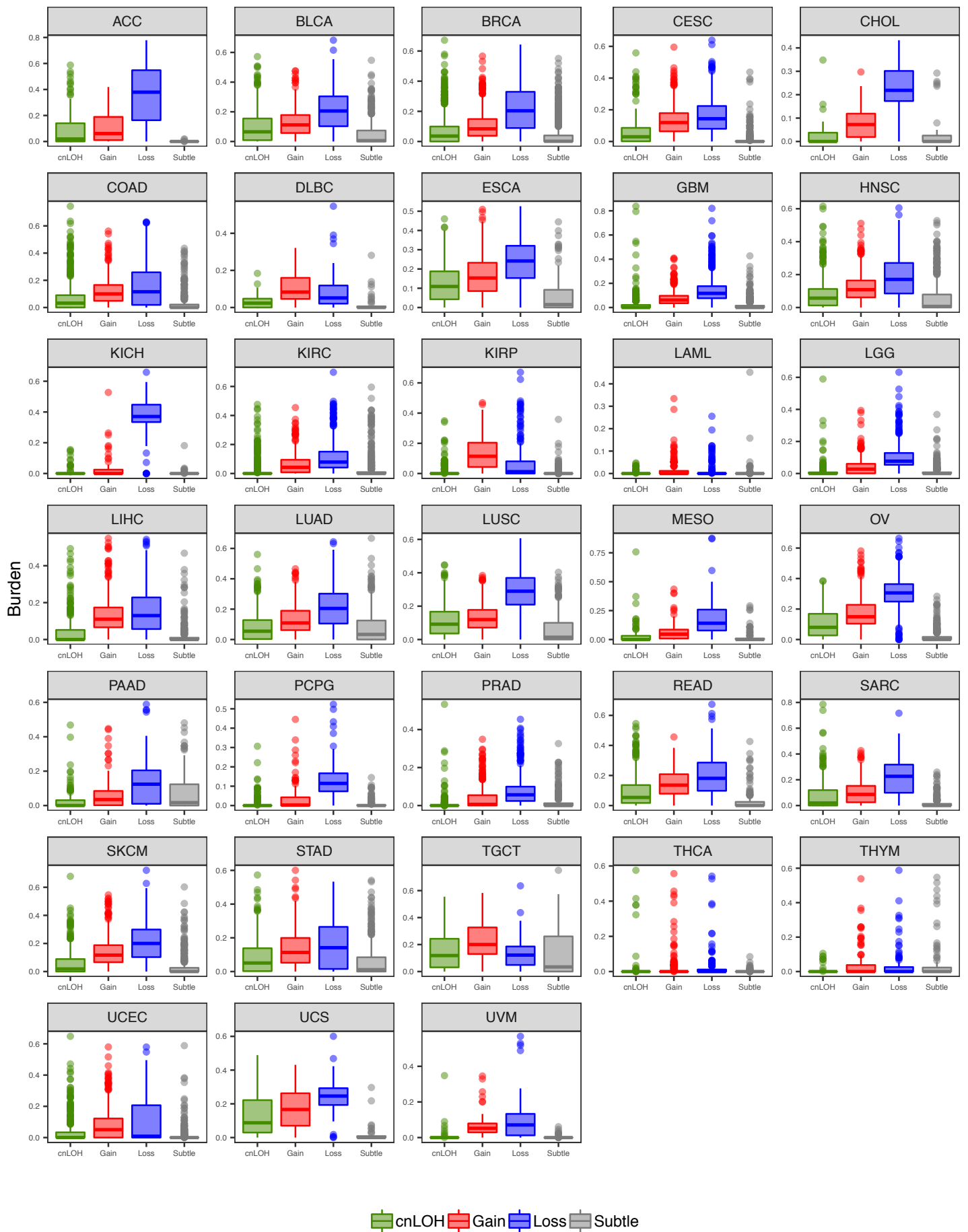


Pan cancer patterns of allelic imbalance from chromosomal alterations in 33 tumor types

Smruthy Sivakumar*, F Anthony San Lucas*, Yasminka A Jakubek, Zuhail Ozcan, Jerry Fowler, Paul Scheet

*Equally contributing authors

SUPPLEMENTARY FIGURES



■ cnLOH
 ■ Gain
 ■ Loss
 ■ Subtle

Figure S1. Distribution of allelic imbalance burdens across 33 tumor sites in TCGA. For each tumor site, the proportion of the genome under allelic imbalance leading to each event type are shown. Boxplots represent the distribution of burdens across all samples profiled for each tumor site, with gains shown in red, losses shown in blue, cnLOH shown in green and subtle, unclassifiable events shown in grey. Tumor sites exhibiting varying patterns of enrichment across these different event types were identified.

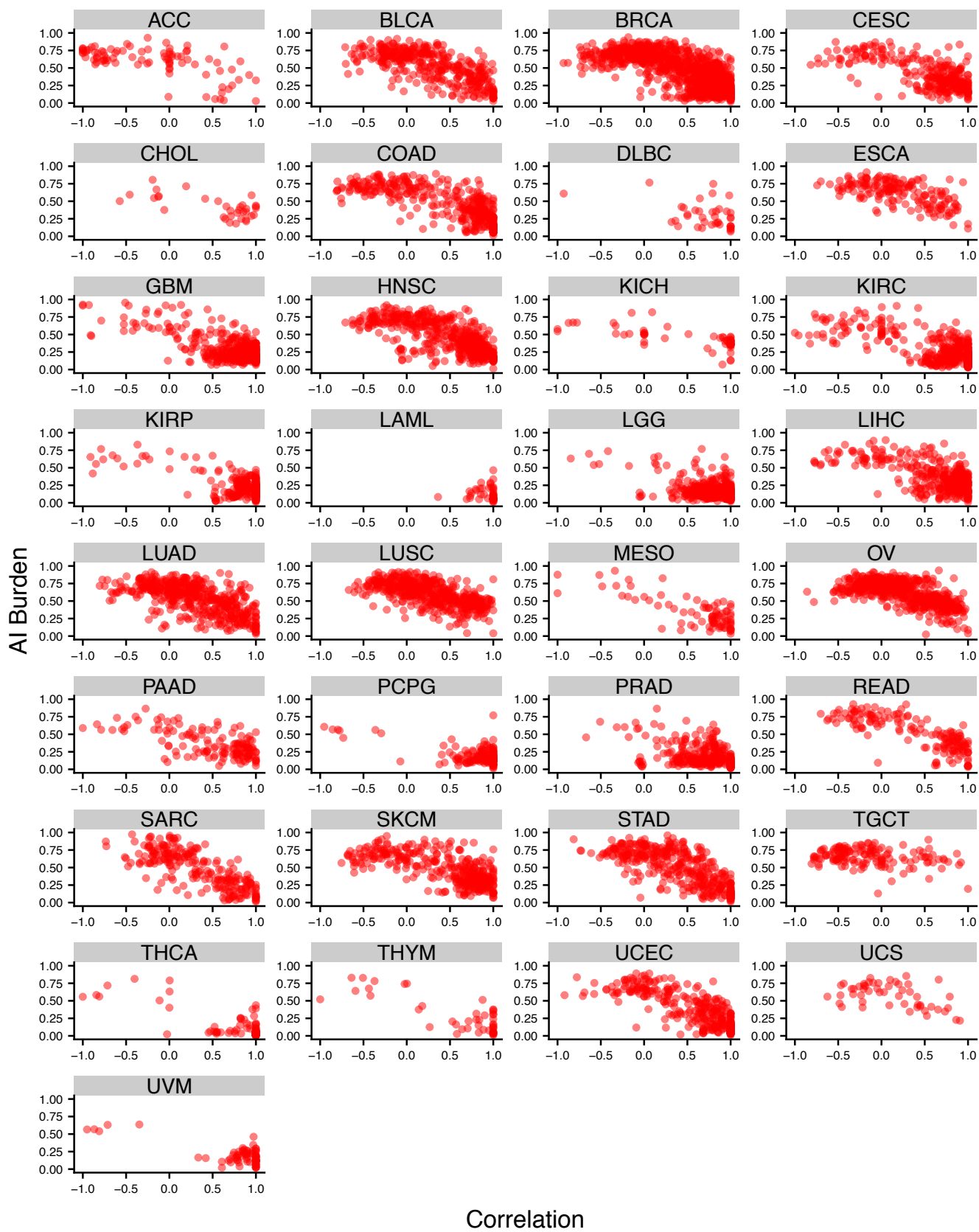


Figure S2. Relationship between overall allelic imbalance genomic burden and concordance between call sets. Shown here are scatter plots for each tumor site, with the x-axis corresponding to the extent of concordance between SCNA calls inferred from hapLOH and those reported in TCGA and the y-axis corresponding to the allelic imbalance genomic burden inferred based on hapLOH. Across tumor sites, a trend of higher negative correlations, thereby discordance, was observed in cases with higher overall burden of allelic imbalance. This suggests that tumor genomes that exhibited higher chromosomal instability/aneuploidy showed patterns of discordance between call sets, possibly due to the difficulty in estimation of copy-neutral or normal regions for the accurate calibration of SCNAs.

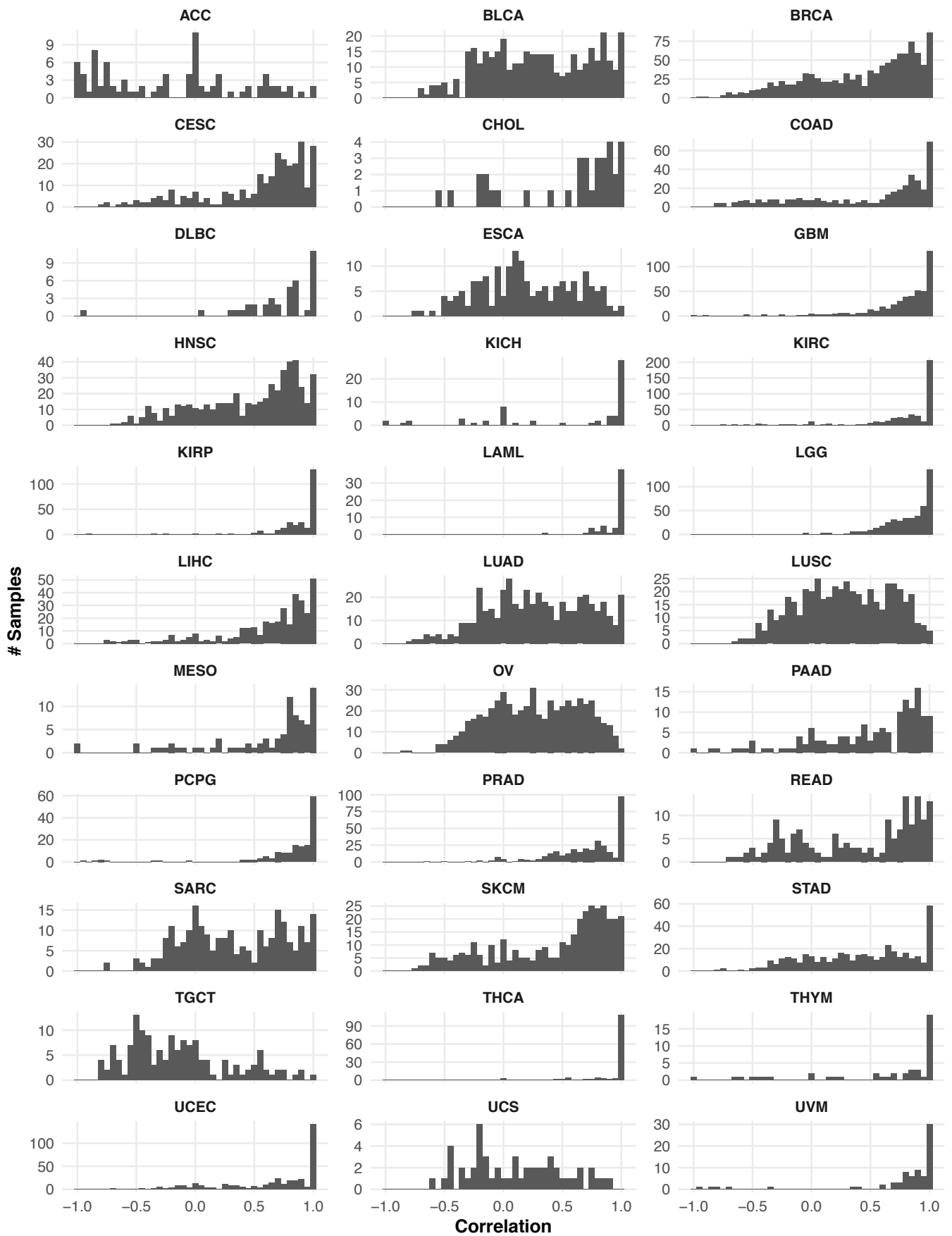


Figure S3. Distribution of concordance statistics between SCNA call sets. In each tumor type assessed, the extent of concordance between SCNA calls inferred from hapLOH and those reported in TCGA, as determined by the correlation statistic, were grouped into bins (bin size = 0.05). Shown here are histograms for each tumor type, with the x-axis corresponding to the different bins of correlation values and the y-axis depicting the number of samples in each bin. The overall distribution varied by tumor type. While some tumors displayed a tight peak closer to a correlation of 1 (high concordance), a subset of sites exhibited bimodal or sufficiently high dispersion that spanned the range (-1,1).