

Supporting Information

Karczewski et al. 10.1073/pnas.1219099110

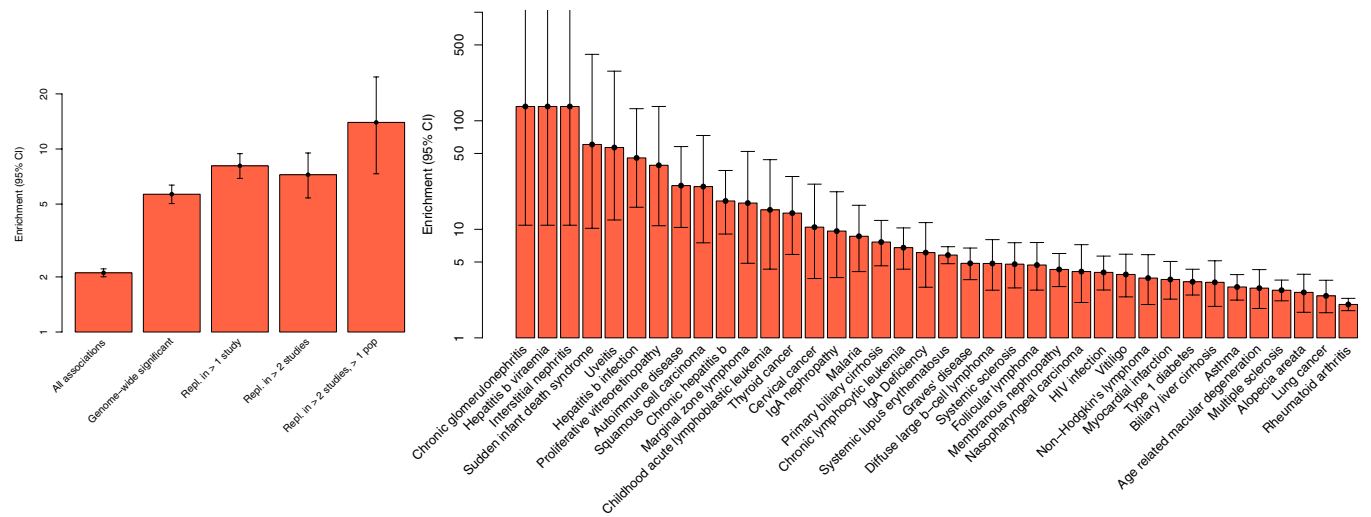


Fig. S1. Regulatory RNA polymerase II (PolII) regions are enriched for disease-associated variants. Generally, regulatory regions are also enriched for disease-associated variants, but the diseases enriched are a more general set than those for NFκB (Fig. 2C); the PolII regions are enriched for some inflammatory diseases as well, but also many others.

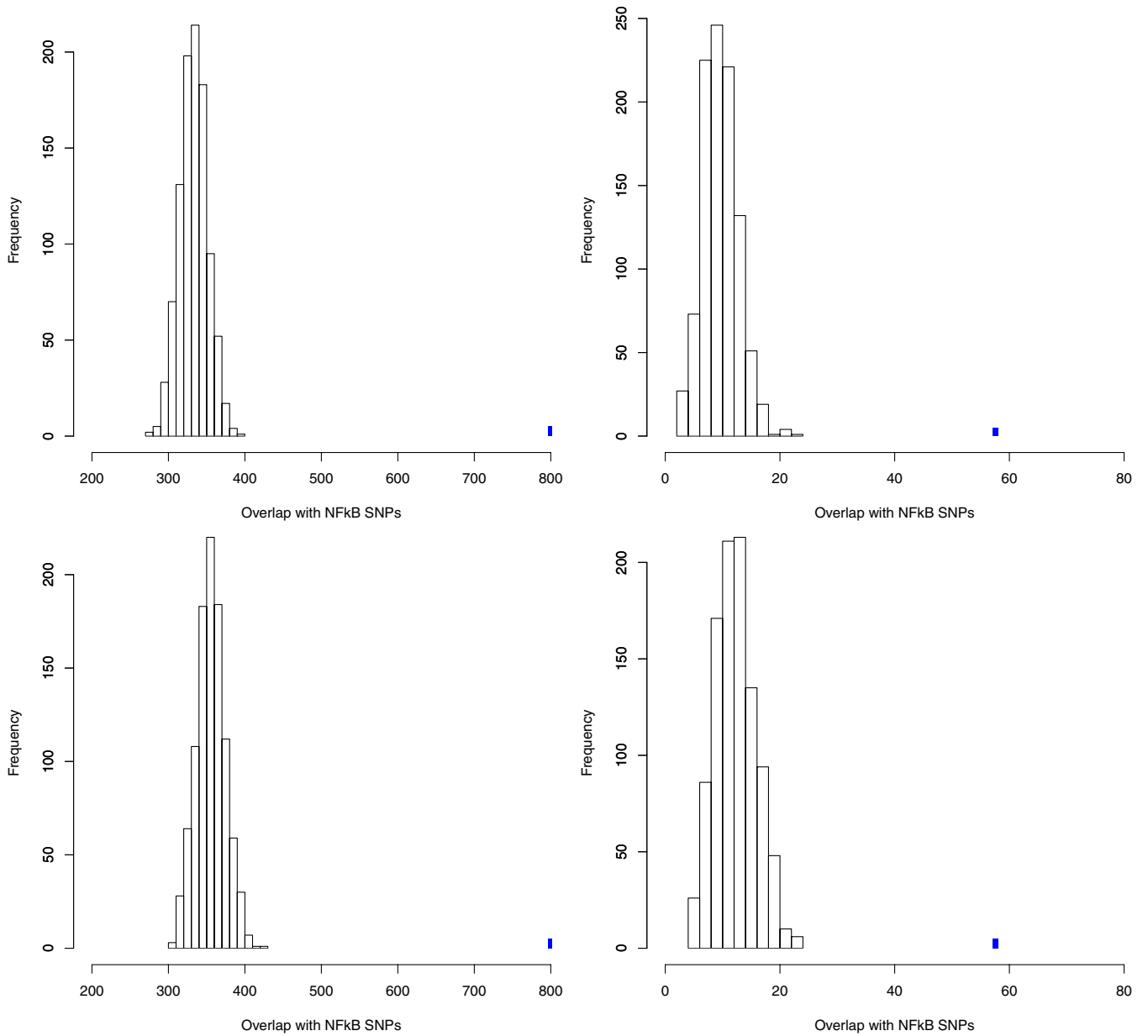


Fig. S2. NFkB regions are enriched for disease-associated variants. To assess any bias due to minor allele frequency (MAF) or distance to transcription start site (TSS) of the enrichment of disease-associated variants in NFkB regions, a background distribution of variants, matched to MAF and distance to TSS, was generated. (Left) All associations. (Right) Stringent associations ($P < 10^{-7}$). (Upper) Background distribution matched to disease-associated variants. (Lower) Background distribution matched to all NFkB variants (Methods).

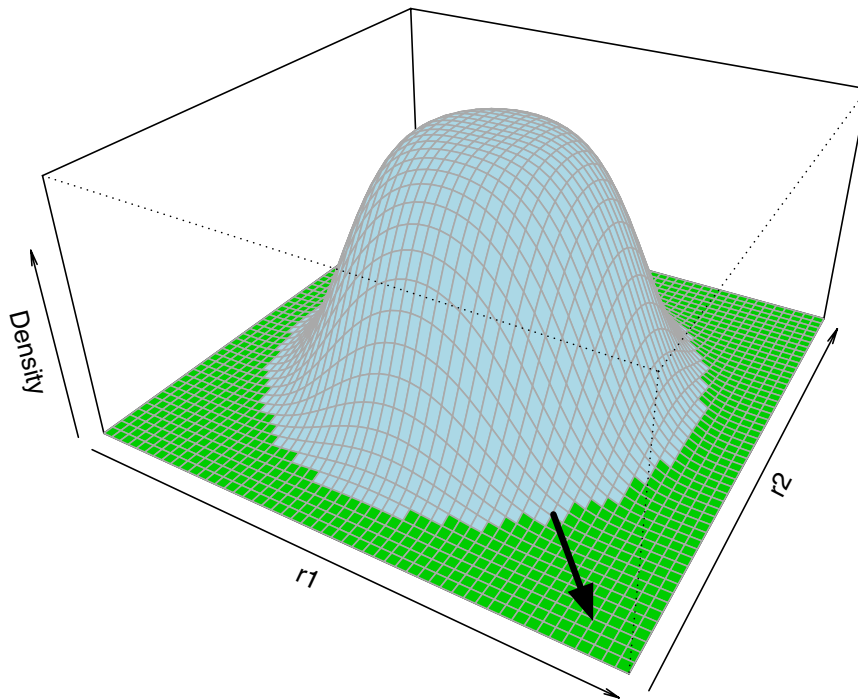


Fig. S7. Permutation testing across two statistical tests. To generate the null distribution for both the variant-binding association and the binding-expression correlation, we permuted the variant, binding, and expression across individuals. We ran association tests for the permuted sets as before and this schematic demonstrates an ideal distribution based on normally distributed r values. As with a typical permutation test, the empirical P value is estimated by the number of random sets more extreme than the observed value. In this case, we considered the extremity on two dimensions, using the sum of squared correlation values ($r_{\text{binding}}^2 + r_{\text{expression}}^2$) as a distance metric. We compared the distance metrics for the true value to all permuted sets more extreme than this value (displayed in green) to estimate an empirical P value.

Dataset S1. All diseases

[Dataset S1 \(XLSX\)](#)

Enrichments are shown for all associations, as well as per-disease enrichments (data underlying Fig. 2 A and C and Fig. S1).

Dataset S2. Variants in binding sites associated with NFκB binding and linked to disease-associated SNPs

[Dataset S2 \(XLSX\)](#)

Significant associations (317; $P < 0.05$) between the variant and NFκB binding are shown, as well as the strongest correlation between the binding site strength and gene within 100 kb (64 of which are significant; $P < 0.05$). Additionally, there are seven heterozygous allele-specific binding events ($P < 0.01$).

Dataset S3. One hundred forty-three diseases used for linking disease-associated SNPs, binding, and expression of differentially expressed genes in the same disease (Fig. 5)

[Dataset S3 \(XLSX\)](#)