

COMPUTATIONAL OPTIMISATION OF TARGETED DNA SEQUENCING FOR CANCER DETECTION

Pierre Martinez, Nicholas McGranahan, Nicolai Juul Birkbak, Marco Gerlinger, Charles Swanton*

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1:

Illustration of the hotspot region definition process. For different distance thresholds d , mutations were iteratively grouped into clusters that determined hotspot regions. In the example, the blue and yellow mutations are 13 base pairs apart and there is no reported mutation between them. Using at threshold $d=10$, these mutations are thus in two different hot spot regions. Increasing d to 20, all 4 mutations will be clustered together in a single hotspot region. Even if the green and red mutations are 21 base pairs apart, all 4 mutations are within 20 base pairs of at least one other mutation in the hotspot region. Any mutation less than 20 base pairs upstream of the red mutation or downstream of the green mutation would be included in the hotspot region.

Supplementary Figure 2:

Computationally estimated sensitivity of targeted sequencing screen for the detection of specific tumor types using specific candidate genes. Left panels: percentage of tumors for each tumor type in each cohort that have at least one mutation in a gene included in panels defined by combining the best 1 to 25 specific candidate genes. Right panels: percentage of tumors for each tumor type that have at least one mutation in a gene included in the best panel of specific genes defined for a maximum length varying from 100 to 60,000 nucleotides.

Supplementary Figure 3:

Proportion of in-frame mutations that are detectable and not detectable by hotspot region targeting approaches in each cohort. Each panel represents a different distance threshold d used to define the hotspot regions. P-values are given by two-tailed Fisher's exact tests.

Supplementary Figure 4:

Computationally estimated sensitivity of targeted SNV sequencing screen for the detection of specific tumour types using pan-cancer candidate SNVs. Percentage of tumours for each tumour type in each cohort that have at least one mutation in panels defined by iteratively combining the best 1,000 pan-cancer single nucleotide variants.







