VIPER Supplementary Materials

This document was created using R markdown. The source document and data are available in the "viper-Data" R package (1). There are three sections:

- Study Designs
- Sizing
- Histograms of reader scores and ROC curves.

Study Designs

In Fig. S1A through Fig. S1E we graphically show the reader and case sampling of the five VIPER reader studies. Each (row, column) pixel corresponds to a reader/radiologist (row) and a case/patient (column). The readers may be different from study to study. The cases are organized by type: BIRADS 0 by full-field digital mammography (FFDM), BIRADS 0 by screen-film mammography (SFM), BIRADS 1-2 by FFDM, BIRADS 1-2 by SFM, and cancer. If the pixel is white, the corresponding reader evaluated the corresponding case. If the pixel is gray, the corresponding reader did not read the corresponding case. These "missing" observations were either by design or were accidentally skipped by readers during data collection. On average readers missed less than one case per modality and sub-study; there was a total of 138 missing observations out of 20,520 planned.

For example, in the cancer block of Fig. S1B we see the prototypical split-plot design with two missing cases. Readers 1-5 read the first 30 cases, readers 6-10 read the next 31 cases, readers 11-15 read the next 29 cases, and readers 16-20 read the last 25 cases. In the "BIRADS 0 by FFDM" block of Fig. S1B, every reader read 4 cases. During the study, these four cases were split into two case sets so that each reading session had two cases. For readers 1-5, these cases appear sequentially in Fig. S1B. For the other readers, the cases appear separated into two locations.

We could reorder each block to show the prototypical split-plot design, but the order would be different for each reader study. Instead, we wanted to keep the same order for all the study designs, even if that meant some blocks didn't appear as the prototypical split-plot design. Using the same order for all the study designs shows how the studies fit together. For example, most readers that participated in the screeningLowP reader study also participated in the screeningMedP and challengeMedP studies. If you trace any line down through these study designs, you will find that no reader group read the same case more than once.

Lastly, we can distinguish the screening studies (Figs. S1A-S1C) from the challenge studies (Figs. S1D-S1E) by the sparse inclusion of BIRADS 0 cases and heavy inclusion of BIRADS 1-2 cases. We can also see changes in prevalence as the number of non-cancer cases decreases from Fig. S1A to Fig. S1C and from Fig. S1D to Fig. S1E.

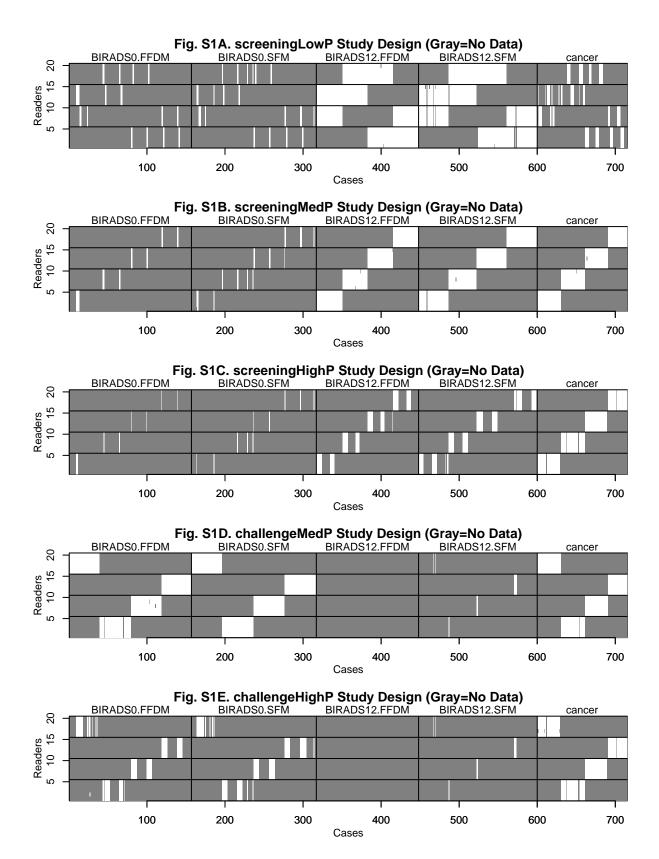


Figure 1: Figure S1: Graphical depiction of VIPER reader study designs.

Sizing

In what follows, we sketch out the steps we took to size the VIPER reader studies. We targeted a standard error of 0.03, believing that would allow ample precision for testing an effect size of 0.11, the difference between the area under the receiver operating characteristic curves (AUCs, ROCs) from FFDM and SFM observed in DMIST (2). The power for a significance level of 0.01 (0.05 split evenly among 5 reader studies) is 0.86 under a normal approximation.

The process to size VIPER started with DMIST reader study data:

- Hendrick2008_Radiology_v247p38 (3): 'Accuracy of soft-copy digital mammography versus that of screen-film mammography according to digital manufacturer: ACRIN DMIST retrospective multireader study.'
 - There were four machines studied, labeled below as hendrick2008.GE, hendrick2008.Fuji, hendrick2008.Fischer, and hendrick2008.Hologic. The Hologic data is not included since only a few cancer cases were available.
- An unpublished study evaluating the effect of breast density on diagnostic accuracy of FFDM and SFM
 - There were four density levels studied, corresponding to the four BIRADS density levels, labeled as breastDensity.A, breastDensity.B, breastDensity.C, breastDensity.D. The study included 320 women (109 cancers) stratified by original SFM breast density determination, 8 readers reading FFDM, and 8 readers reading SFM.
- An unpublished study evaluating performance in a reader study at 50% prevalence. The scope of the study has been described as including 30 readers assigned to read a total of 300 cases by FFDM and SFM.
 - There was one data set, labeled below as prevalence 50.
- Nishikawa2009_Radiology_v251p41, 'Comparison of Soft-Copy and Hard-Copy Reading for Full-Field Digital Mammography.' This data was not used here since it did not compare FFDM to SFM.

Tables S1 and S2 show the U-statistics-based performance results as calculated by version 4.0 of the iMRMC application developed at the FDA (4).

Table S1: Performance Results for DMIST Reader Studies

	NR	N0	N1	AUC.FFDM	SE(AUC.FFDM)	AUC.SFM	SE(AUC.SFM)
hendrick2008.GE	12	72	48	0.773	0.030	0.805	0.030
hendrick2008.Fuji	12	71	27	0.720	0.047	0.762	0.047
hendrick2008.Fischer	6	73	42	0.718	0.041	0.742	0.041
breastDensity.A	16	61	19	0.852	0.042	0.890	0.042
breastDensity.B	16	40	39	0.767	0.036	0.758	0.036
breastDensity.C	16	41	39	0.828	0.030	0.818	0.030
breastDensity.D	16	69	12	0.548	0.051	0.672	0.051
prevalence50	29	156	144	0.765	0.019	0.788	0.019

Table S1 footnote: NR is the number of readers, N0 is the number of non-cancers, N1 is the number of cancers, AUC.FFDM and AUC.SFM are the AUCs for FFDM and SFM evaluations, and SE is the standard error.

Table S2: Performance Differences for DMIST Reader Studies

	NR	N0	N1	AUC.FFDM minus AUC.SFM	SE(AUC.FFDM minus AUC.SFM)
hendrick2008.GE	12	72	48	-0.032	0.024
hendrick2008.Fuji	12	71	27	-0.042	0.026
hendrick2008.Fischer	6	73	42	-0.025	0.024
breastDensity.A	16	61	19	-0.037	0.033
breastDensity.B	16	40	39	0.009	0.026
breastDensity.C	16	41	39	0.010	0.028
breastDensity.D	16	69	12	-0.123	0.055
prevalence50	29	156	144	-0.022	0.017

Table S2 footnote: NR is the number of readers, N0 is the number of non-cancers, N1 is the number of cancers, AUC.FFDM and AUC.SFM are the AUCs for FFDM and SFM evaluations, and SE is the standard error

The total variance of AUCs or differences in AUCs from a fully-crossed study can be expressed as (5,6)

Equation 1

$$V = \frac{var_0}{N_0} + \frac{var_1}{N_1} + \frac{var_{01}}{N_0N_1} + \frac{var_R}{N_R} + \frac{var_{0R}}{N_0N_R} + \frac{var_{1R}}{N_1N_R} + \frac{var_{01R}}{N_0N_1N_R},$$

where N_0 , N_1 , N_R are the number of normal (non-cancer) cases, diseased (cancer) cases, and readers, var_0 , var_1 , and var_R are the variances due to the normal cases (non-cancers), the diseased cases (cancers), and the readers. The other variance components correspond to variances arising from interactions between the main variance components.

When the study is not fully crossed, the weights for each variance component are different and calculable (7). In particular, we can calculate the weights for a split-plot study design. For example, consider a split-plot study with N_G groups that have the same number of readers and cases $(N_0, N_1, \text{ and } N_R \text{ are all divisible by } N_G)$. Assuming a study with readers and cases paired across the modalities, the variance of the difference in AUCs is given by

Equation 2

$$V = \frac{var_0}{N_0} + \frac{var_1}{N_1} + \frac{N_G var_{01}}{N_0 N_1} + \frac{var_R}{N_R} + \frac{N_G var_{0R}}{N_0 N_R} + \frac{N_G var_{1R}}{N_1 N_R} + \frac{N_G^2 var_{01R}}{N_0 N_1 N_R}.$$

In addition to performance results, the DMIST reader studies provide estimates of the MRMC variance components for the AUC differences. We can use these to size a study to compare FFDM to SFM for women with dense breasts.

Table S3: Variance Components of AUC Differences for DMIST Reader Studies

-	var_0	var_1	var_{01}	var_R	var_{0R}	var_{1R}	var_{01R}
	0470	car ₁	047 01	car _R	car or	ear _{1R}	
machA	0.0082	0.0186	0.0029	-0.0009	0.0287	0.0583	0.0437
machB	0.0019	0.0102	0.0012	0.0004	0.0315	0.0573	0.0470
machC	0.0025	0.0141	0.0016	-0.0017	0.0541	0.0869	0.0616
mach4	0.0039	0.0111	0.0011	0.0012	0.0136	0.0385	0.0329
densityA	0.0060	0.0096	0.0018	-0.0005	0.0425	0.0620	0.0453
densityB	0.0014	0.0097	0.0018	0.0016	0.0258	0.0609	0.0402
densityC	0.0080	0.0299	0.0018	-0.0067	0.0561	0.1083	0.0426
prevalence 50	0.0019	0.0108	0.0019	0.0020	0.0435	0.0684	0.0515
sizingModel	0.0100	0.0300	0.0050	0.0010	0.0550	0.0900	0.0650

The last row of Table S3 shows the variance components in the model that we used to size VIPER. We generally took round numbers that were between the mean of the variance components observed or higher. This can be seen in Figure S2 where the circles show the variance components from the DMIST reader studies. The "x" symbols connected by lines show the variance components in the model that we used to size VIPER.

Variance Components of AUC Differences

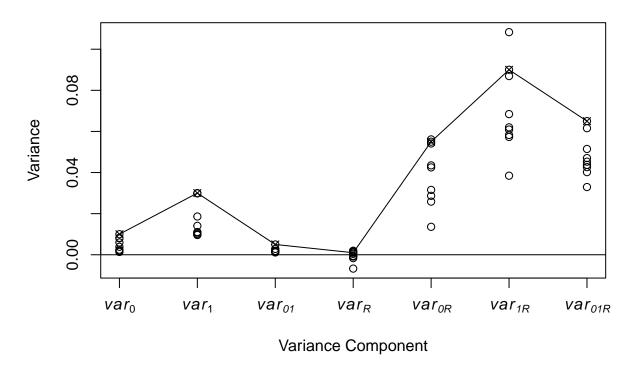


Figure 2: Figure S2: Variance components from DMIST reader studies (circles) and for the model to size the VIPER studies (x's connected by lines).

The original plan for VIPER included a large fully-crossed study (N_0 =400, N_1 =100, N_R =20). Given the model parameters discussed above and Eq. 1, we get the following estimates of uncertainty:

#> Total variance: 0.0004270812
#> Standard error: 0.02066594
#> Half width of 95% CI: 0.04133189

Before VIPER data collection started, a paper about split-plot study designs was published (8), and we decided to design split-plot studies for VIPER. Given the model parameters discussed above and Eq. 2, we get the following estimates of uncertainty for a split-plot study with four groups (N_G =4):

#> Total variance: 0.0005843
#> Standard error: 0.0241723
#> Half width of 95% CI: 0.0483446

The split-plot design was going to save about 75% of the reading time (75% of our costs) and with a moderate impact on precision. With this savings, we wanted to explore more study conditions (prevalence levels and types of non-cancers). We decided to design a split-plot study at ~10% prevalence per reader in which each reader would read N_1 =20 cancers and N_0 =160 non cancers. Given the model parameters discussed above and Eq. 2, we get the following estimates of uncertainty for a split-plot study with four groups (N_G =4):

#> Total variance: 0.0006842188
#> Standard error: 0.02615758
#> Half width of 95% CI: 0.05231515

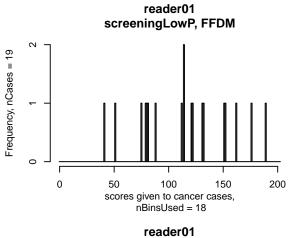
This study design was acceptable except that it required 50% more non-cancer cases than we had originally considered. This increase could not be accommodated by the available cases. After some consideration, we decided that we would split the non-cancer cases into four *overlapping* groups instead of four *independent* groups (See blocks "BIRADS12.FFDM" and "BIRADS12.SFM" in Fig. S1A). Such a study would only require 320 cases instead of 640, and the impact on the uncertainty is negligible. The standard error of the difference in AUCs is insensitive to the number of non-cancers because there are so many non-cancers. The standard error is driven by the number of readers and cancers.

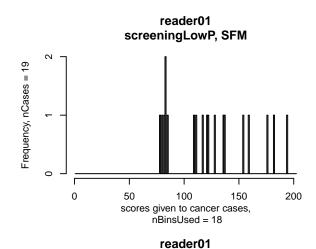
Given our initial budget and the precision achievable by a study at approximately 11% prevalence, we determined that we could conduct five reader studies in total. After splitting our significance level by five (Bonferroni Correction), the power to detect a difference between FFDM and SFM was 0.86. This was acceptable.

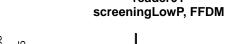
Reader Scores and ROC Curves

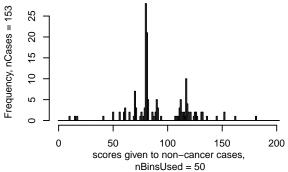
In the following pages we give histograms of the reader scores and the corresponding ROC curves and operating points. They are organized by reader study (screeningLowP, screeningMedP, screeningHighP, challengeMedP, and challengeHighP). Then for each reader study, we show the figures for each reader. In the histograms (one for FFDM and one for SFM), the scores from cancer and non-cancer cases are combined, and we indicate the number of cases (nCases) and the number bins used by each reader (nBinsUsed). The ROC curves are the empirical ROC curves obtained by considering all possible thresholds from 0 to 202. To these curves we add the operating points based on the readers' binary recall decision. The operating points fall on the ROC curves as a result of the two step scoring method summarized in the main paper.

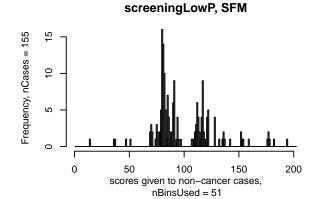
screeningLowP



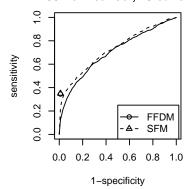


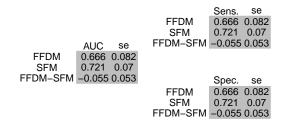


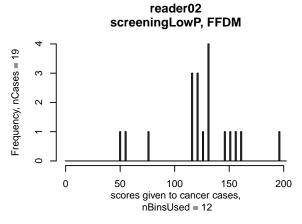




reader01, screeningLowP 153 non-cancer, 19 cancer







Frequency, nCases = 155

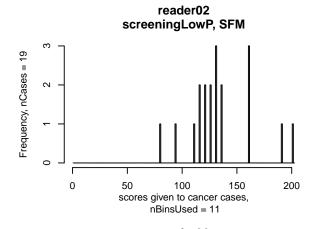
10 ω

9 4

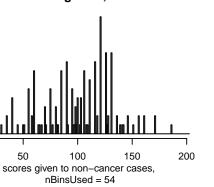
0 0

0

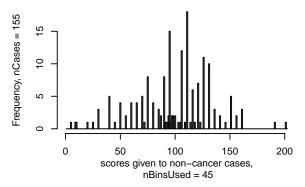
50





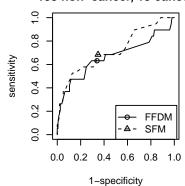


reader02 screeningLowP, SFM

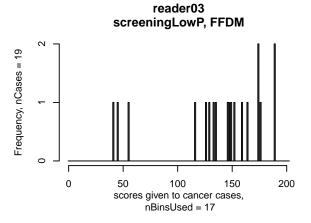


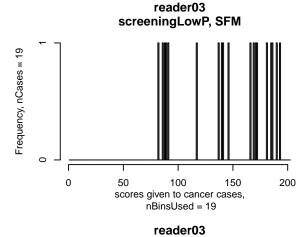
reader02, screeningLowP 155 non-cancer, 19 cancer

100

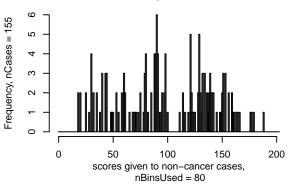


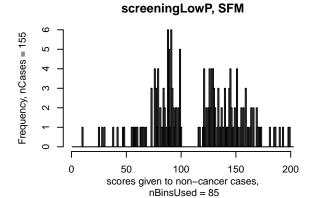




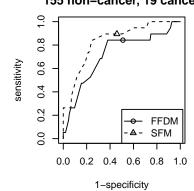


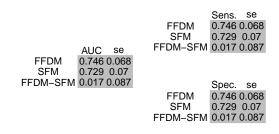


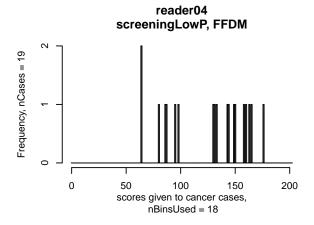


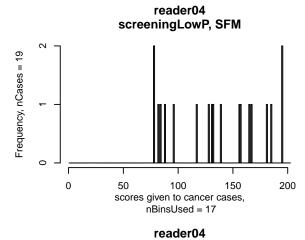


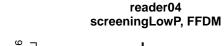
reader03, screeningLowP 155 non-cancer, 19 cancer

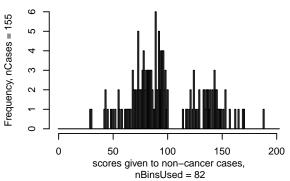


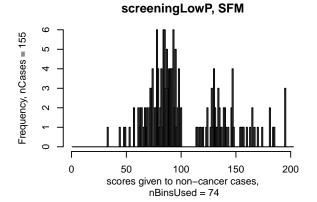




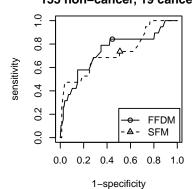


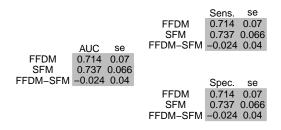


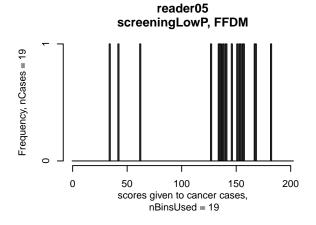


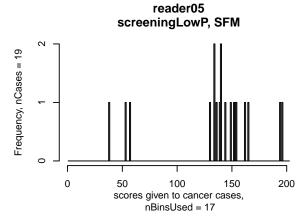


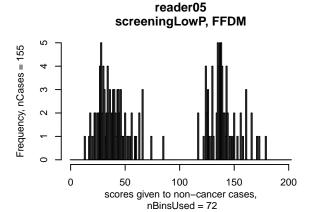
reader04, screeningLowP 155 non-cancer, 19 cancer

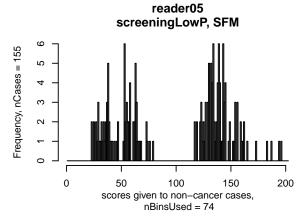


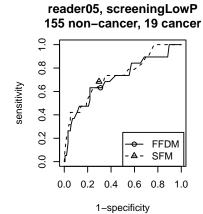


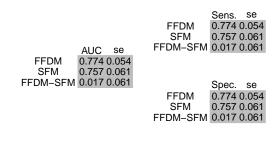


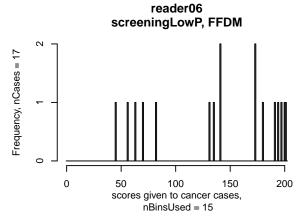


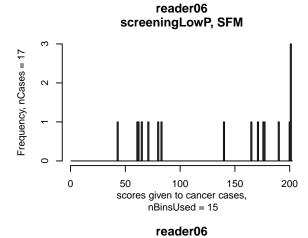


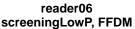


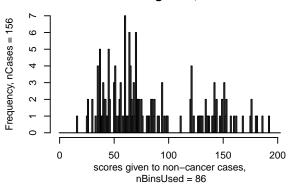


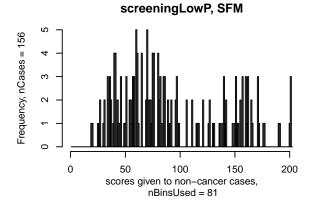




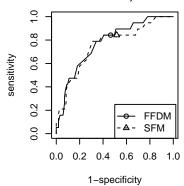


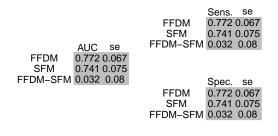


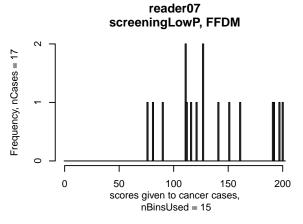


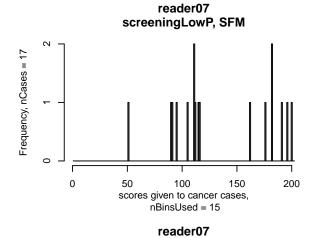


reader06, screeningLowP 156 non-cancer, 17 cancer

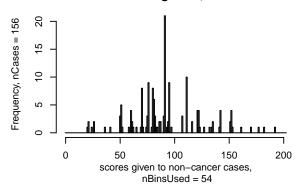


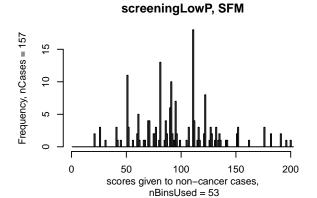




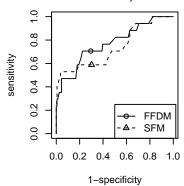


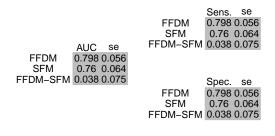


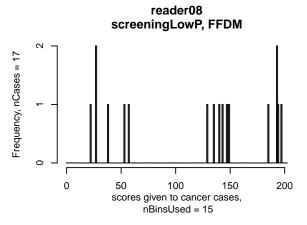


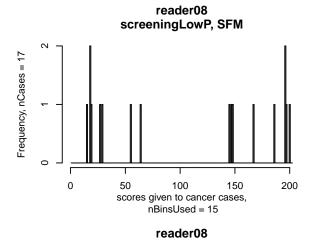


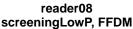
reader07, screeningLowP 156 non-cancer, 17 cancer

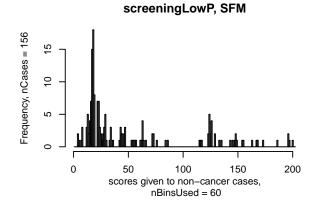


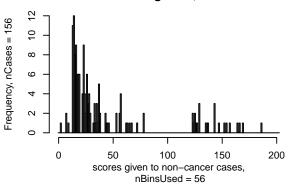


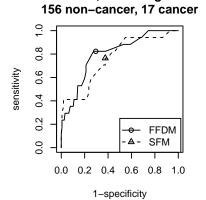




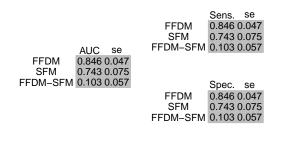


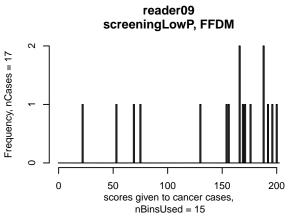


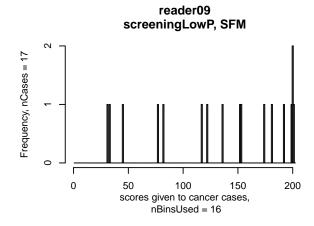




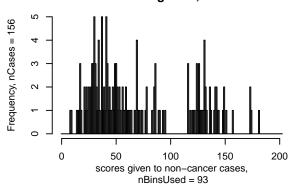
reader08, screeningLowP

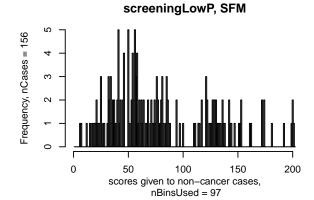






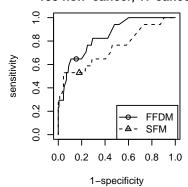


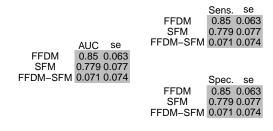


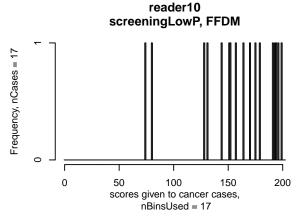


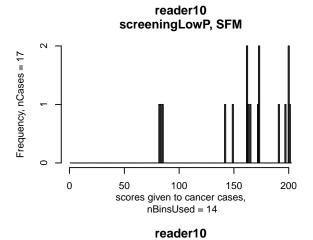
reader09

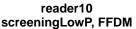
reader09, screeningLowP 155 non-cancer, 17 cancer

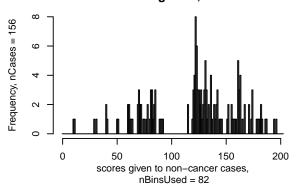




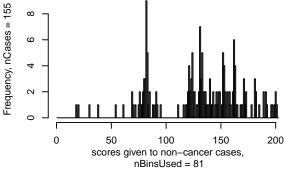




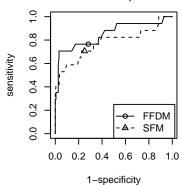


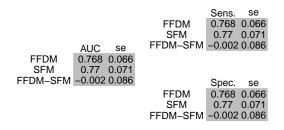


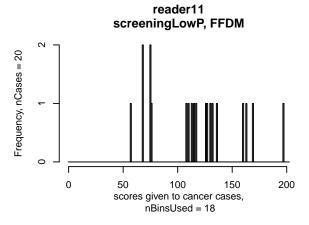


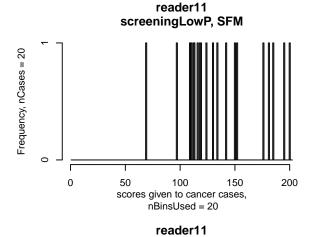


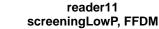
reader10, screeningLowP 155 non-cancer, 17 cancer

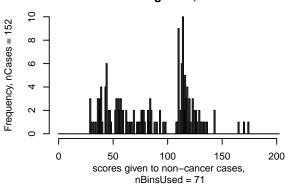


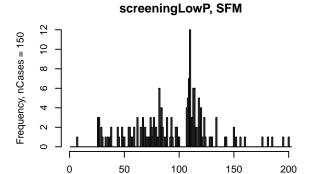






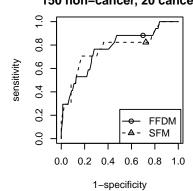


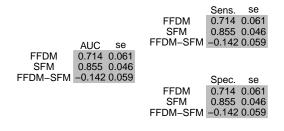


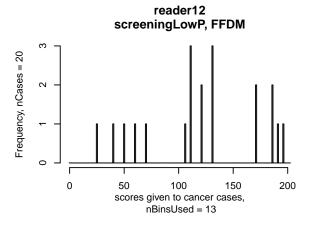


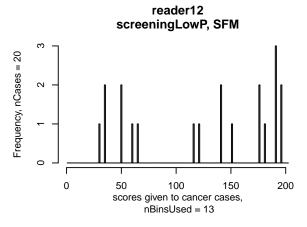
scores given to non-cancer cases, nBinsUsed = 69

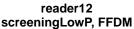
reader11, screeningLowP 150 non-cancer, 20 cancer

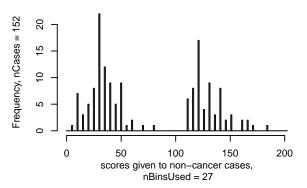


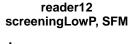


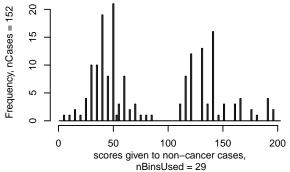




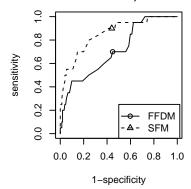


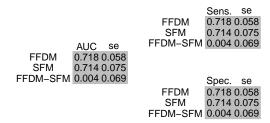


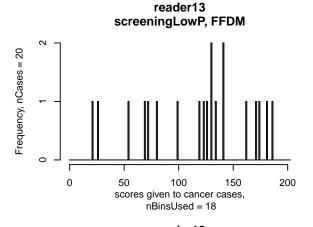


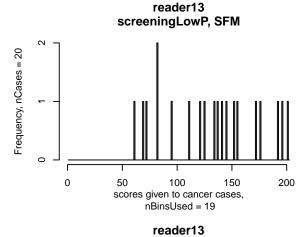


reader12, screeningLowP 152 non-cancer, 20 cancer

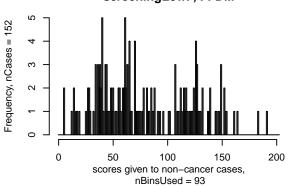


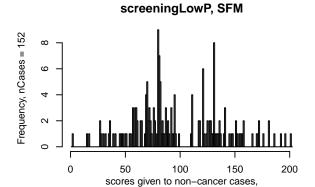






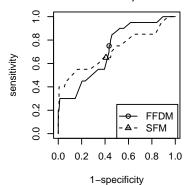


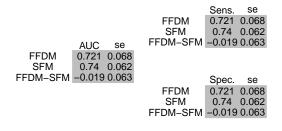


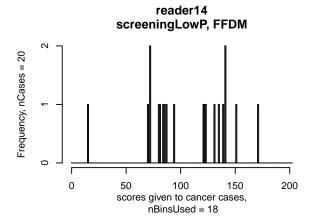


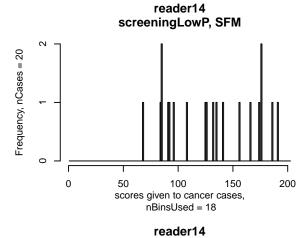
nBinsUsed = 84

reader13, screeningLowP 152 non-cancer, 20 cancer

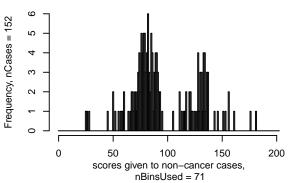


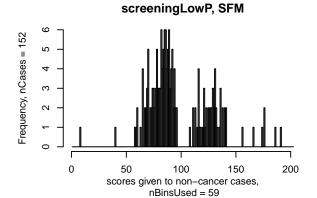




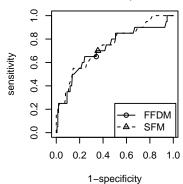


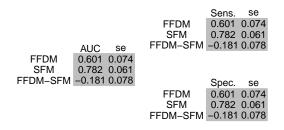


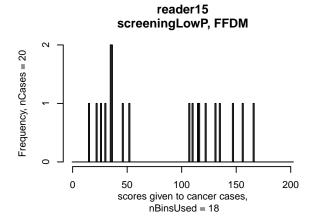


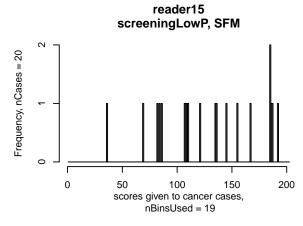


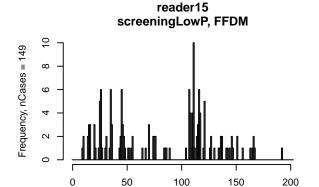
reader14, screeningLowP 152 non-cancer, 20 cancer

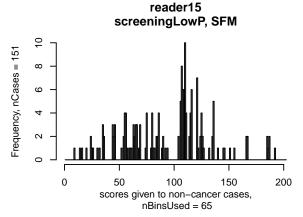


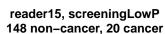




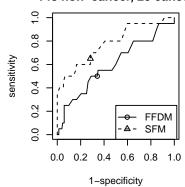


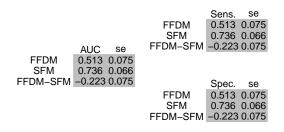


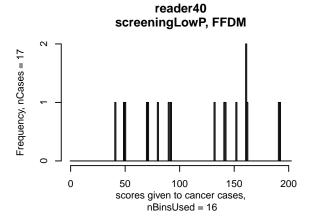


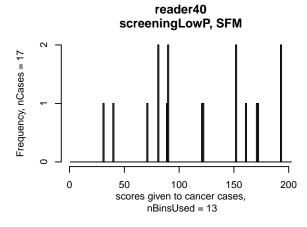


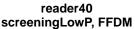
scores given to non-cancer cases, nBinsUsed = 68

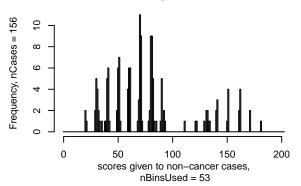


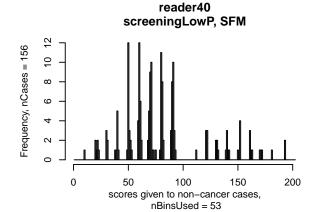




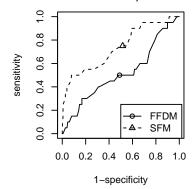


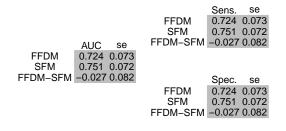


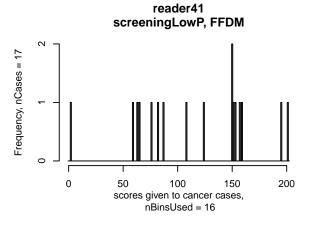


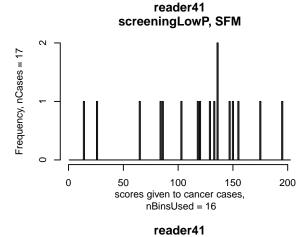


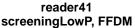
reader40, screeningLowP 156 non-cancer, 17 cancer

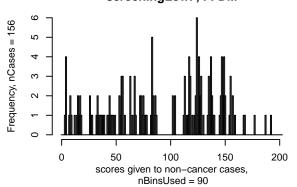


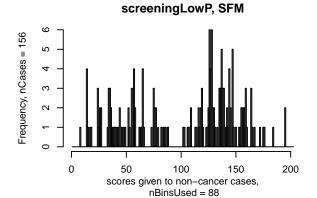




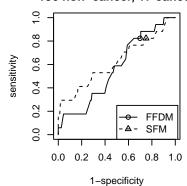


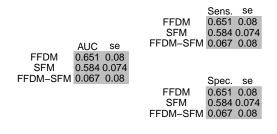


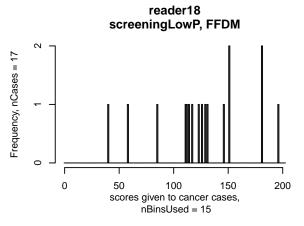


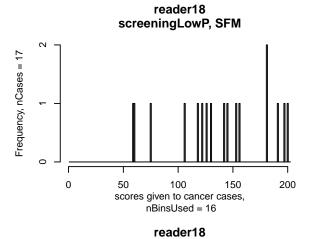


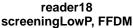
reader41, screeningLowP 156 non-cancer, 17 cancer









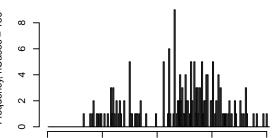


Frequency, nCases = 156

ω

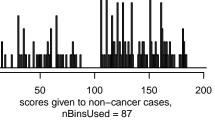
N

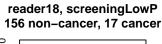


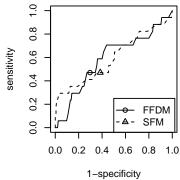


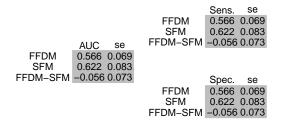
screeningLowP, SFM

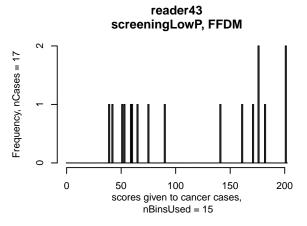
scores given to non-cancer cases, nBinsUsed = 78

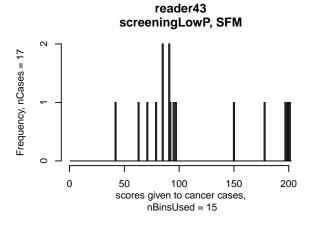




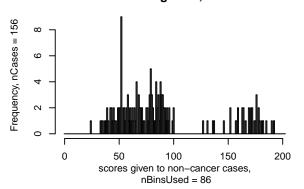


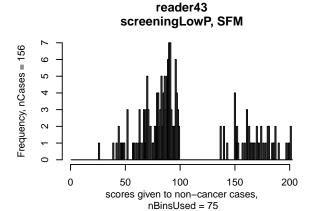




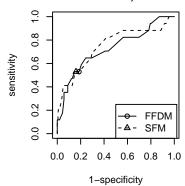


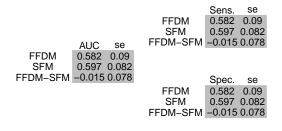


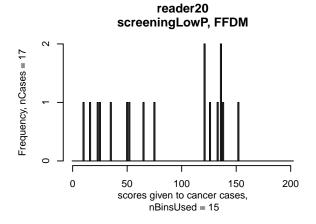


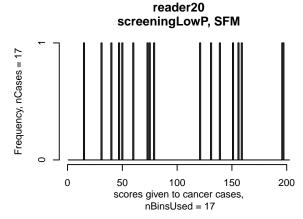


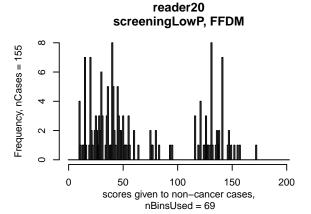
reader43, screeningLowP 156 non-cancer, 17 cancer

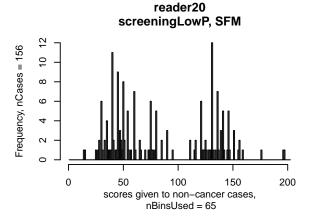


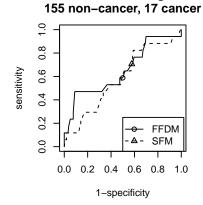




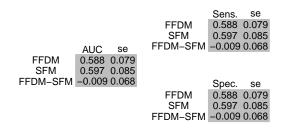




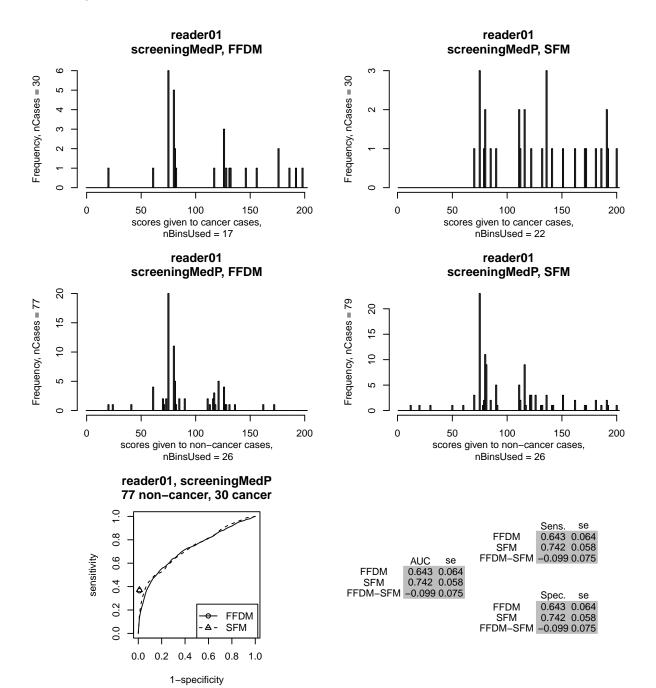


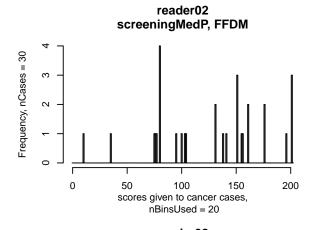


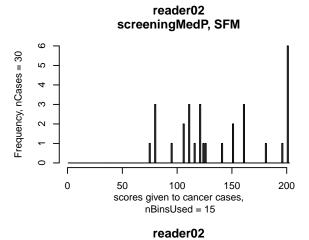
reader20, screeningLowP

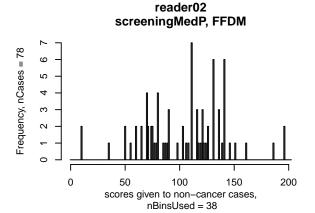


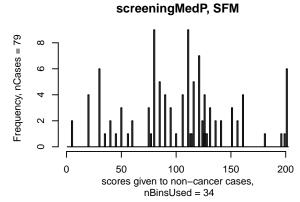
screeningMedP

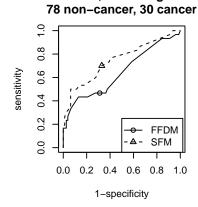




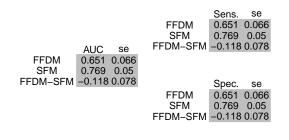


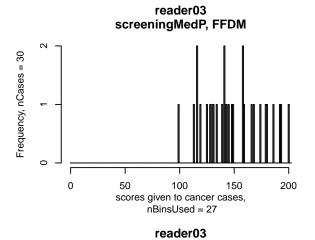


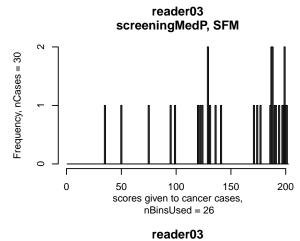


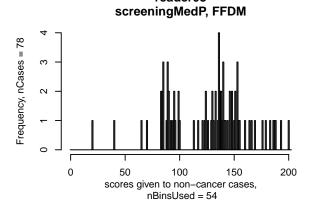


reader02, screeningMedP

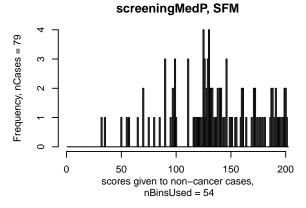


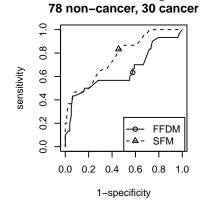


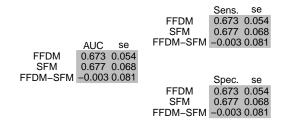


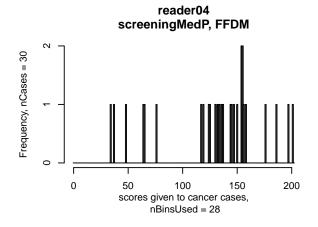


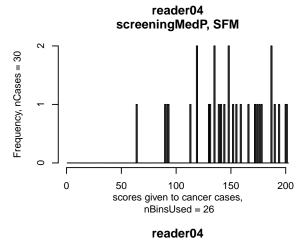
reader03, screeningMedP

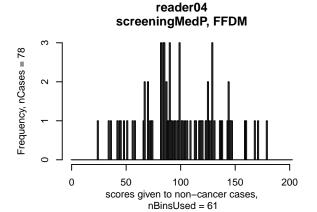


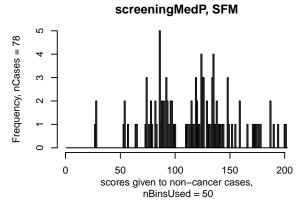


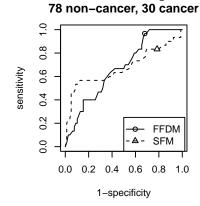




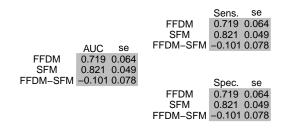


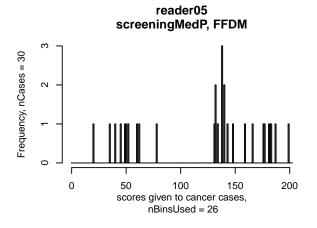


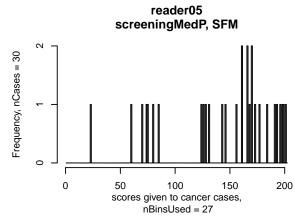


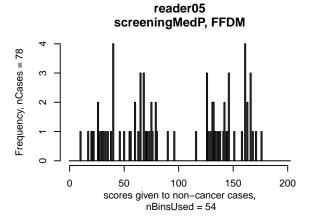


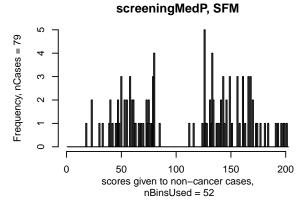
reader04, screeningMedP



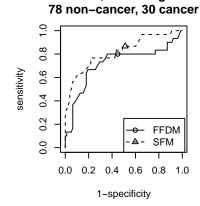




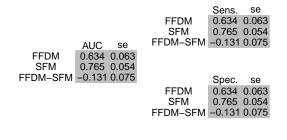


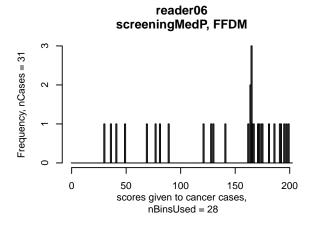


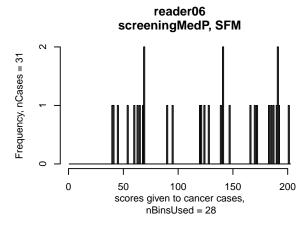
reader05

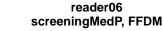


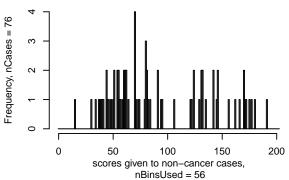
reader05, screeningMedP

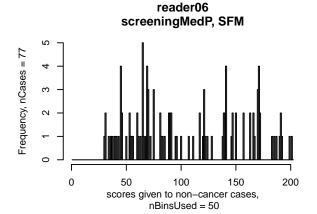




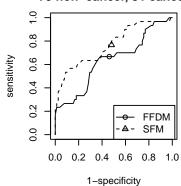


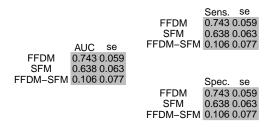


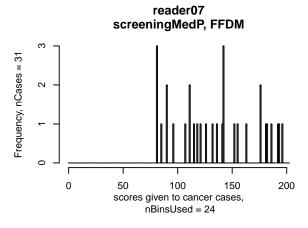


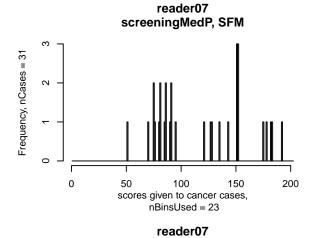


reader06, screeningMedP 75 non-cancer, 31 cancer

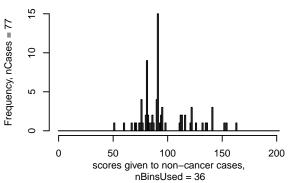


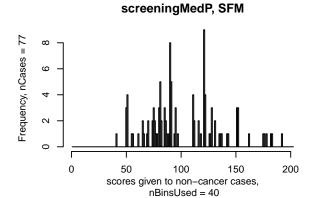




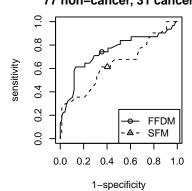


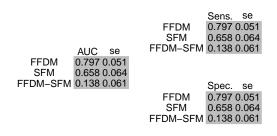


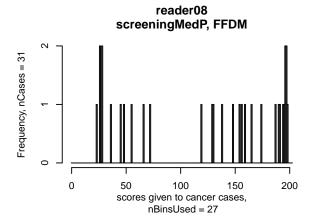


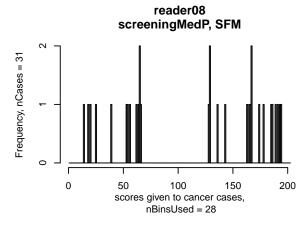


reader07, screeningMedP 77 non-cancer, 31 cancer

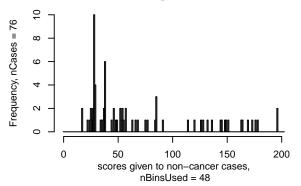


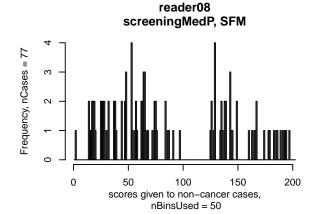




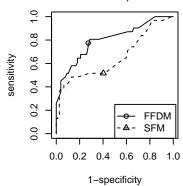


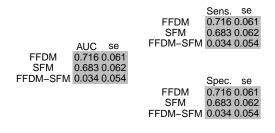


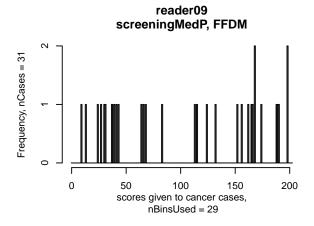


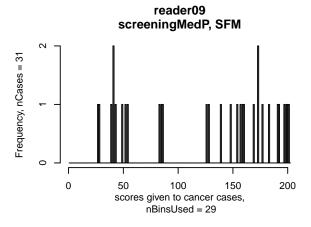


reader08, screeningMedP 76 non-cancer, 31 cancer

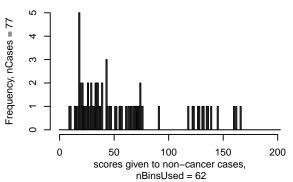


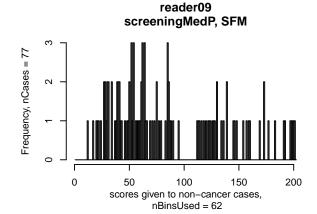




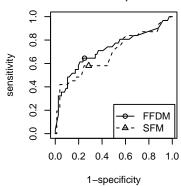


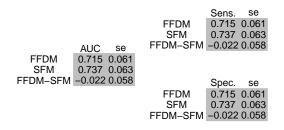


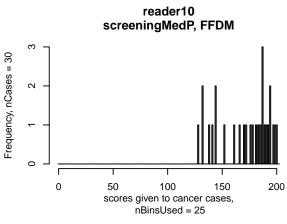


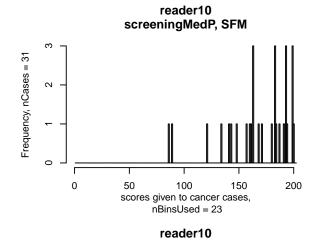


reader09, screeningMedP 77 non-cancer, 31 cancer

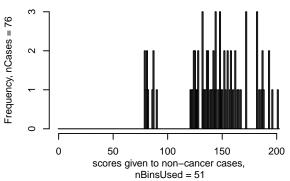


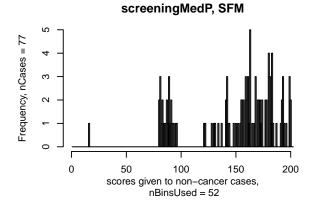




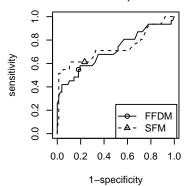




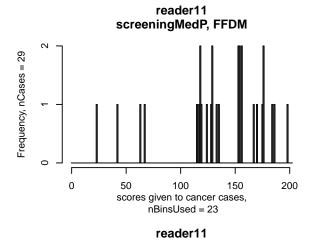


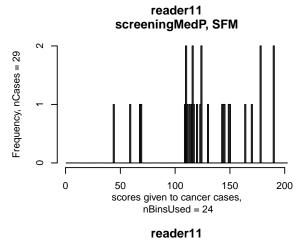


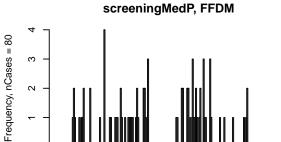
reader10, screeningMedP 76 non-cancer, 30 cancer













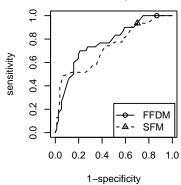
Frequency, nCases = 80
0 1 2 3 4 5

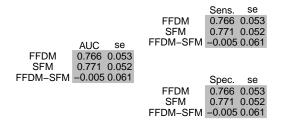
screeningMedP, SFM

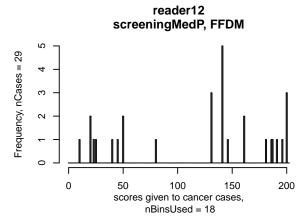
scores given to non-cancer cases, nBinsUsed = 55

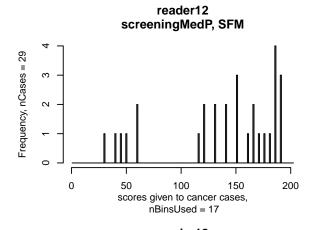
reader11, screeningMedP 80 non-cancer, 29 cancer

scores given to non-cancer cases, nBinsUsed = 55

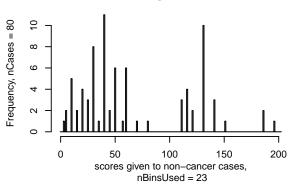


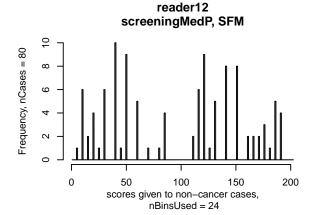




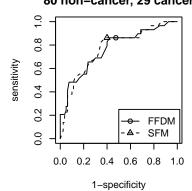


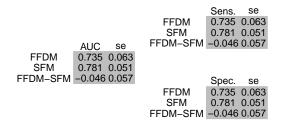


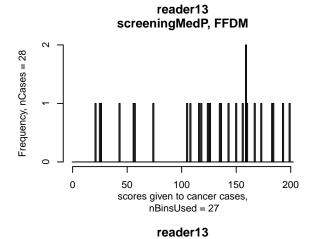


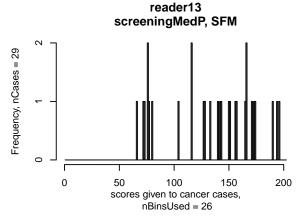


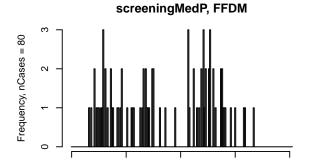
reader12, screeningMedP 80 non-cancer, 29 cancer









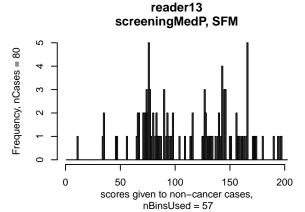


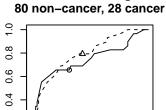
50

0

sensitivity

0.2





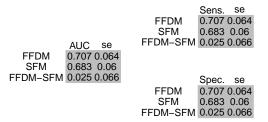
100

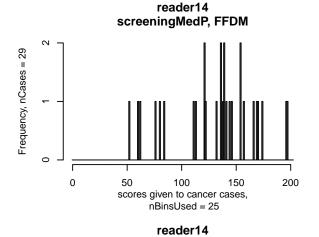
scores given to non-cancer cases, nBinsUsed = 57

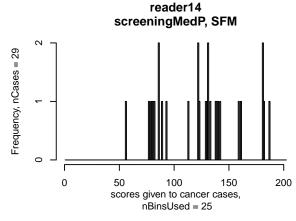
reader13, screeningMedP

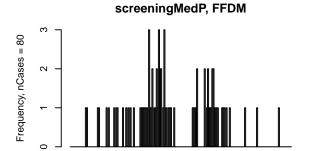
150

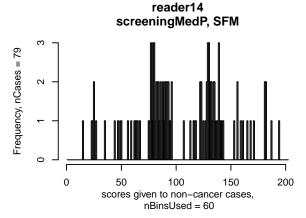
200

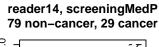






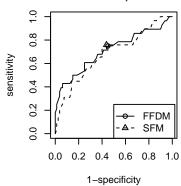


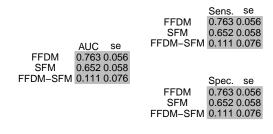


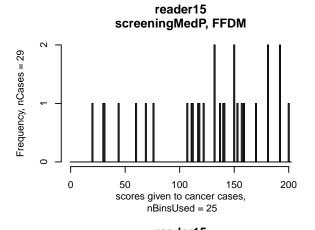


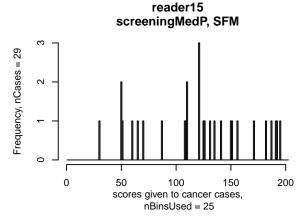
scores given to non-cancer cases,

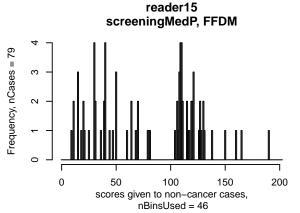
nBinsUsed = 65



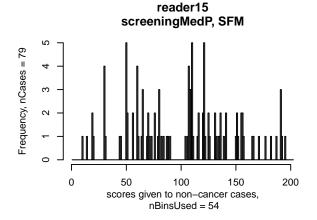


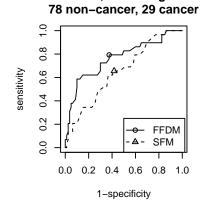


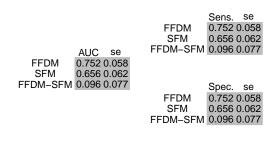


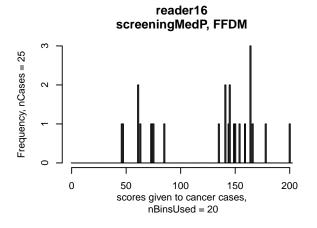


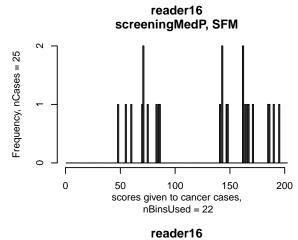
reader15, screeningMedP

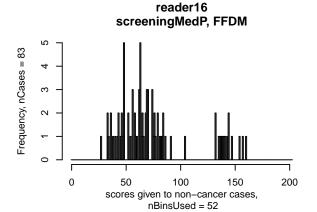


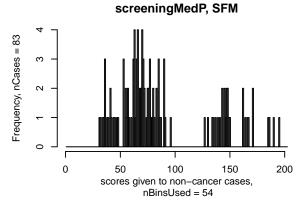


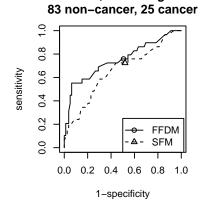




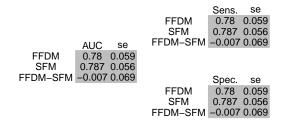


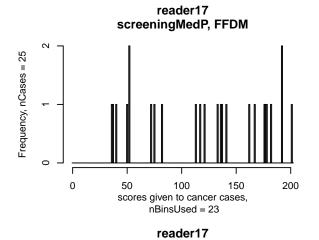


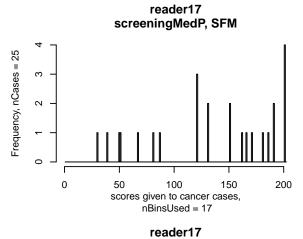


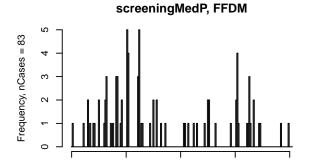


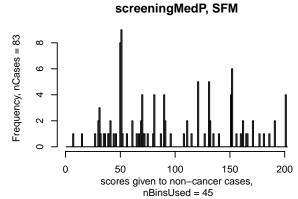
reader16, screeningMedP

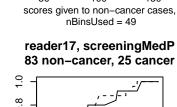


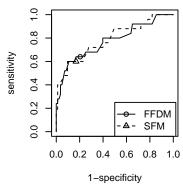


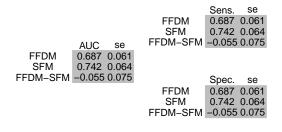


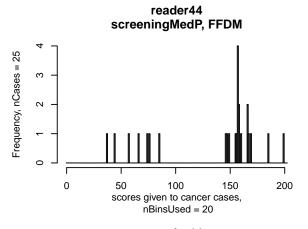


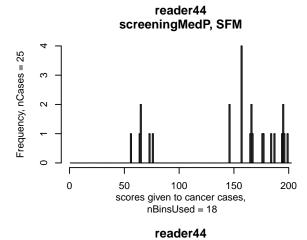


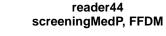


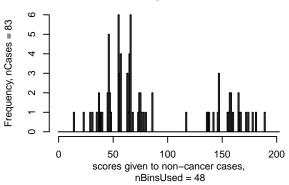


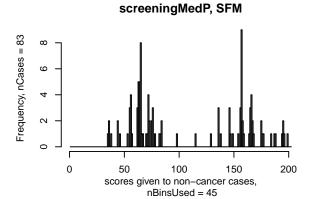




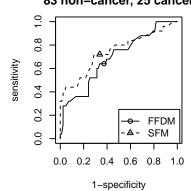


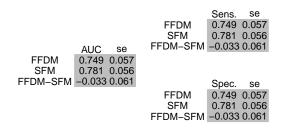


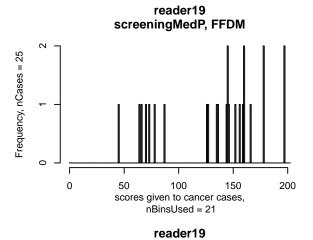


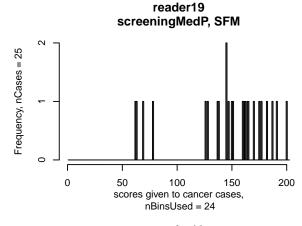


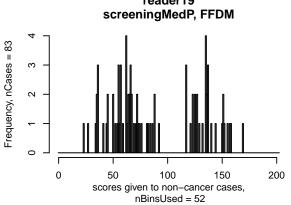
reader44, screeningMedP 83 non-cancer, 25 cancer

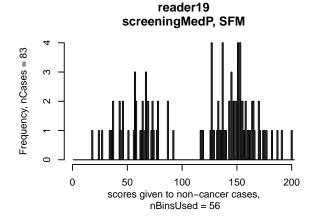


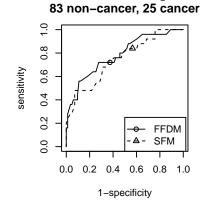






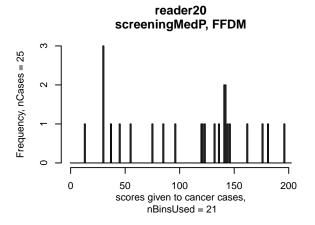


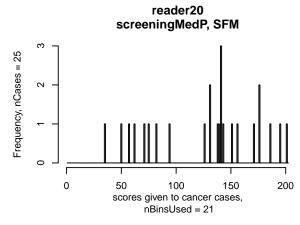


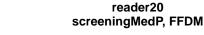


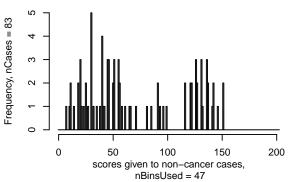
reader19, screeningMedP

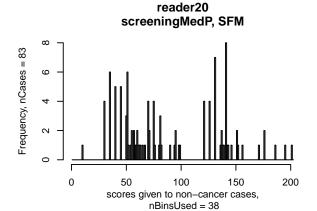




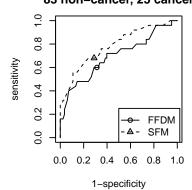


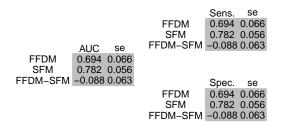




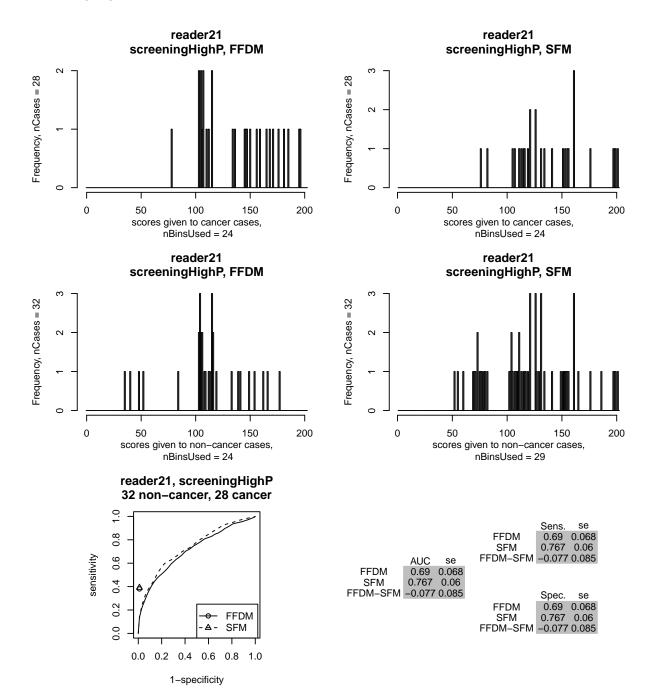


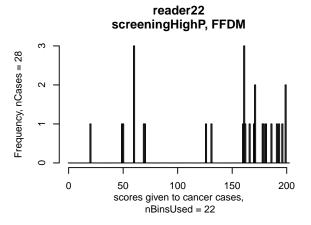
reader20, screeningMedP 83 non-cancer, 25 cancer

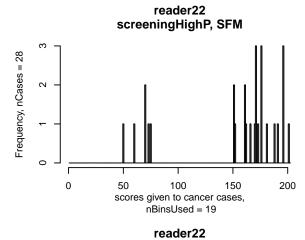




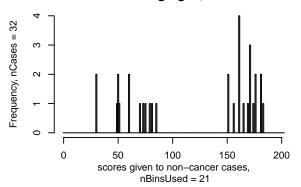
screeningHighP

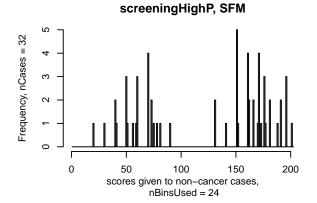




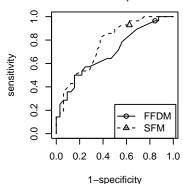


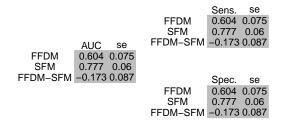


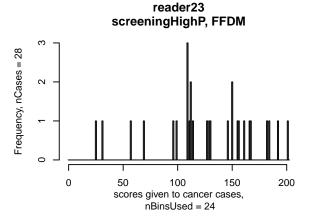


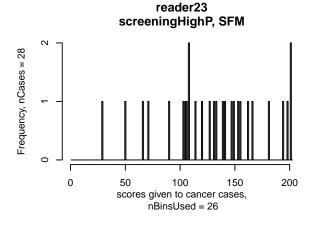


reader22, screeningHighP 32 non-cancer, 28 cancer

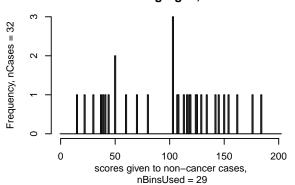


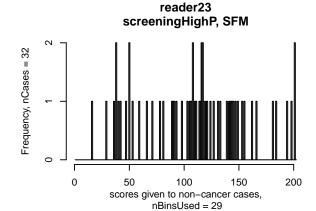




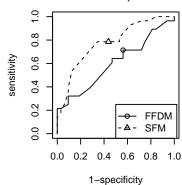


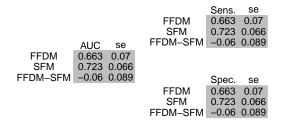


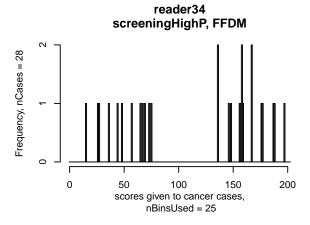


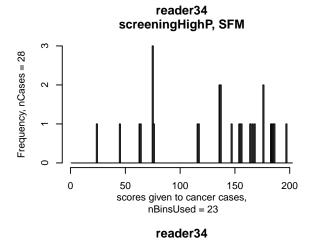


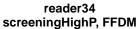
reader23, screeningHighP 32 non-cancer, 28 cancer

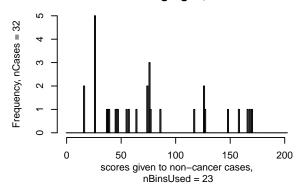


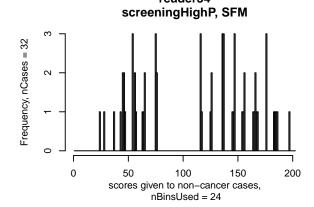




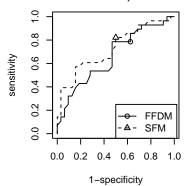


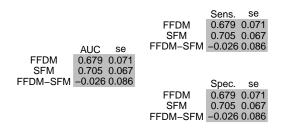


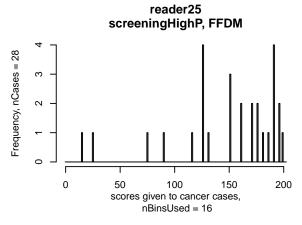


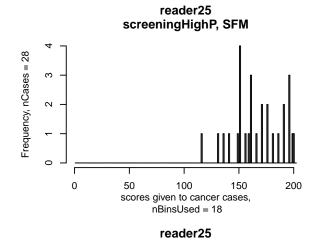


reader34, screeningHighP 32 non-cancer, 28 cancer

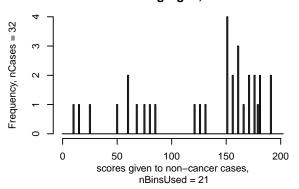


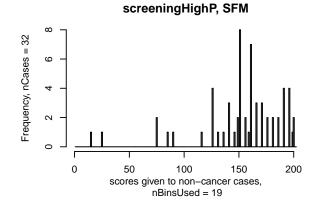




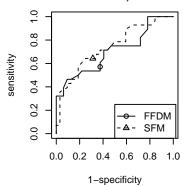




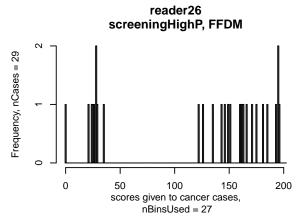


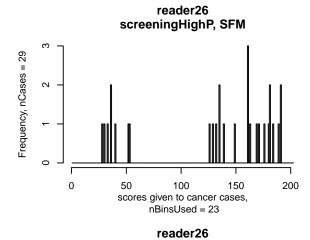


reader25, screeningHighP 32 non-cancer, 28 cancer

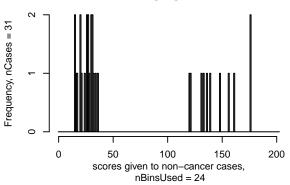


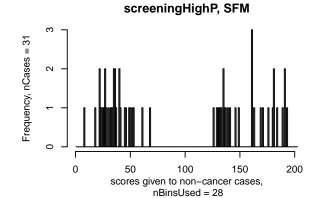




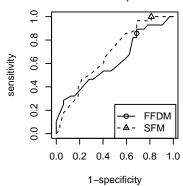


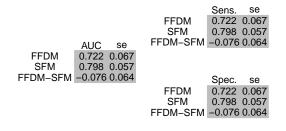


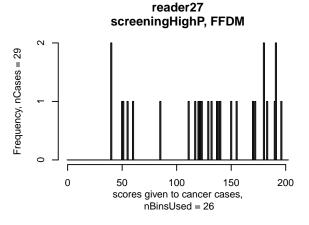


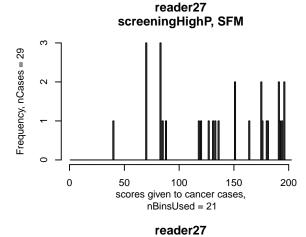


reader26, screeningHighP 31 non-cancer, 29 cancer

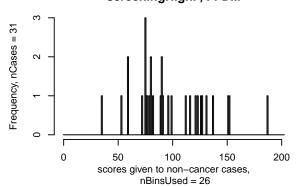


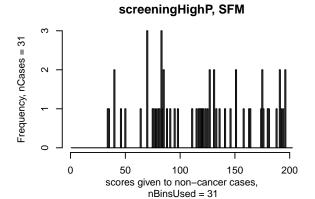




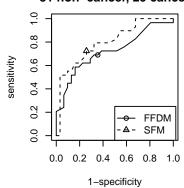


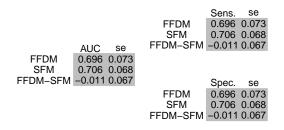


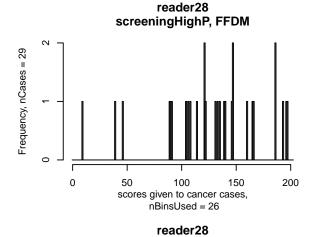


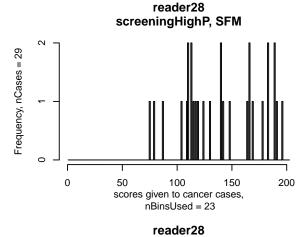


reader27, screeningHighP 31 non-cancer, 29 cancer

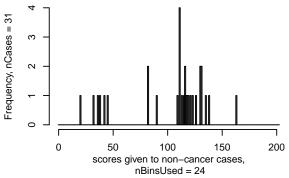


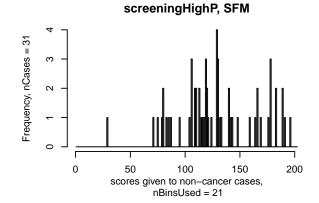




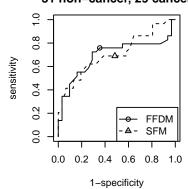


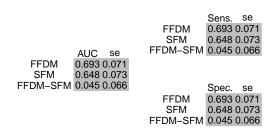


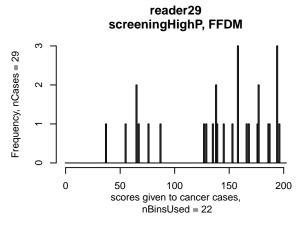


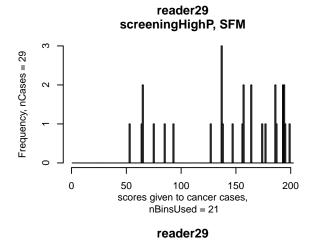


reader28, screeningHighP 31 non-cancer, 29 cancer

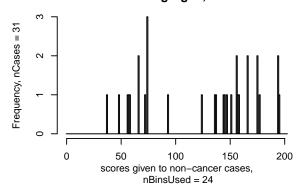


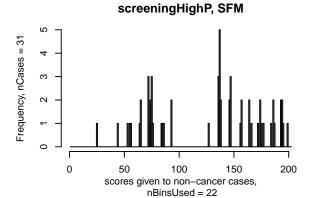




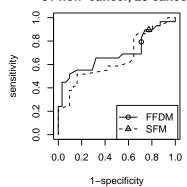


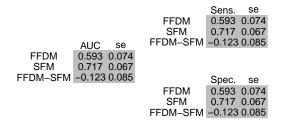


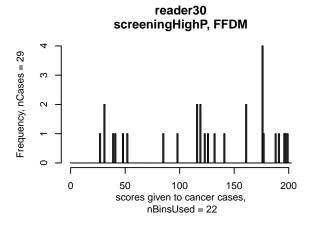


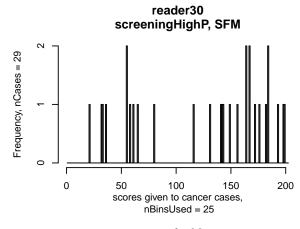


reader29, screeningHighP 31 non-cancer, 29 cancer

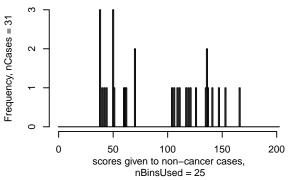


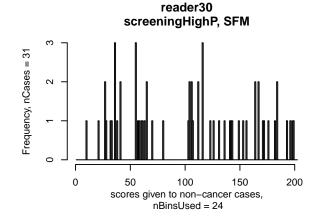




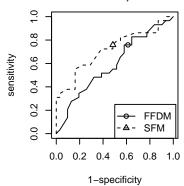


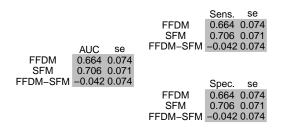


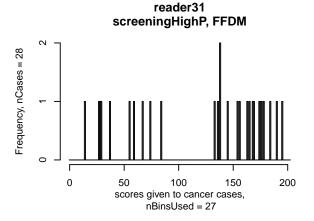


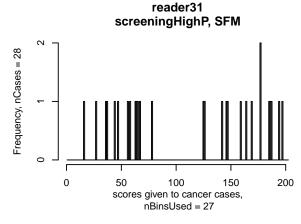


reader30, screeningHighP 31 non-cancer, 29 cancer

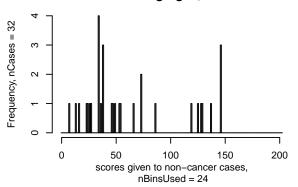


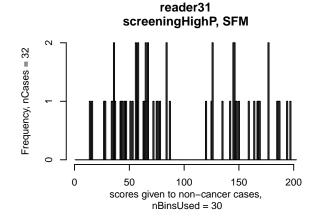




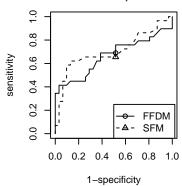


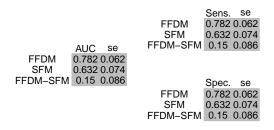


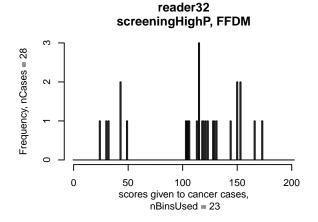


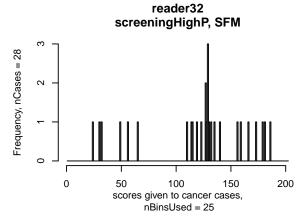


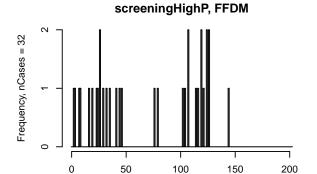
reader31, screeningHighP 32 non-cancer, 28 cancer



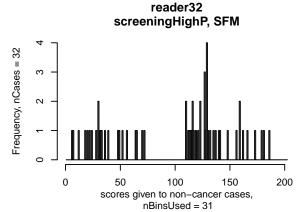


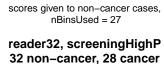


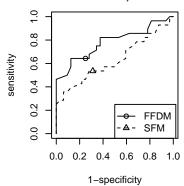


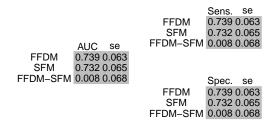


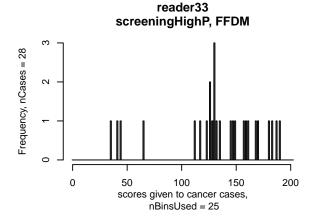
reader32

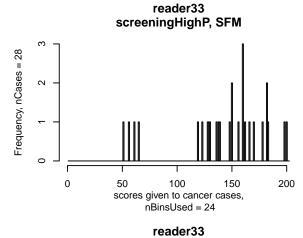


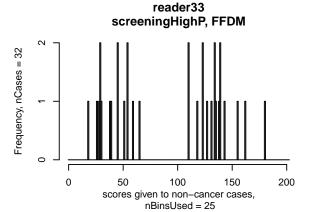


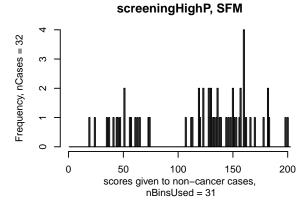


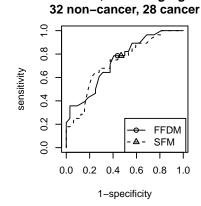




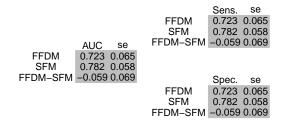


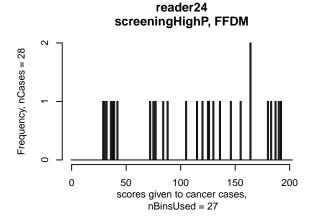


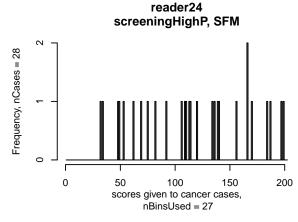


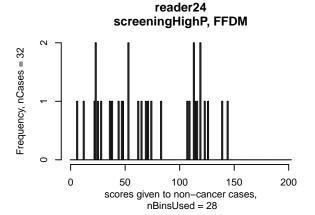


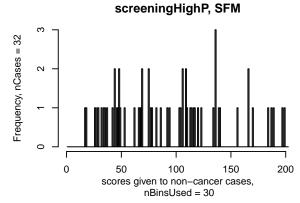
reader33, screeningHighP



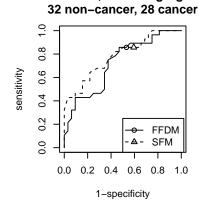




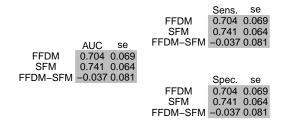


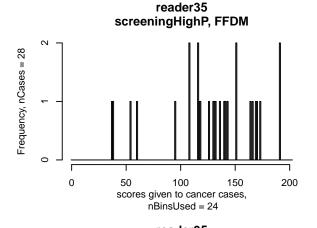


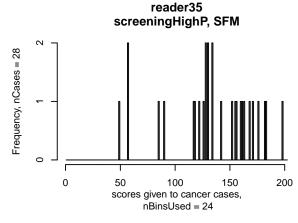
reader24

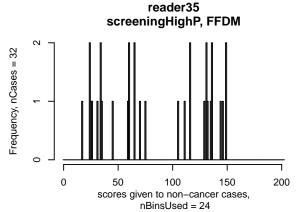


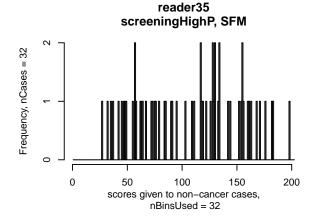
reader24, screeningHighP

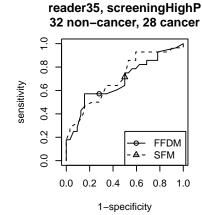


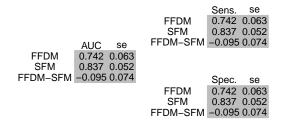


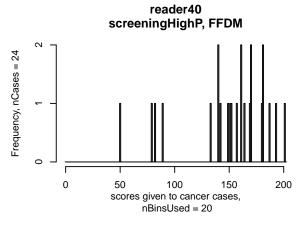


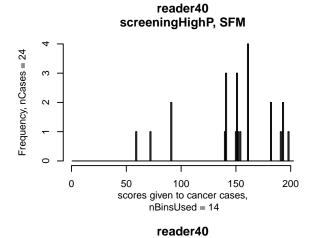




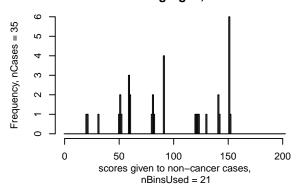


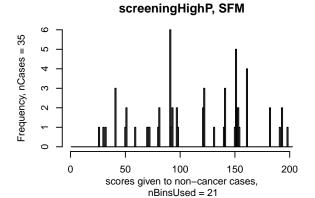




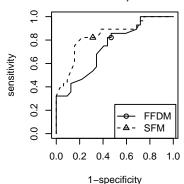


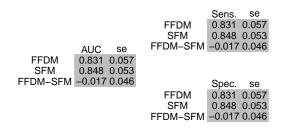


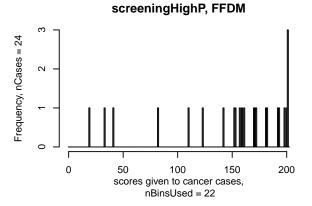




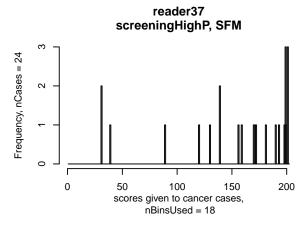
reader40, screeningHighP 35 non-cancer, 24 cancer

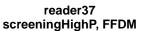


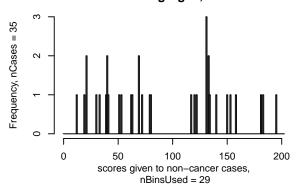




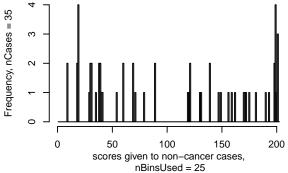
reader37



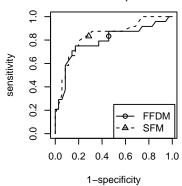


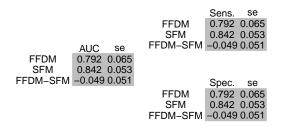


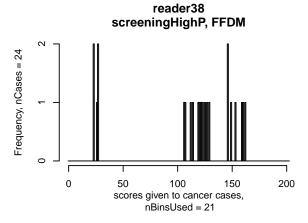


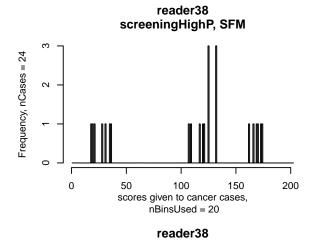


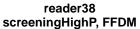
reader37, screeningHighP 35 non-cancer, 24 cancer

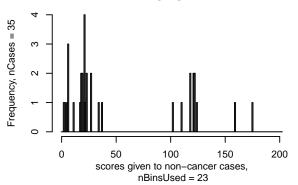


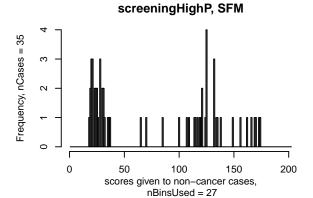




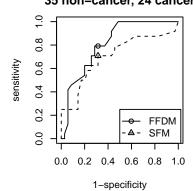


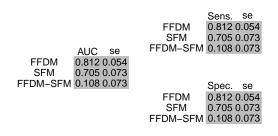


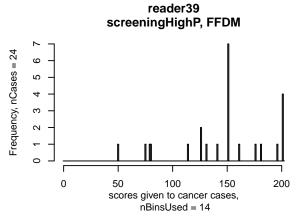


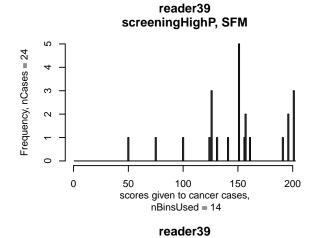


reader38, screeningHighP 35 non-cancer, 24 cancer

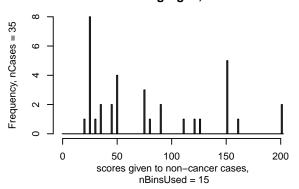


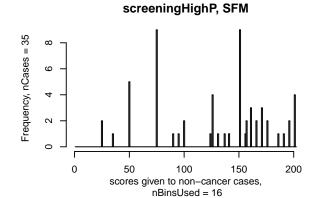




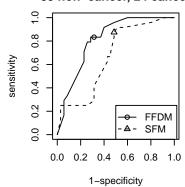


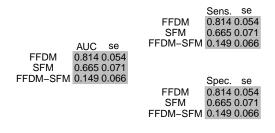


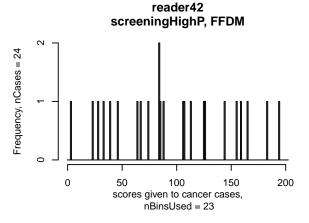


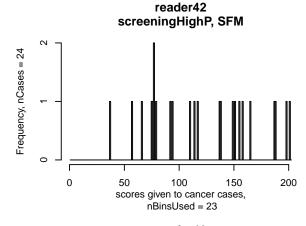


reader39, screeningHighP 35 non-cancer, 24 cancer

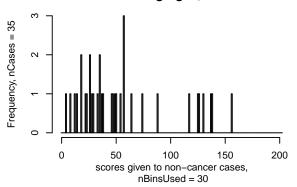


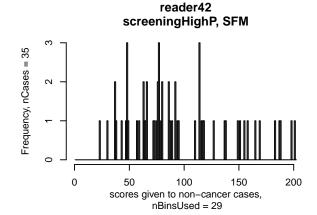




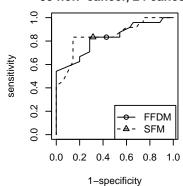


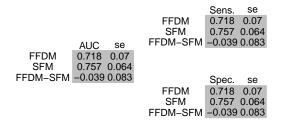




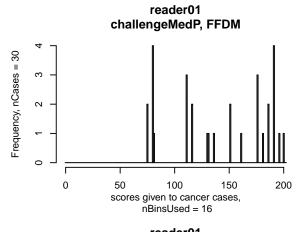


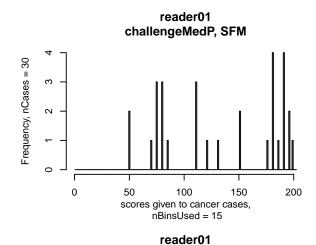
reader42, screeningHighP 35 non-cancer, 24 cancer



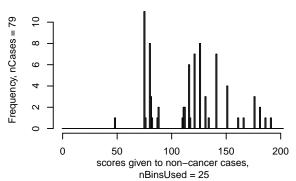


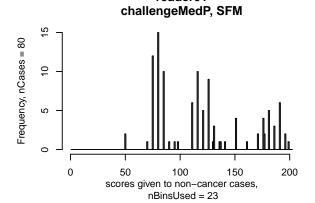
challengeMedP



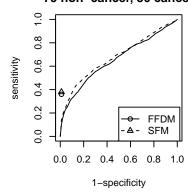


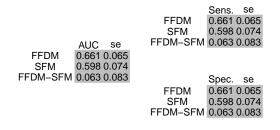


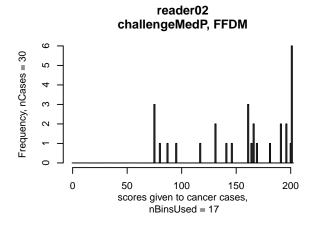


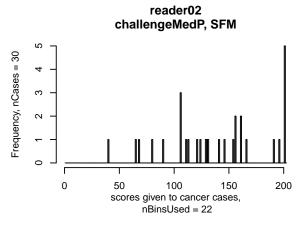


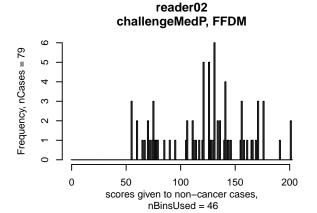
reader01, challengeMedP 79 non-cancer, 30 cancer

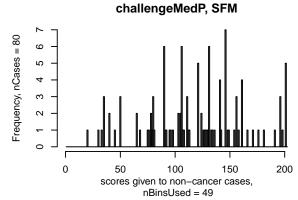




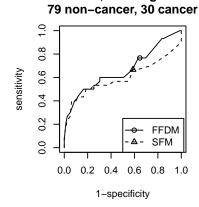




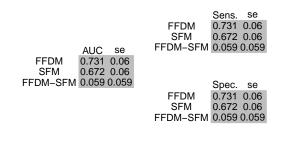


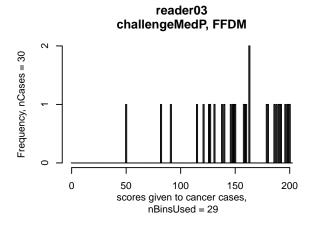


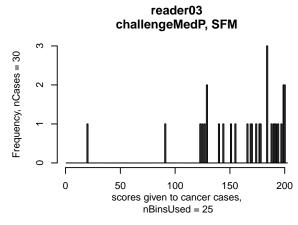
reader02

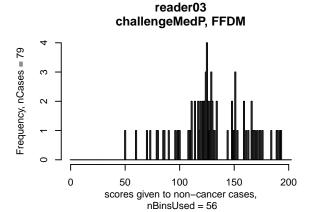


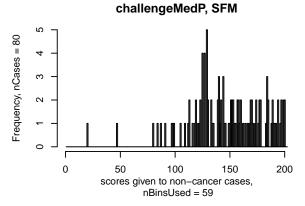
reader02, challengeMedP



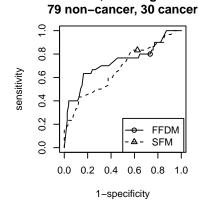




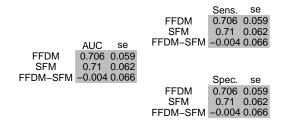


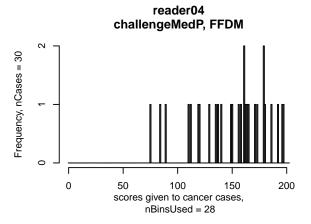


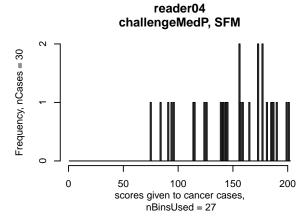
reader03



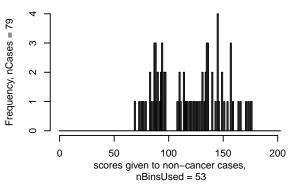
reader03, challengeMedP



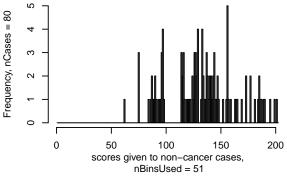




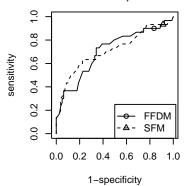


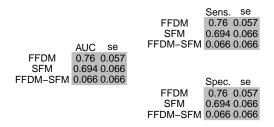


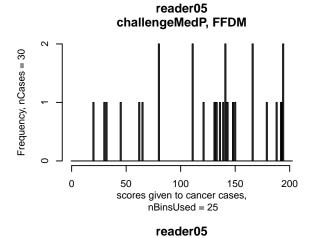


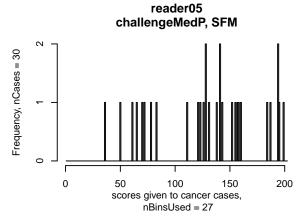


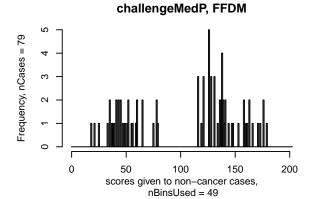
reader04, challengeMedP 79 non-cancer, 30 cancer

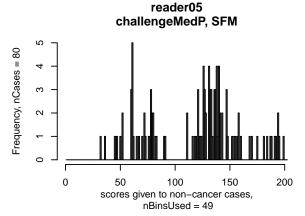


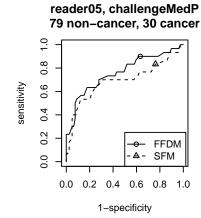


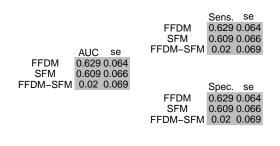


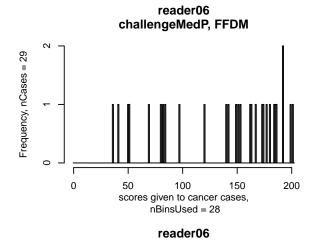


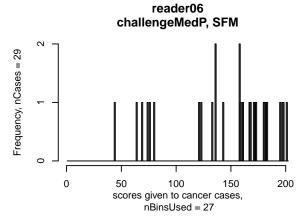


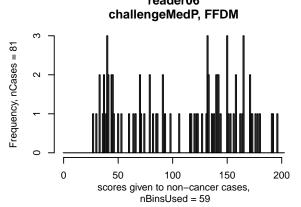


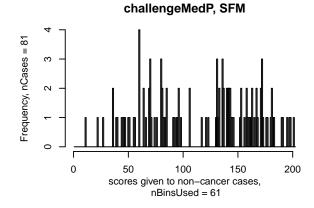




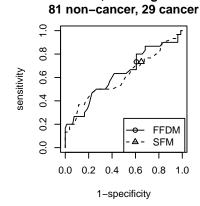




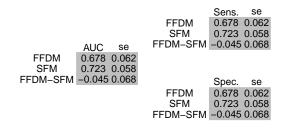


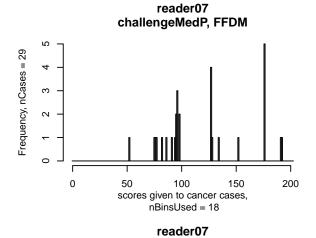


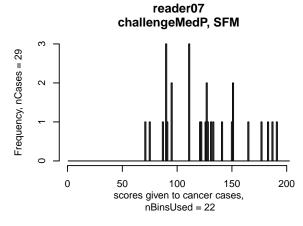
reader06

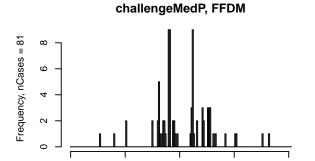


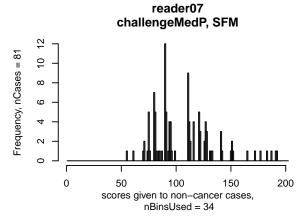
reader06, challengeMedP

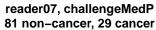




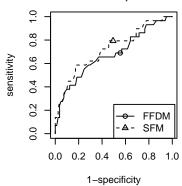


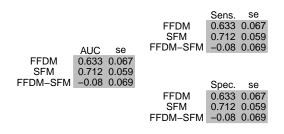


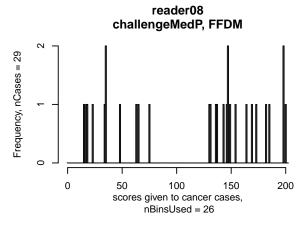


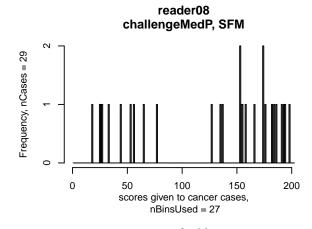


scores given to non-cancer cases, nBinsUsed = 34

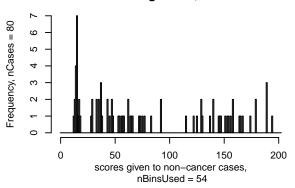


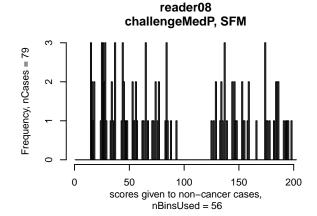




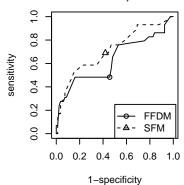


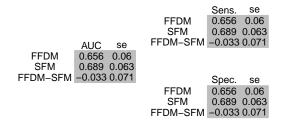


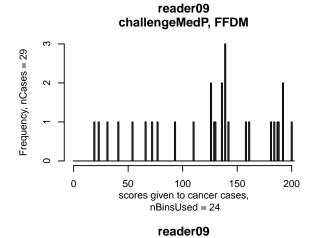


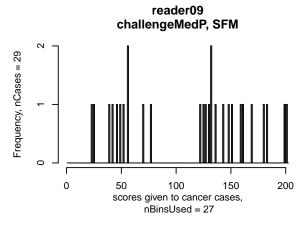


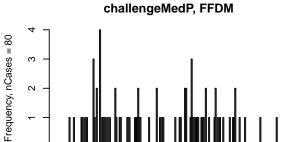
reader08, challengeMedP 78 non-cancer, 29 cancer

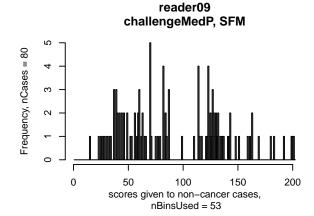


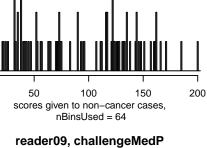


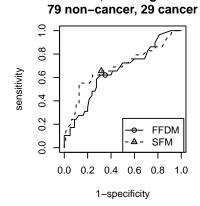


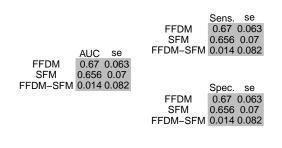


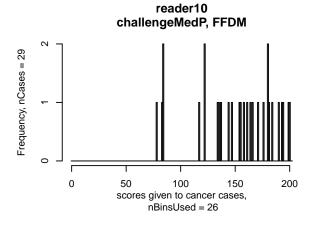


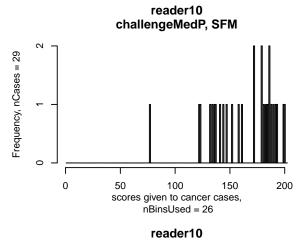


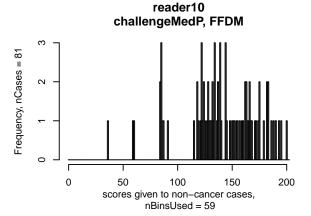




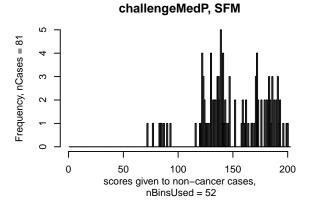


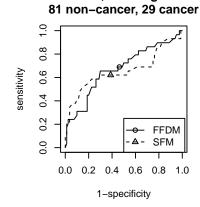


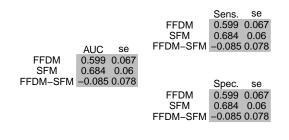


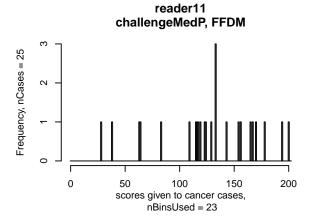


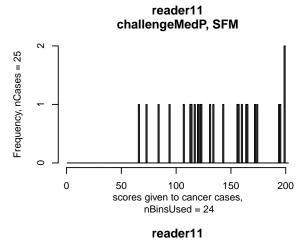
reader10, challengeMedP



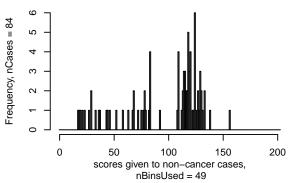


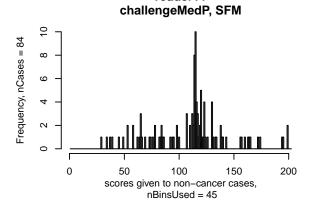




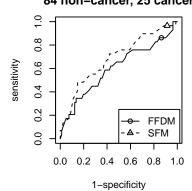


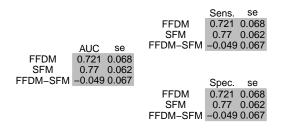


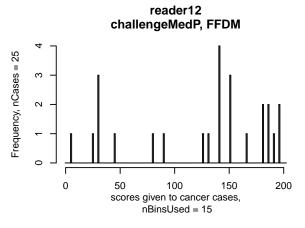


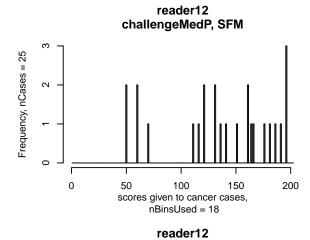


reader11, challengeMedP 84 non-cancer, 25 cancer

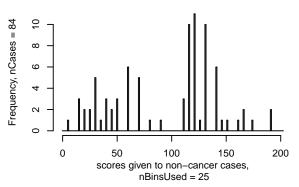


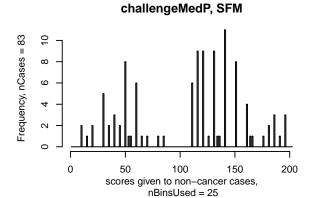




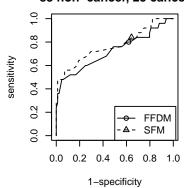


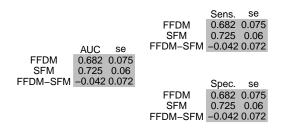


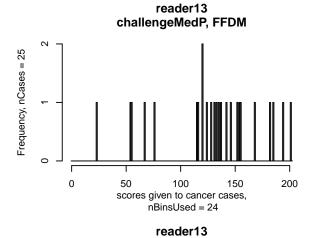


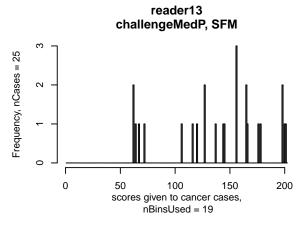


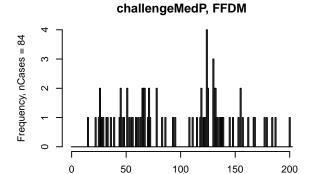
reader12, challengeMedP 83 non-cancer, 25 cancer

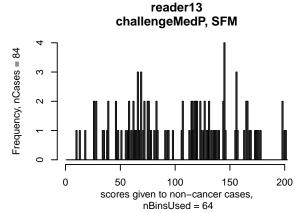


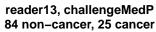




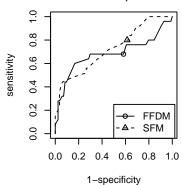


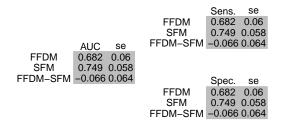


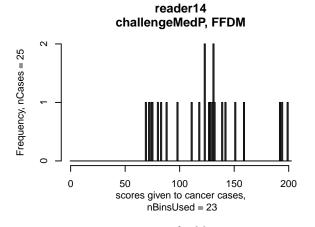


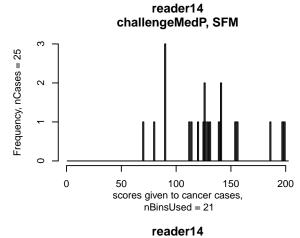


scores given to non-cancer cases, nBinsUsed = 66

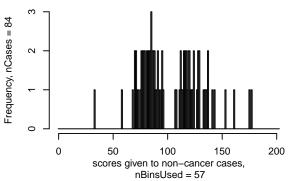


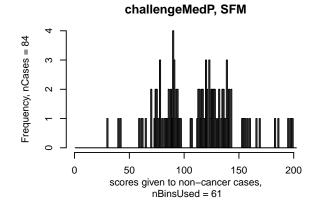




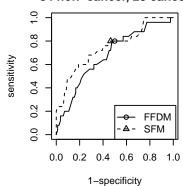


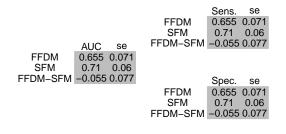


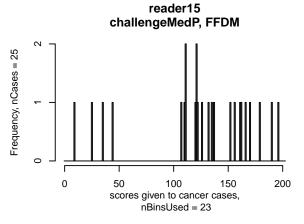


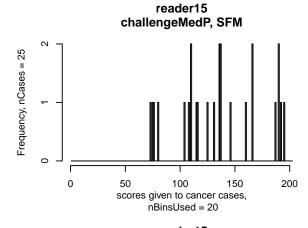


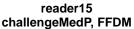
reader14, challengeMedP 84 non-cancer, 25 cancer

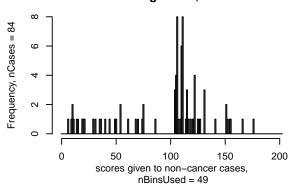


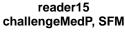


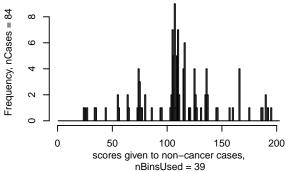




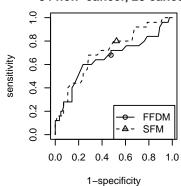


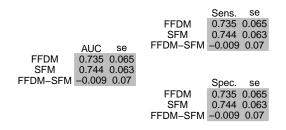


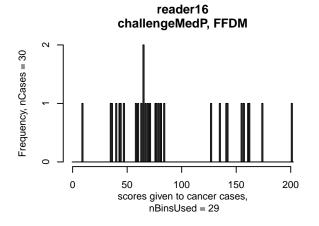


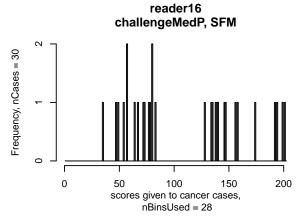


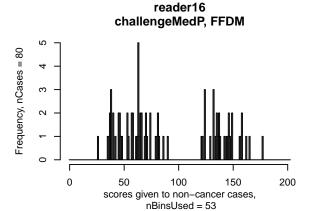
reader15, challengeMedP 84 non-cancer, 25 cancer

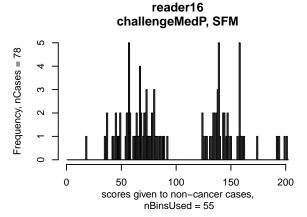


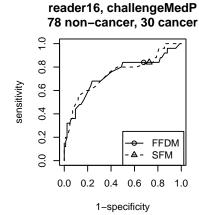


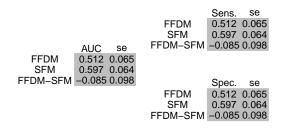


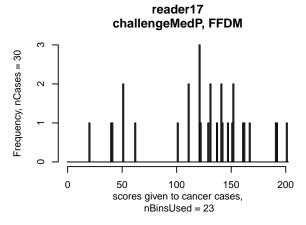


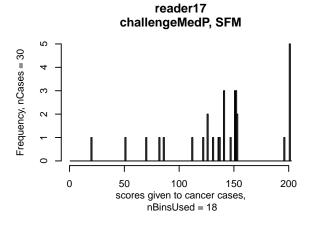


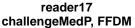


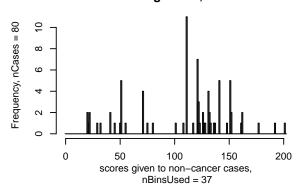


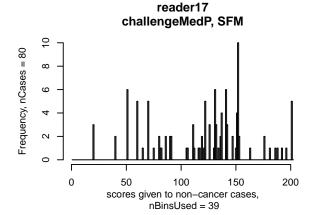




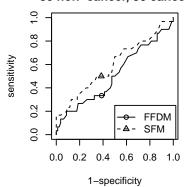


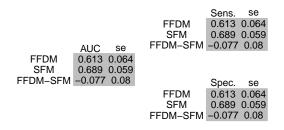


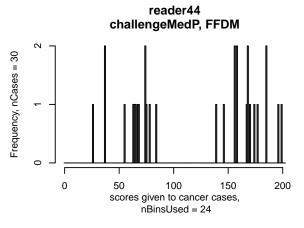


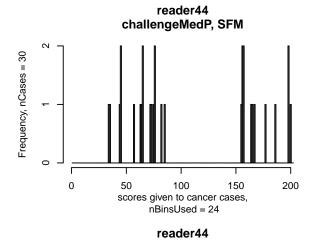


reader17, challengeMedP 80 non-cancer, 30 cancer

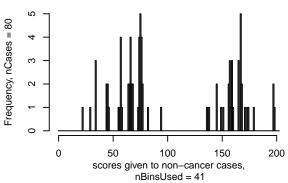


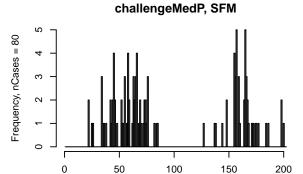






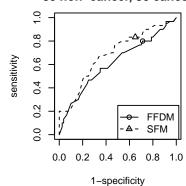


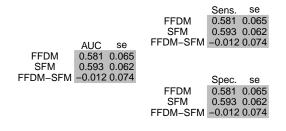


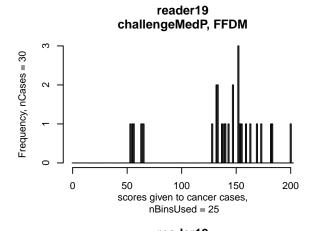


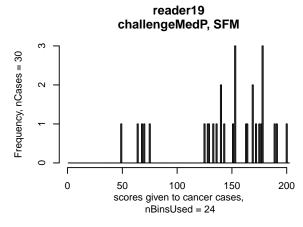
scores given to non-cancer cases, nBinsUsed = 50

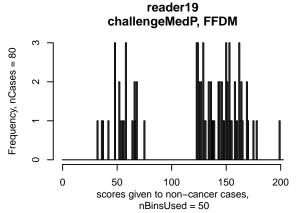
reader44, challengeMedP 80 non-cancer, 30 cancer

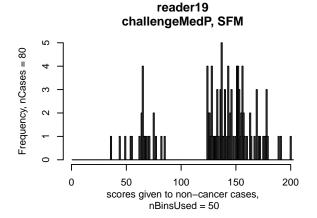


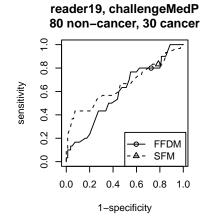


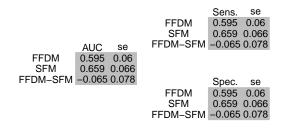


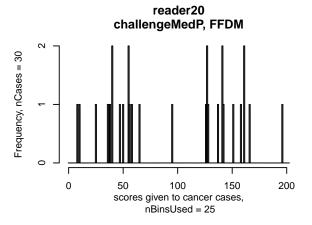


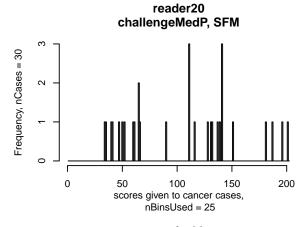


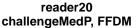


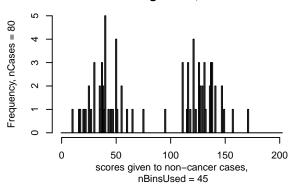


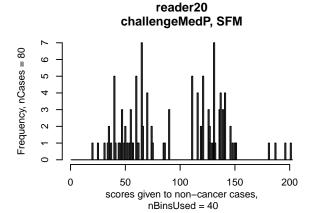




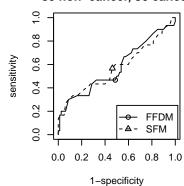


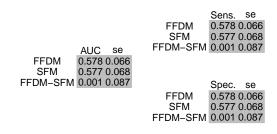




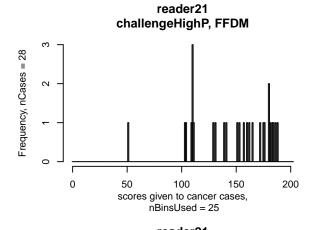


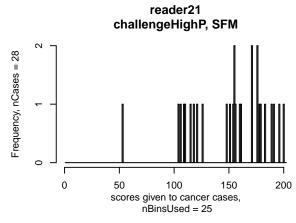
reader20, challengeMedP 80 non-cancer, 30 cancer

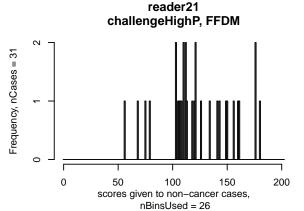


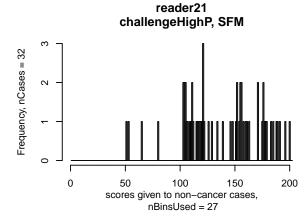


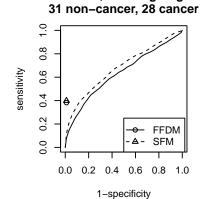
challengeHighP



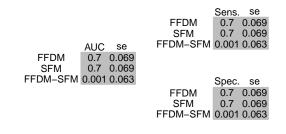


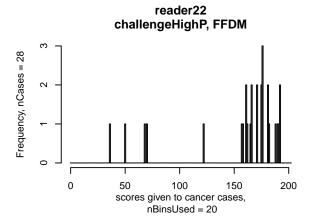


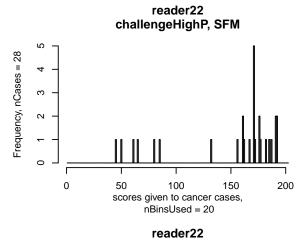




reader21, challengeHighP

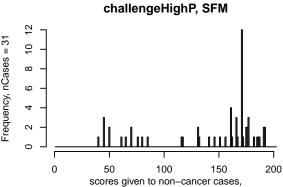




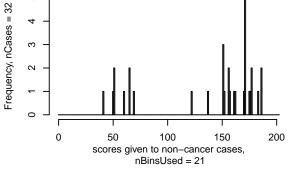


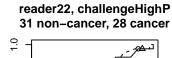


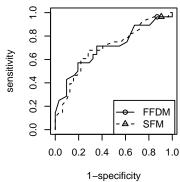


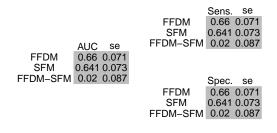


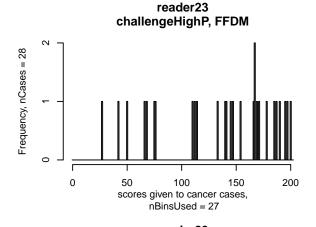
nBinsUsed = 17

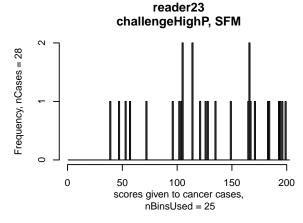


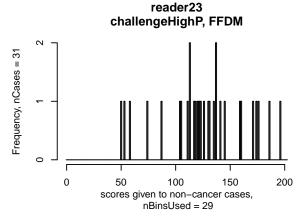


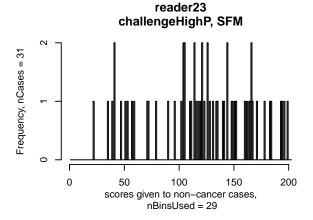


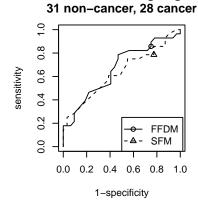




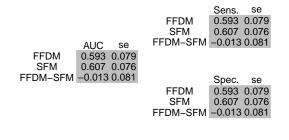


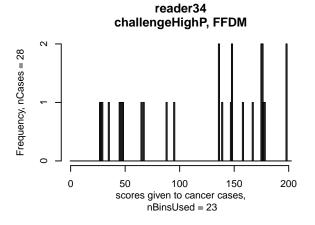


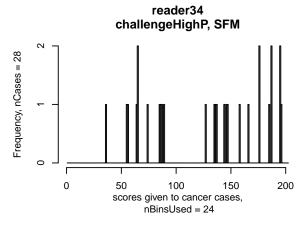


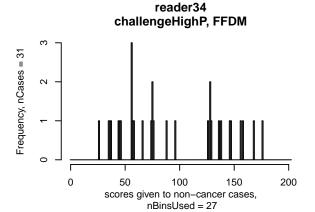


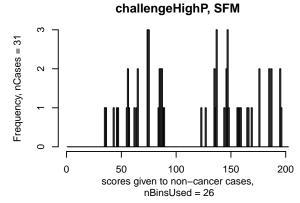
reader23, challengeHighP



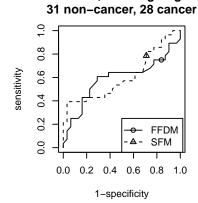




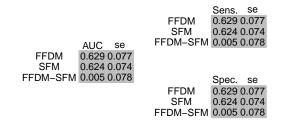


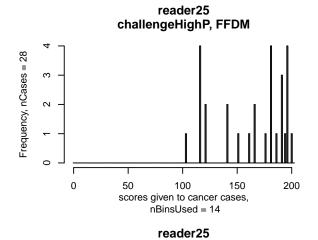


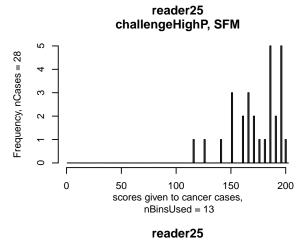
reader34

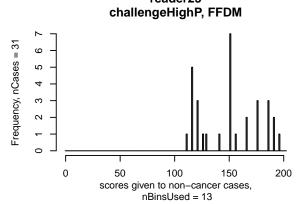


reader34, challengeHighP

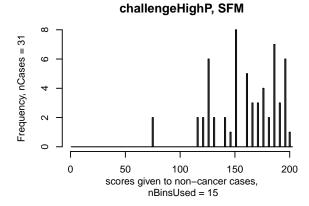


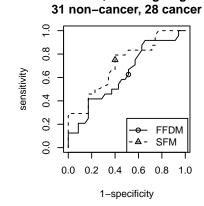


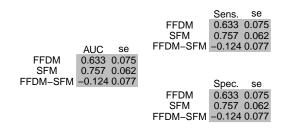


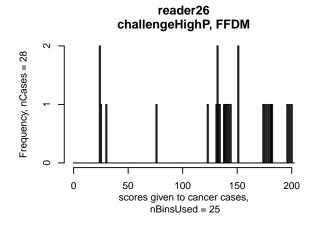


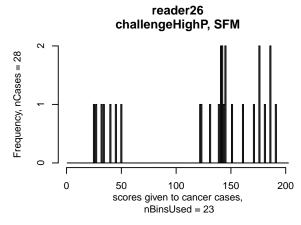
reader25, challengeHighP

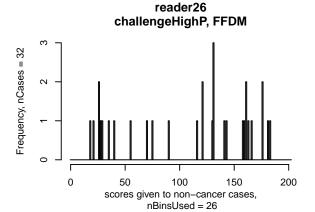


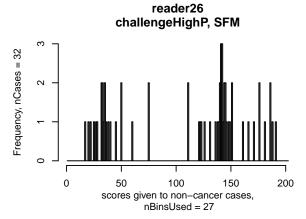


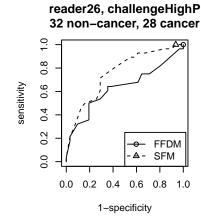


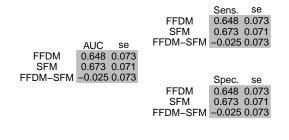


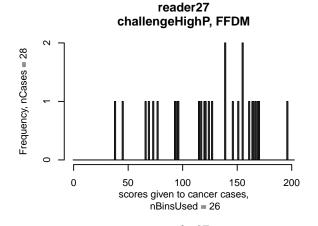


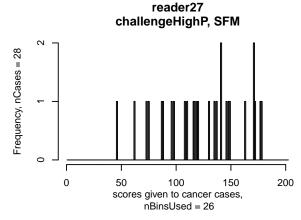


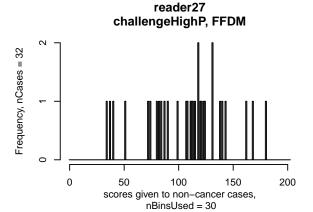


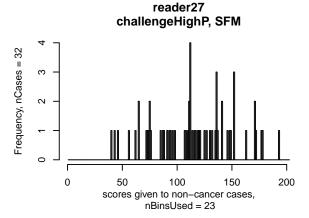


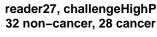


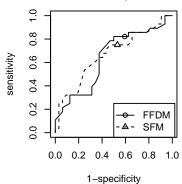


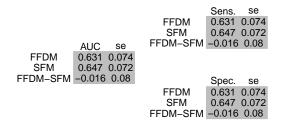


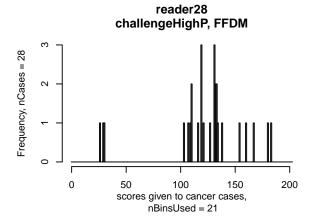


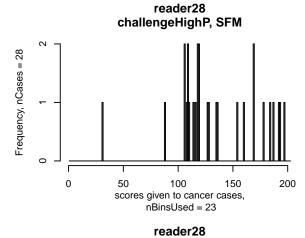


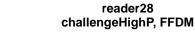


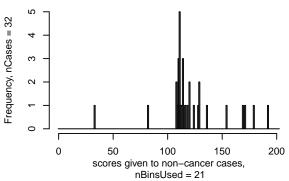


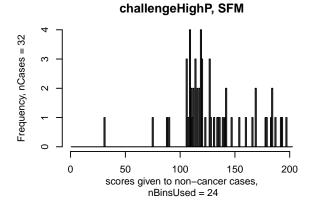




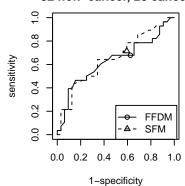


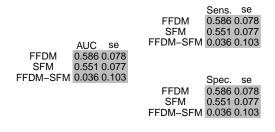


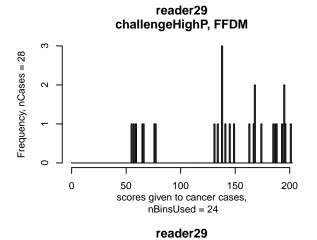


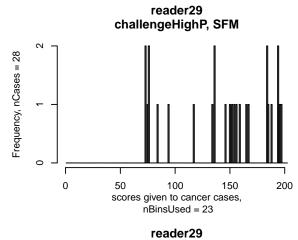


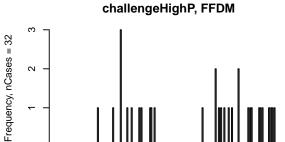
reader28, challengeHighP 32 non-cancer, 28 cancer

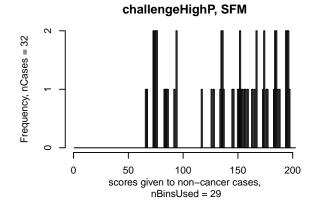


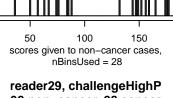




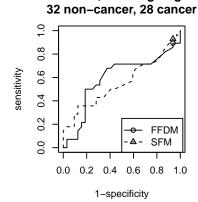


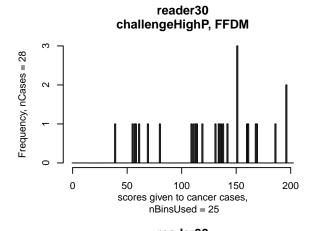


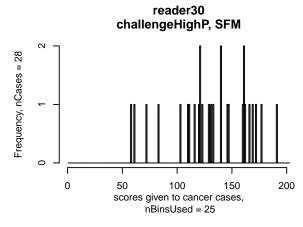


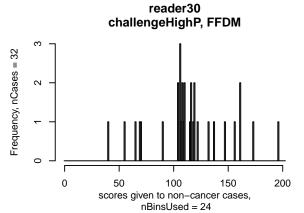


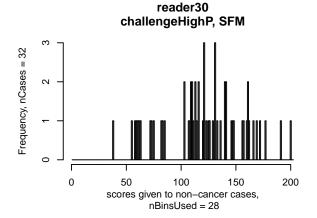


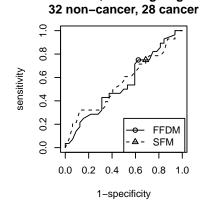




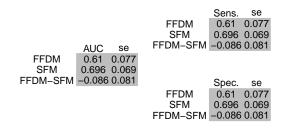


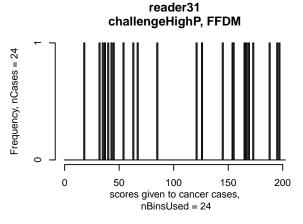


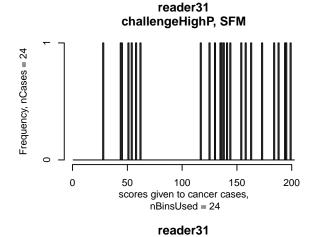




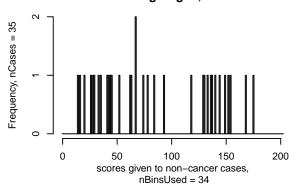
reader30, challengeHighP

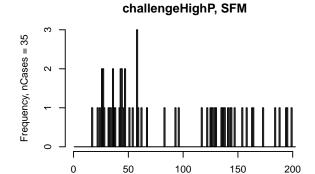






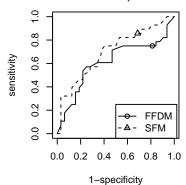


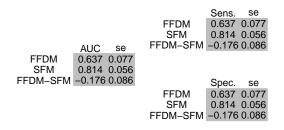


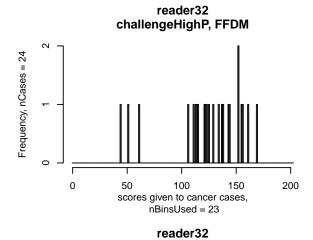


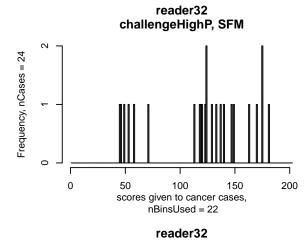
scores given to non-cancer cases, nBinsUsed = 29

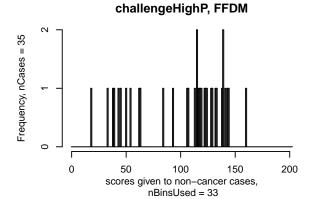
reader31, challengeHighP 35 non-cancer, 24 cancer

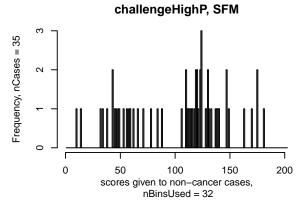




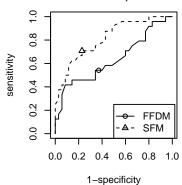


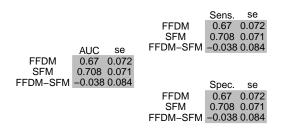


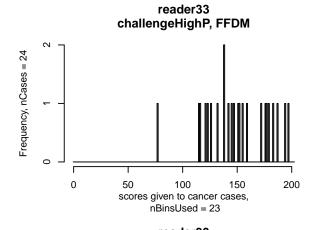


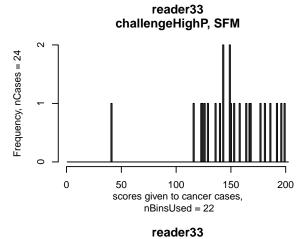


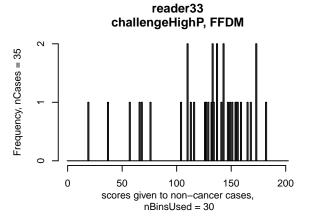




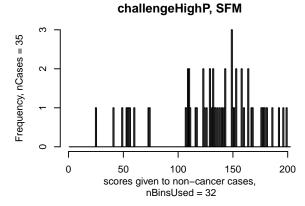


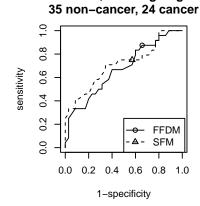


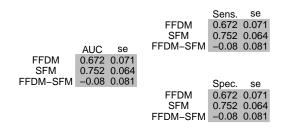


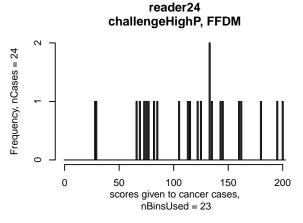


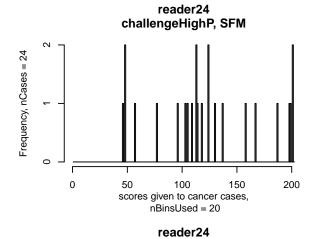
reader33, challengeHighP



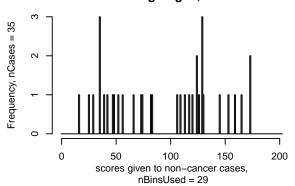


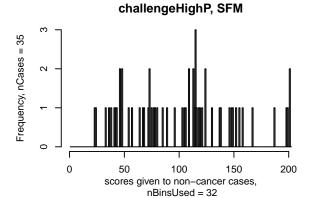




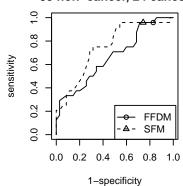


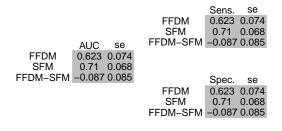


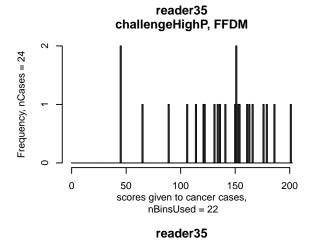


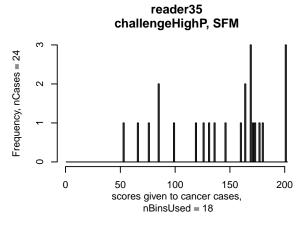


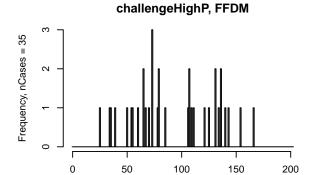
reader24, challengeHighP 35 non-cancer, 24 cancer









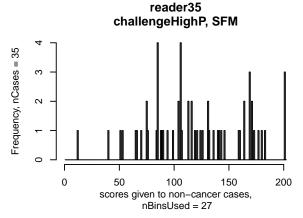


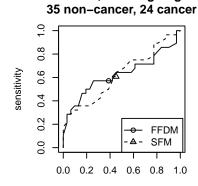
scores given to non-cancer cases,

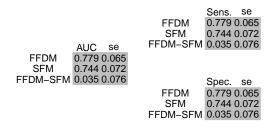
nBinsUsed = 28

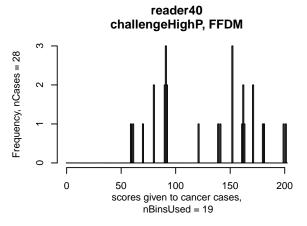
reader35, challengeHighP

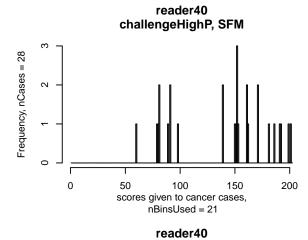
1-specificity



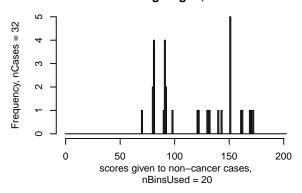


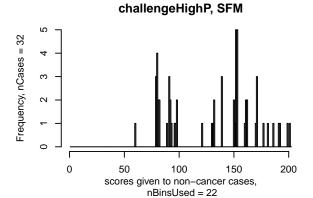




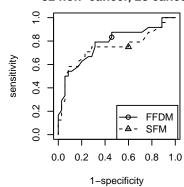


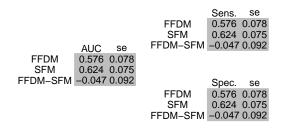


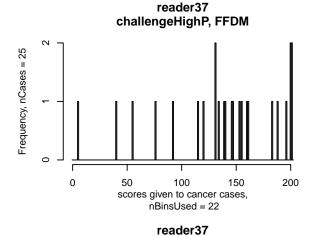


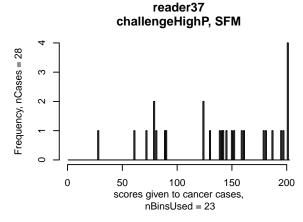


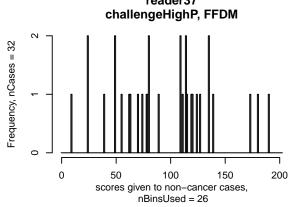
reader40, challengeHighP 32 non-cancer, 28 cancer



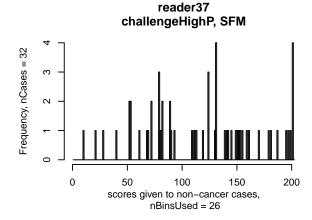


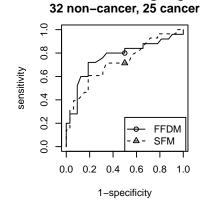


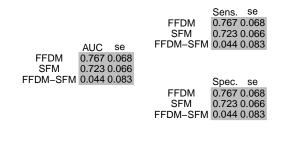


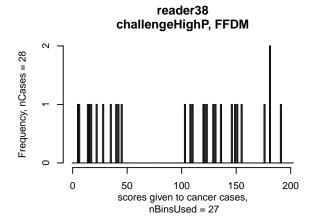


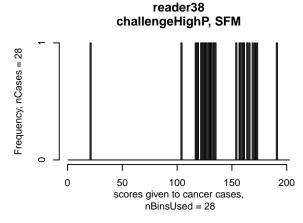
reader37, challengeHighP

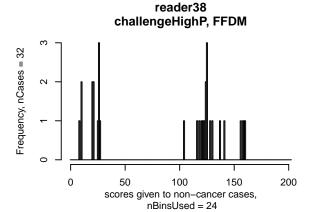


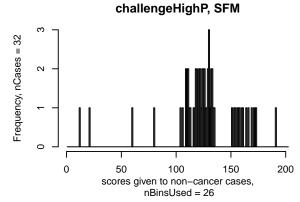




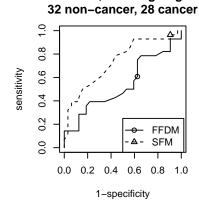




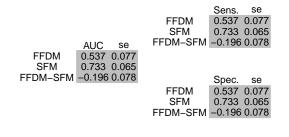


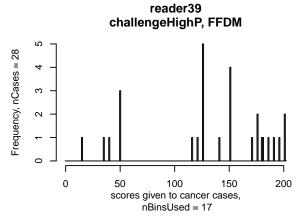


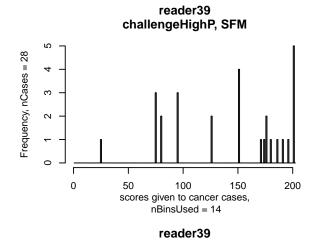
reader38



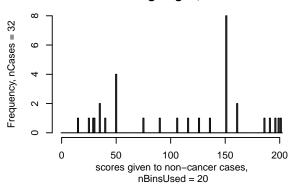
reader38, challengeHighP

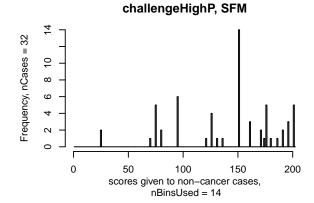




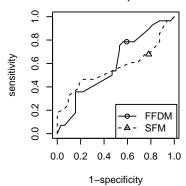




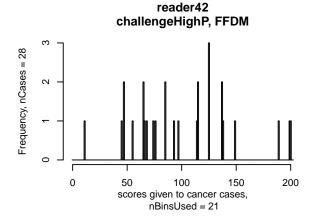


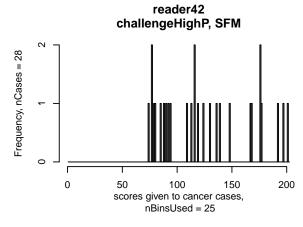


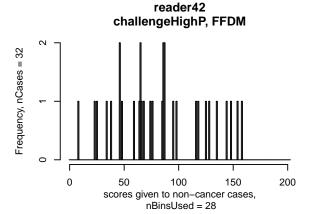
reader39, challengeHighP 32 non-cancer, 28 cancer

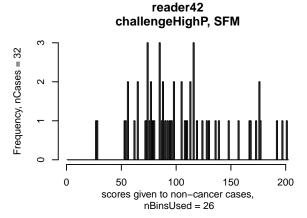


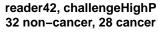


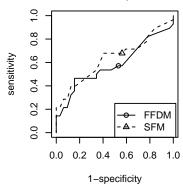


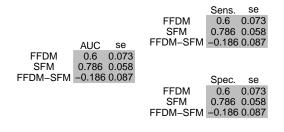












References

- 1. Gallas BD, Chen W, Cole E, et al. Supplementary materials: Impact of prevalence and case distribution in lab-based diagnostic imaging studies. 2018. Accessed December 30, 2018.
- 2. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005;353(17):1773–1783.
- 3. Hendrick RE, Cole EB, Pisano ED, et al. Accuracy of soft-copy digital mammography versus that of screen-film mammography according to digital manufacturer: ACRIN DMIST retrospective multireader study. Radiology. hwestern Univ, Galter Pavilion, 13th Floor, 251 E Huron St, Chicago, IL 60611, USA. edward.hendrick@gmail.com; 2008;247(1):38–48.
- 4. Gallas BD. iMRMC v4.0: Application for analyzing and sizing MRMC reader studies. Silver Spring, MD: Division of Imaging, Diagnostics,; Software Reliability, OSEL/CDRH/FDA; 2017. Accessed January 15, 2018.
- 5. Gallas BD. One-shot estimate of MRMC variance: AUC. Acad Radiol. 2006;13(3):353–362.
- 6. Gallas BD, Bandos A, Samuelson F, Wagner RF. A framework for random-effects ROC analysis: Biases with the bootstrap and other variance estimators. Commun Stat A-Theory. 2009;38(15):2586–2603.
- 7. Gallas BD, Brown DG. Reader studies for validation of CAD systems. Neural Networks Special Conference Issue. 2008;21(2):387–397.
- 8. Obuchowski N, Gallas BD, Hillis SL. Multi-reader ROC studies with split-plot designs: A comparison of statistical methods. Acad Radiol. 2012;19(12):1508–1517.