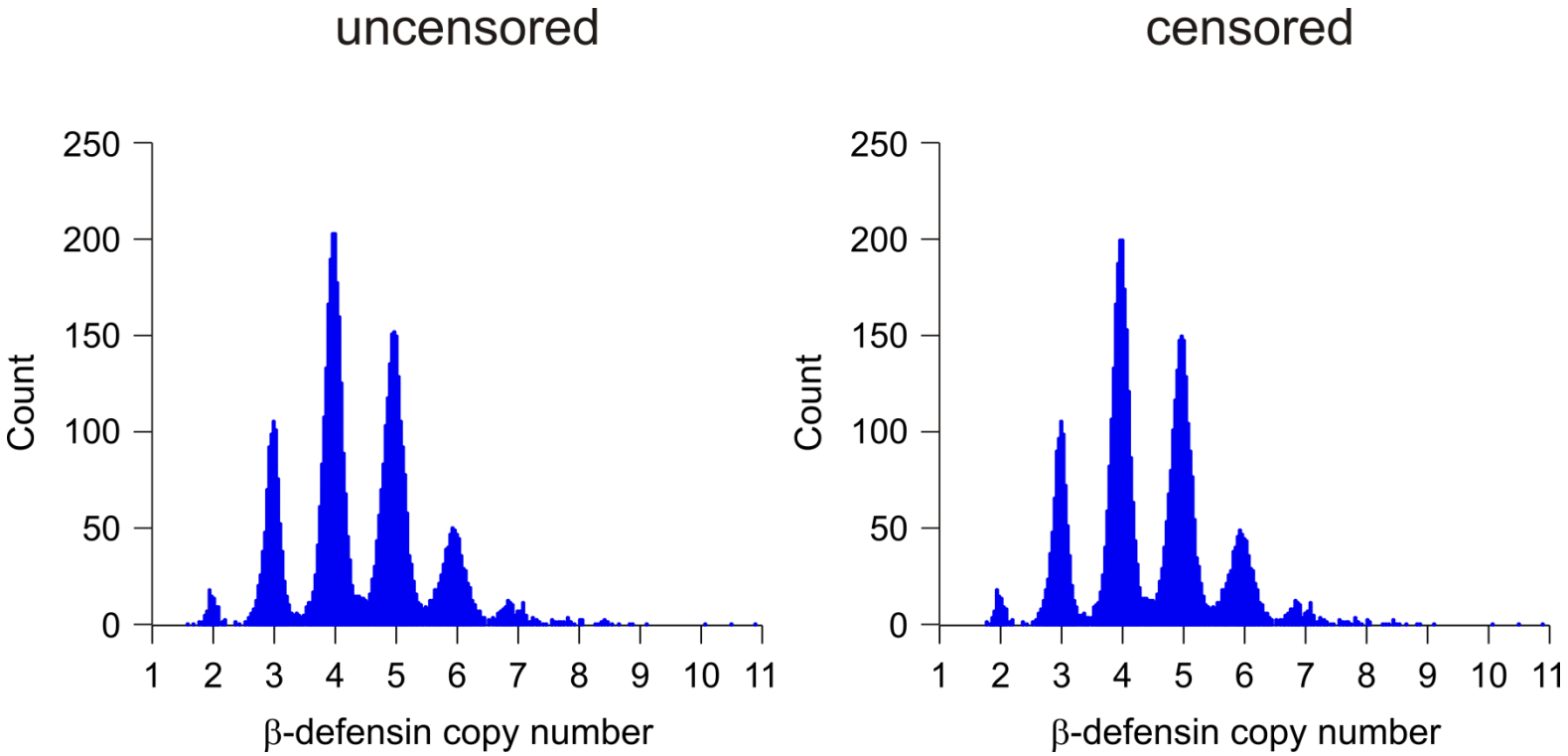
Table S5: Meta-analyses (censored)

stage	cohorts	no. cases/controls	association (fixed effects)		association (random effects)		heterogeneity	
			meta-OR (95% CI)	meta-P	meta-OR (95% CI)	meta-P	Cochran's Q (p-value)	I^2 (95% CI)
discovery	Erlangen (old) + Nijmegen (old, HSPD21)	426 / 399	1.220 (1.061–1.404)	5.3E-03	1.220 (1.061–1.404)	5.3E-03	0.095 (0.76)	0.0 (—)
	Erlangen (old) + Nijmegen (old, PRT107A)	392 / 445	1.211 (1.053–1.392)	7.3E-03	1.211 (1.053–1.392)	7.3E-03	0.015 (0.90)	0.0 (—)
	Erlangen (old) + Nijmegen (old, mean PRT)	406 / 387	1.244 (1.079–1.435)	2.6E-03	1.244 (1.079–1.435)	2.6E-03	0.62 (0.43)	0.0 (—)
	Erlangen (old) + Nijmegen (old, triplex)	406 / 387	1.245 (1.080–1.435)	2.5E-03	1.245 (1.080–1.435)	2.5E-03	0.63 (0.43)	0.0 (—)
replication	Michigan + Erlangen (new)	3867 / 3037	1.059 (1.012–1.107)	1.2E-02	1.059 (1.012–1.107)	1.2E-02	0.008 (0.98)	0.0 (—)
combined	Michigan + Erlangen (all) + Nijmegen (all, HSPD21)	4305 / 3440	1.075 (1.031–1.122)	8.2E-04	1.077 (1.028–1.129)	2.0E-03	2.20 (0.33)	9.3 (0.0–46.0)
	Michigan + Erlangen (all) + Nijmegen (all, PRT107A)	4273 / 3486	1.075 (1.030–1.121)	8.9E-04	1.075 (1.030–1.121)	8.9E-04	1.63 (0.44)	0.0 (0.0–87.2)
	Michigan + Erlangen (all) + Nijmegen (all, mean PRT)	4284 / 3428	1.077 (1.032–1.124)	6.2E-04	1.096 (1.021–1.200)	2.2E-02	4.25 (0.12)	53.0 (0.0–86.5)
	Michigan + Erlangen (all) + Nijmegen (all, triplex)	4284 / 3428	1.077 (1.032–1.124)	6.2E-04	1.094 (1.015–1.180)	2.0E-02	4.04 (0.13)	50.4 (0.0–85.7)

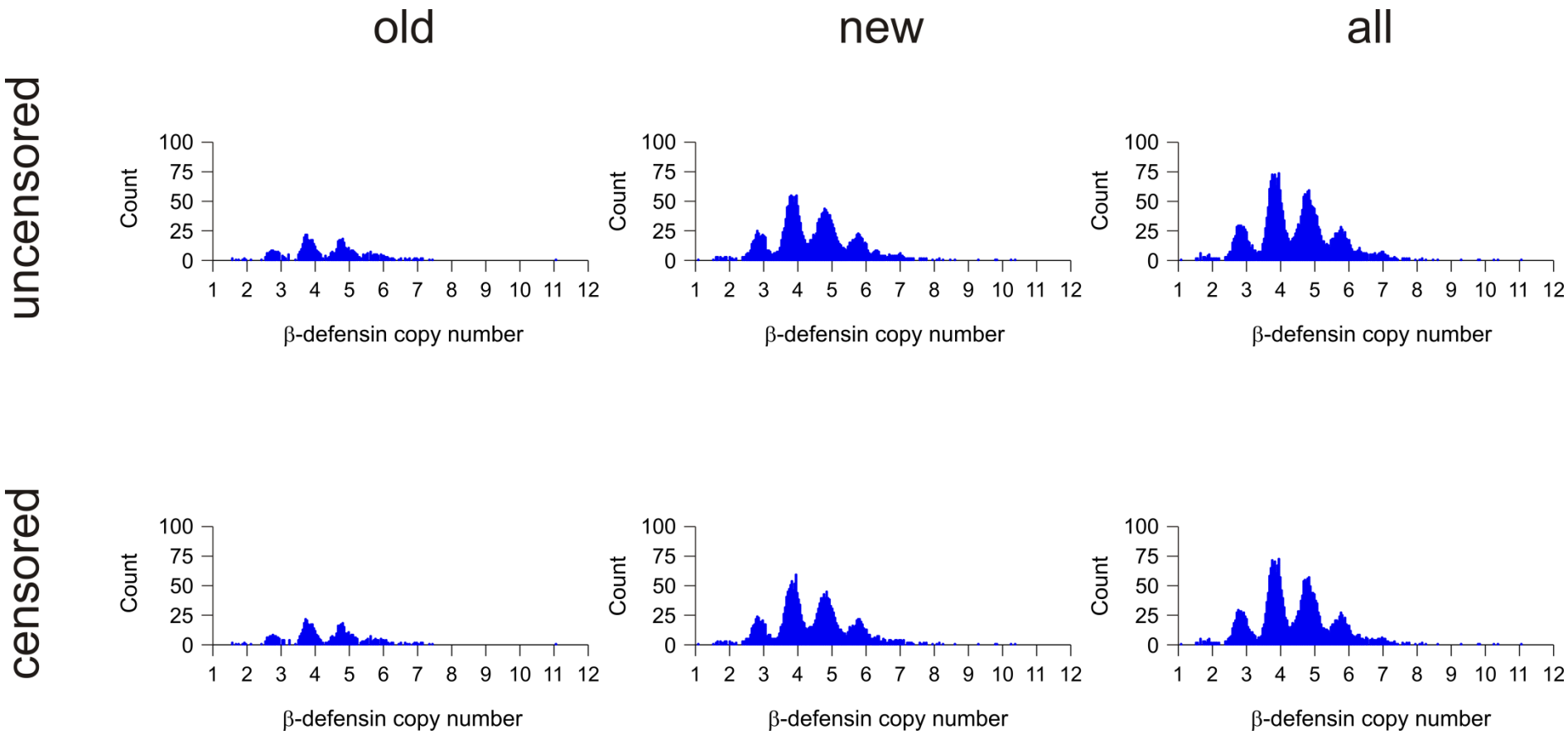
Supplementary Figures

Figure S1: dot density histograms

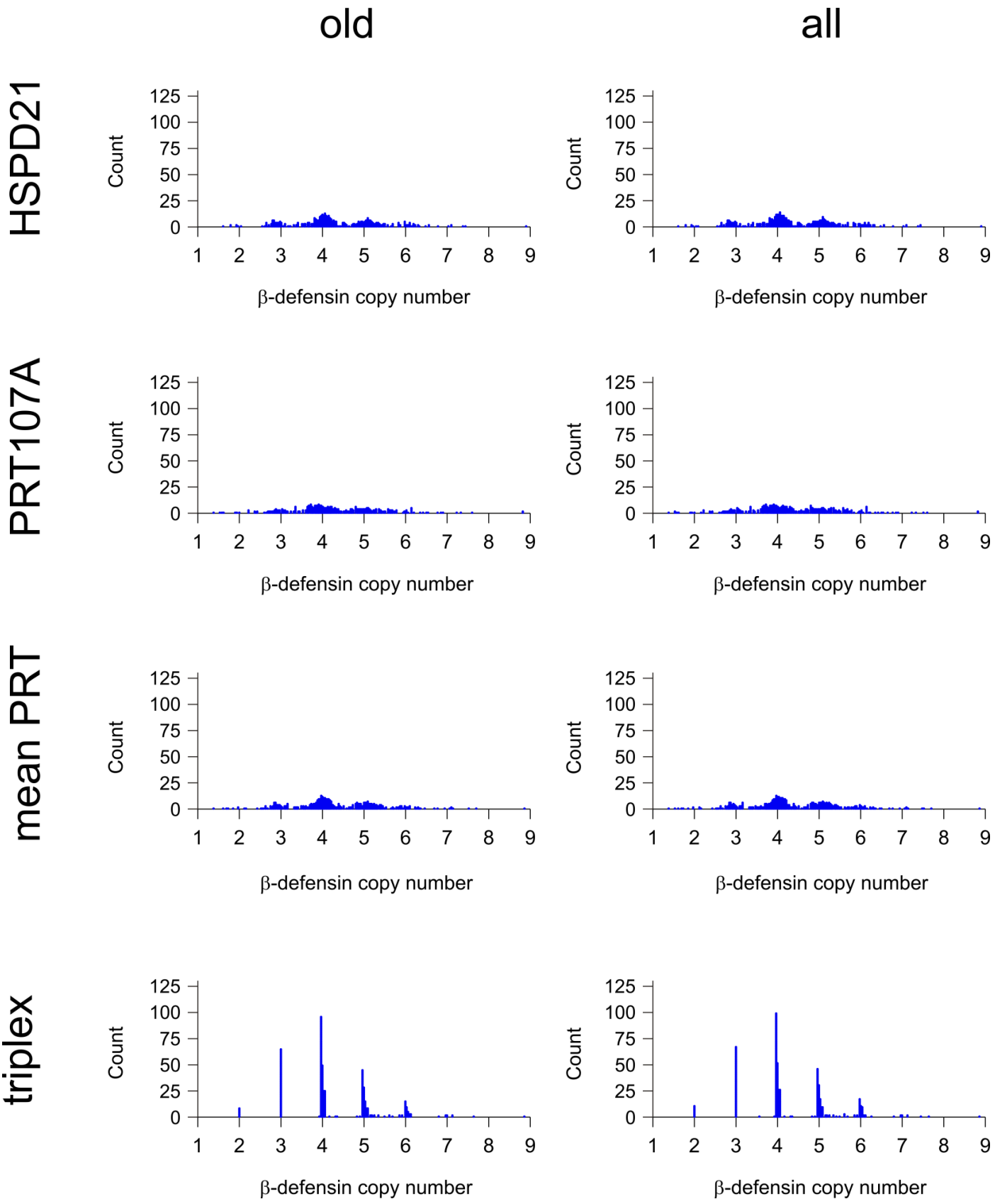
(a) Dot density histograms for Michigan sample



(b) Dot density histograms for Erlangen samples



(c) Dot density histograms for uncensored Nijmegen samples



(d) Dot density histograms for censored Nijmegen samples

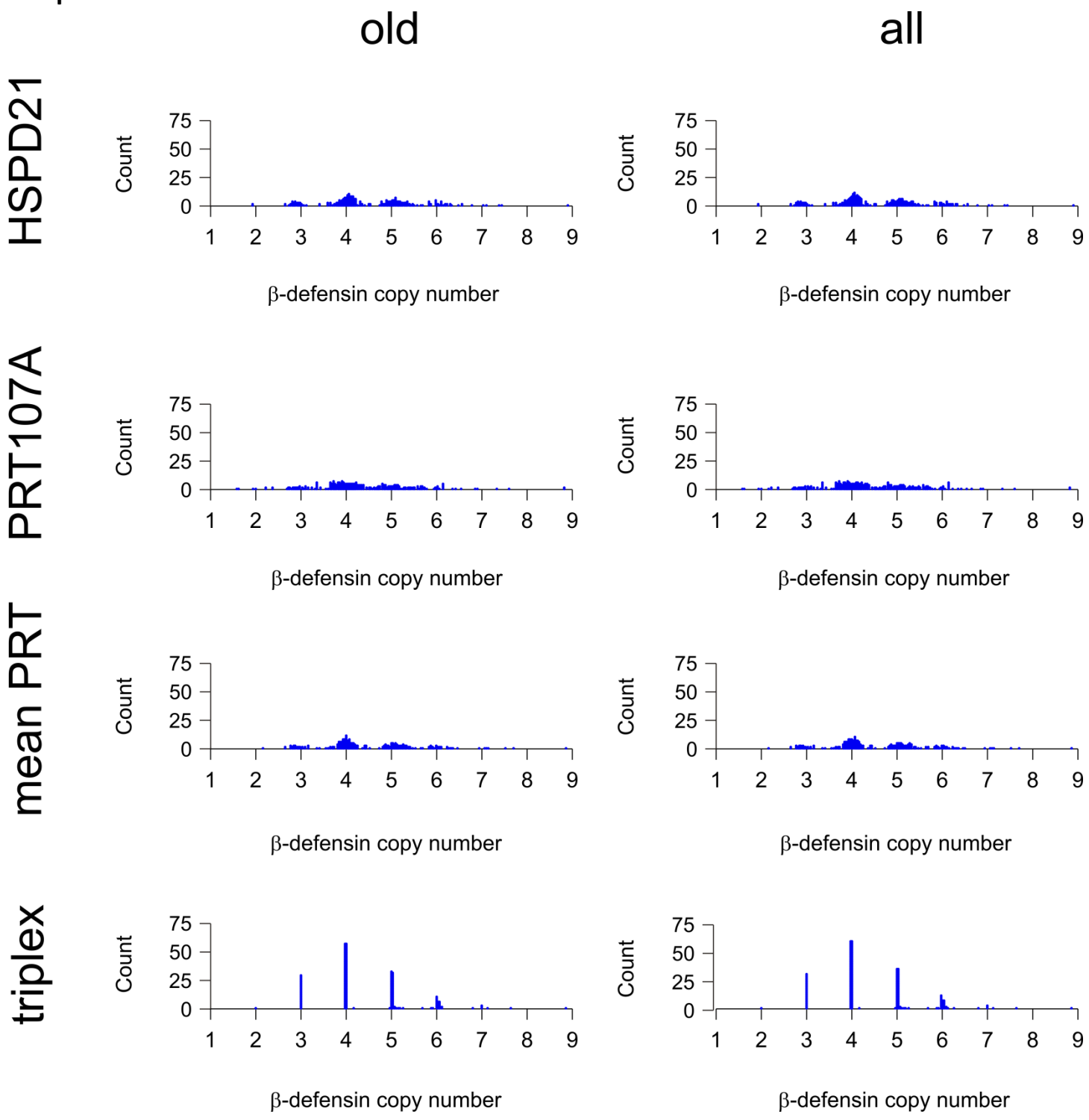
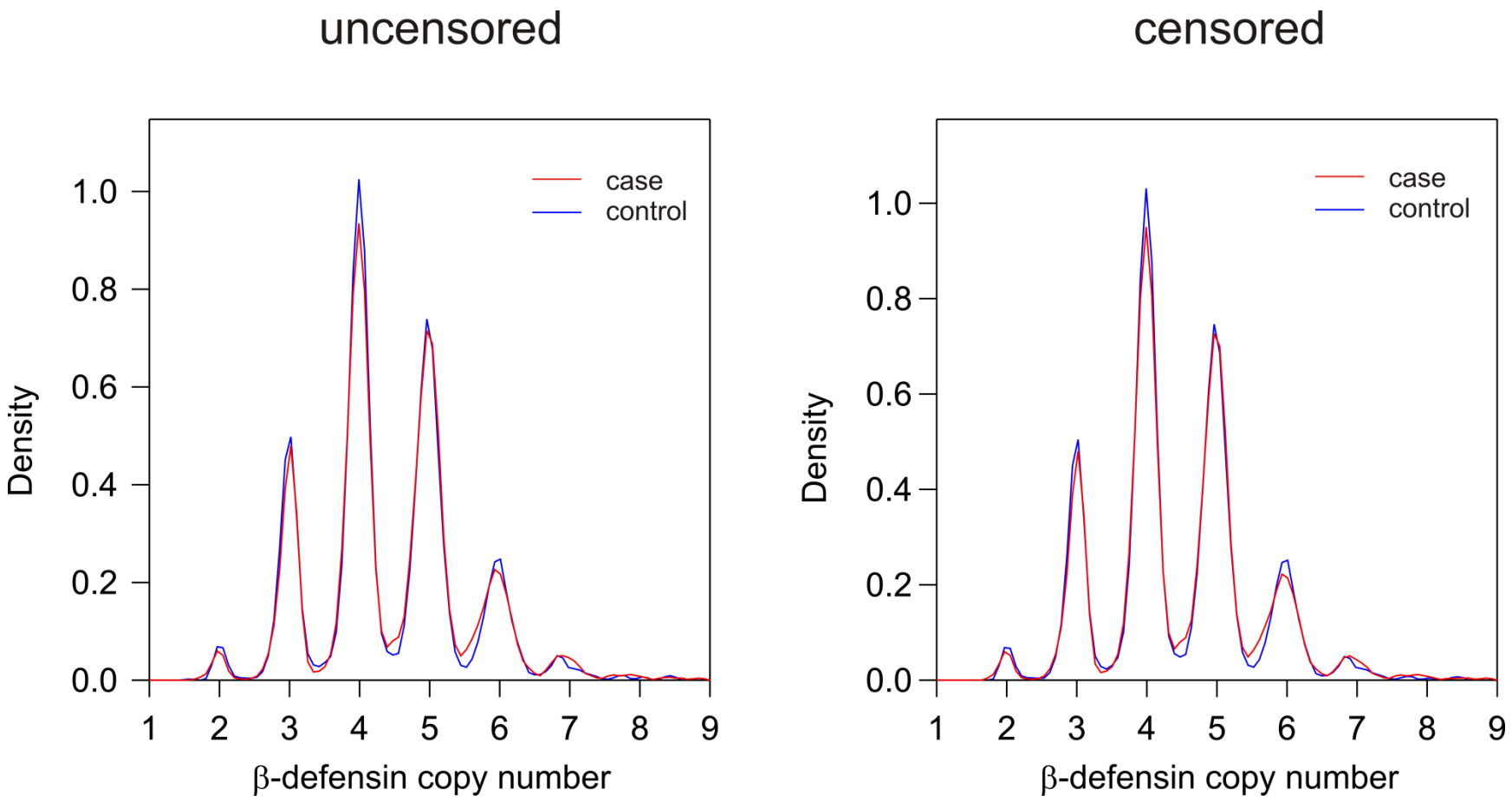
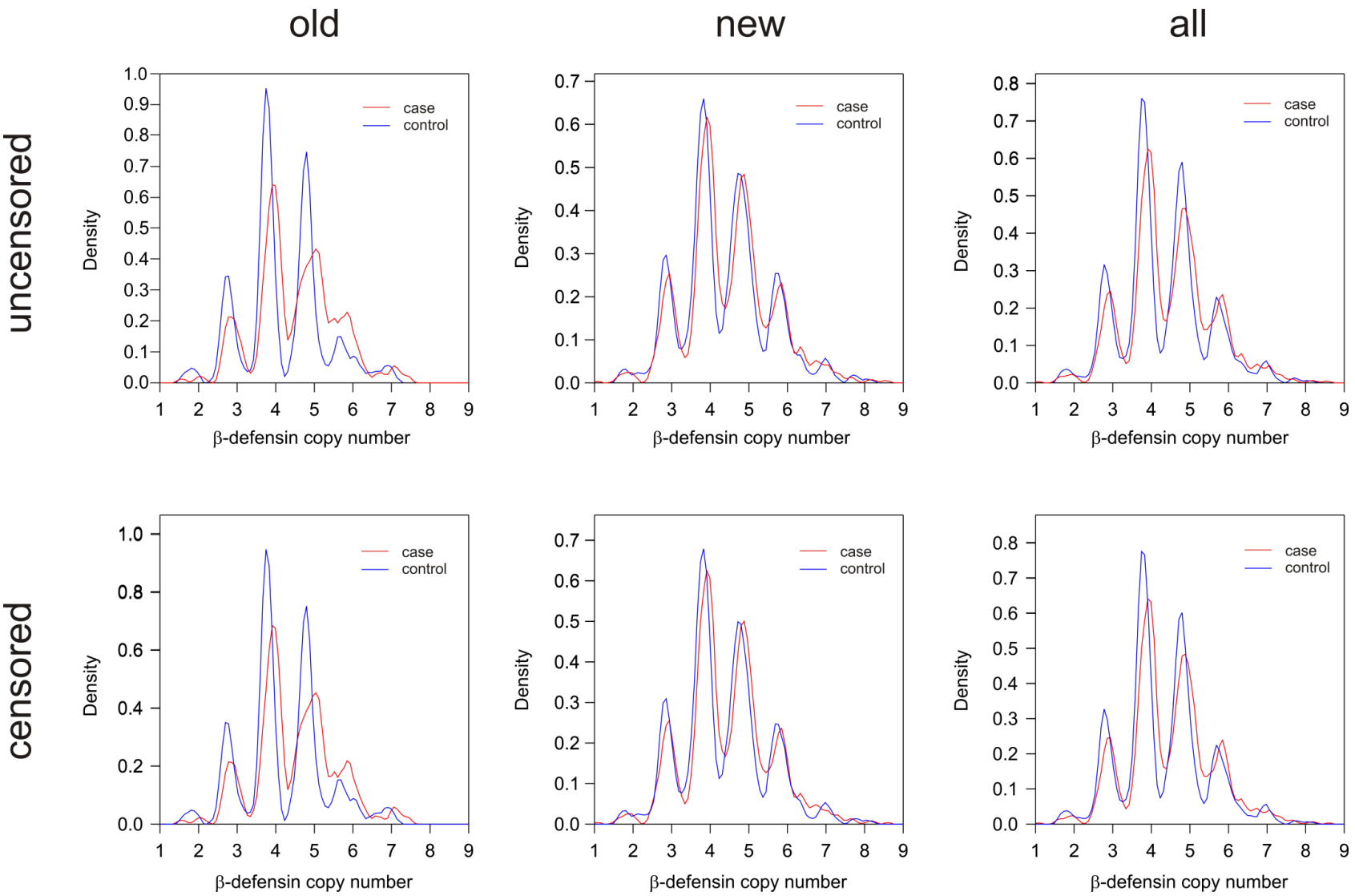


Figure S2: Kernel density plots

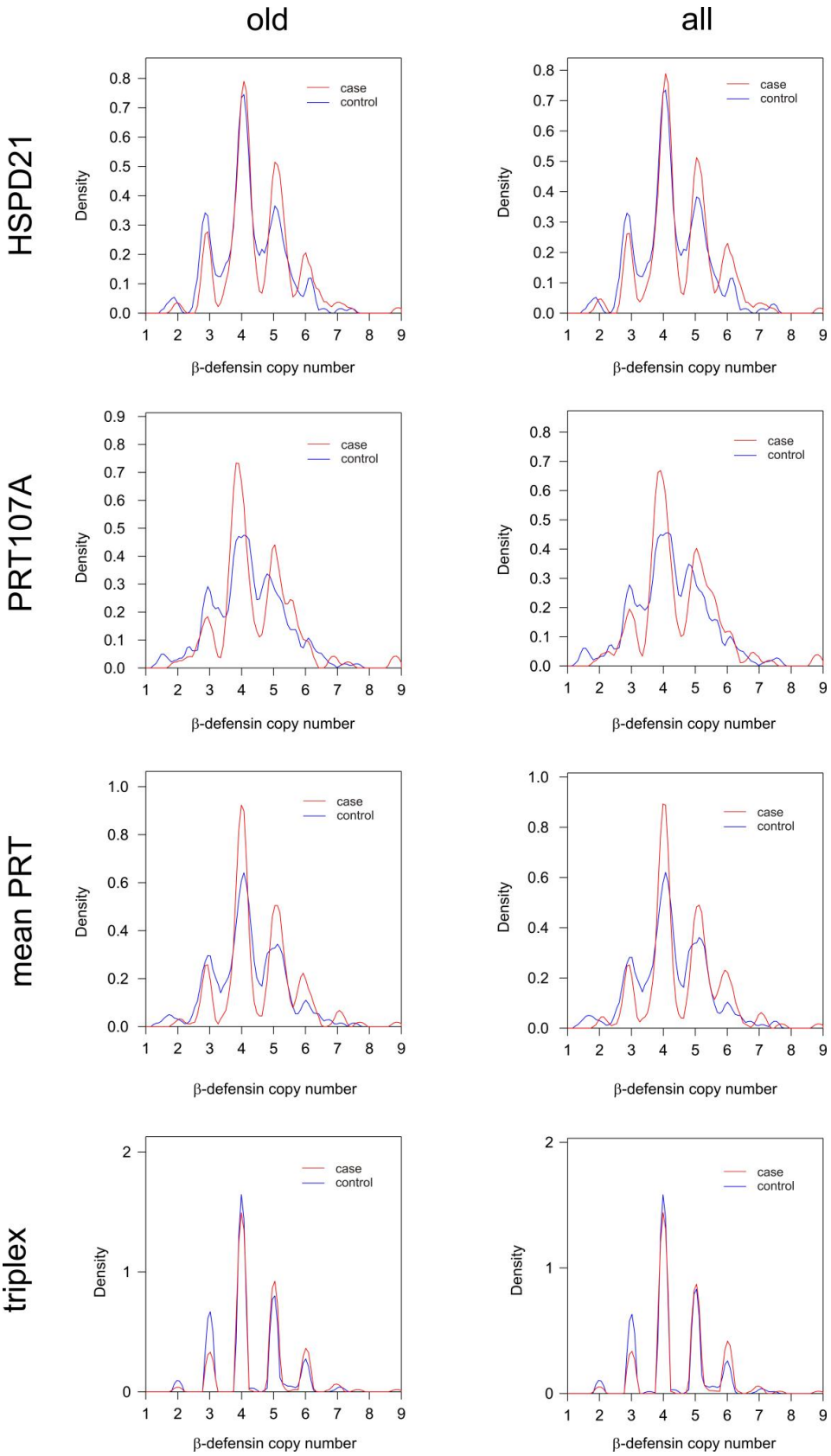
(a) Kernel density plots for cases and controls of the Michigan sample



(b) Kernel density plots of cases and controls of the Erlangen samples



(c) Kernel density plots of cases and controls of the uncensored Nijmegen samples



(d) Kernel density plots of cases and controls of the censored Nijmegen samples

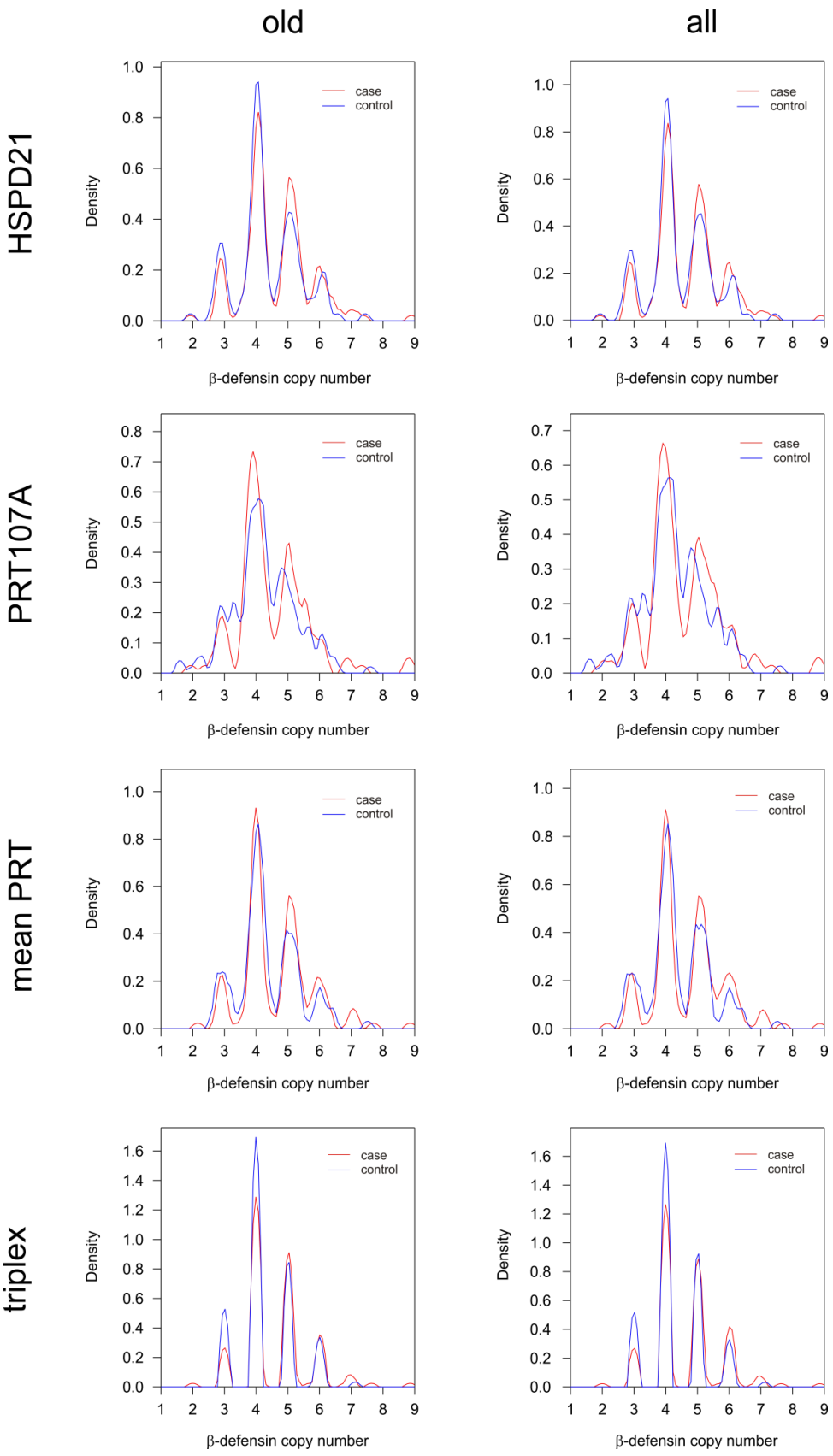
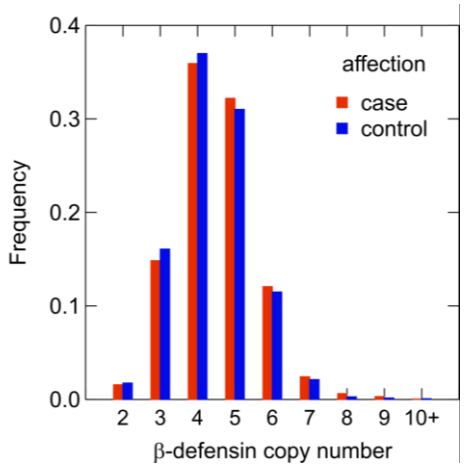


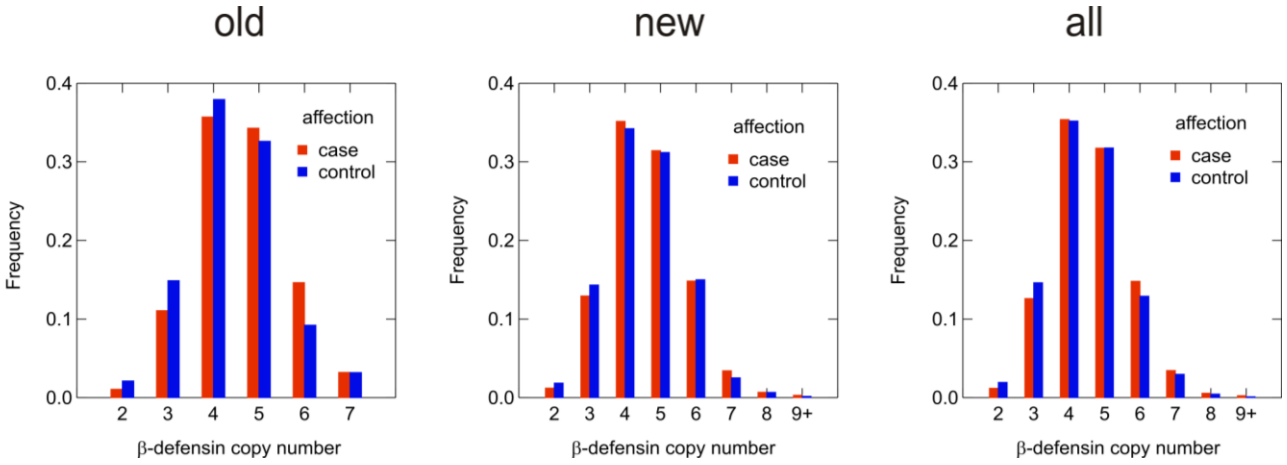
Figure S3

Bar graphs for fitted copy number estimates from censored data

(a) Michigan samples



(b) Erlangen samples



(c) Nijmegen samples