

1 **Supplementary Methods**

2 **Fisher's exact test** This is a statistical test that is used to determine nonrandom
3 associations between 2 categorical variables [1]. Fisher's exact test is similar to the
4 chi-squared test. If the sample size is large, the chi-squared test can be used successfully, but
5 significance values from the chi-squared test are only approximated. Fisher's exact test is a
6 statistical test that is used to analyze contingency tables when the sample size is small [1]. We
7 used Fisher's exact test in the present study. The odds ratio (OR) is defined as $a \times d / (b \times c)$,
8 where a is the number of NC or PD patients with a minor allele, b is the number of CR or PR
9 patients with a minor allele, c is the number of NC or PD patients with a major allele, and d is
10 the number of CR or PR patients with a major allele. The null hypothesis for Fisher's exact
11 test is $OR = 1$.

12 **The permutation test** The permutation test theory evolved from the works of
13 Fisher and Pitman in the 1930s [2]. In this study, p values of multiple-comparison analyses
14 were adjusted by applying the permutation test to two stages of screening. The case-control
15 (or phenotype) labels were randomly shuffled for the two screening stages, and p values were
16 calculated using Fisher's exact test. The lowest p value was selected for the randomized data.
17 This procedure was repeated 100,000 times. Exact p values for the permutation test were
18 calculated based on the distribution of the lowest p values.

19 **Multiple testing correction** Bonferroni correction is a method used to address the
20 problem of multiple comparisons (also known as the multiple testing problem). It is
21 considered the simplest and most conservative method to control the family-wise error rate
22 (FWER). Alternatively, false discovery rate (FDR) controlling procedures, such as the
23 Benjamini-Hochberg (BH) method [3], are more powerful (i.e., less conservative) than the
24 FWER procedures at the cost of increasing false positives within the rejected hypothesis. In
25 the present study, the BH method was used to calculate the q value. The q value is defined as

26 an FDR analog of the p value

27 **Akaike information criterion (AIC)** The AIC is a measure of the relative goodness
28 of fit of a statistical model [4]. A smaller AIC indicates a better fit when comparing fitted
29 objects. The AIC is defined according to the formula $-2 \times (\log \text{likelihood}) + (2 \times n_{par})$, where
30 n_{par} is the number of parameters in the fitted model, and the log likelihood value [5] was
31 obtained from the logistic regression model.

32 **The receiver operating characteristic (ROC)** ROC is a graphical plot which
33 illustrates the performance of a binary classifier system as its discrimination threshold is
34 varied. It is built by plotting sensitivity (the number of true positive results divided by the
35 number of true positive samples) against (1 minus specificity) at various threshold settings.
36 Specificity is the number of true negative results divided by the number of true negative
37 samples. The area under the curve (AUC) of a ROC curve is an indicator representing
38 expected performance. A higher AUC is more desirable, with a value of 1.00 denoting perfect
39 performance (sensitivity and specificity are both 100%), while a value of 0.50 indicates
40 random performance.

41

42

43

44

45

46

47

48

49

50

51 **References**

- 52 1. Fisher RA: **On the interpretation of χ^2 from contingency tables, and the**
53 **calculation of P.** *J Roy Statistical Society* 1922, **85**(1):87-94.
- 54 2. Pitman EJG: **Significance tests which may be applied to samples from any**
55 **population. Part III. The analysis of variance test.** *Biometrika* 1938, **29**:322-335.
- 56 3. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and**
57 **powerful approach to multiple testing.** *J R Statist Soc serB* 1995, **57**(1):298-300.
- 58 4. Akaike H: **A new look at the statistical model identification.** *IEEE T Automat Contr*
59 1974, **19**(6):716-723.
- 60 5. Sakamoto Y, Ishiguro M, Kitagawa G: **Akaike Information Criterion Statistics.**
61 Dordrecht: Reidel Publishing Company; 1986.

62

63