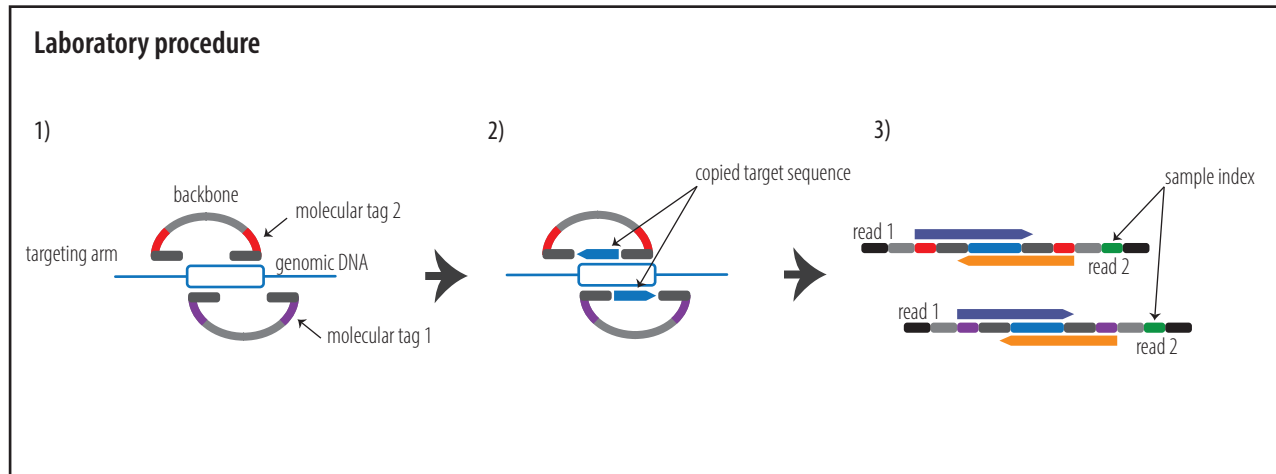
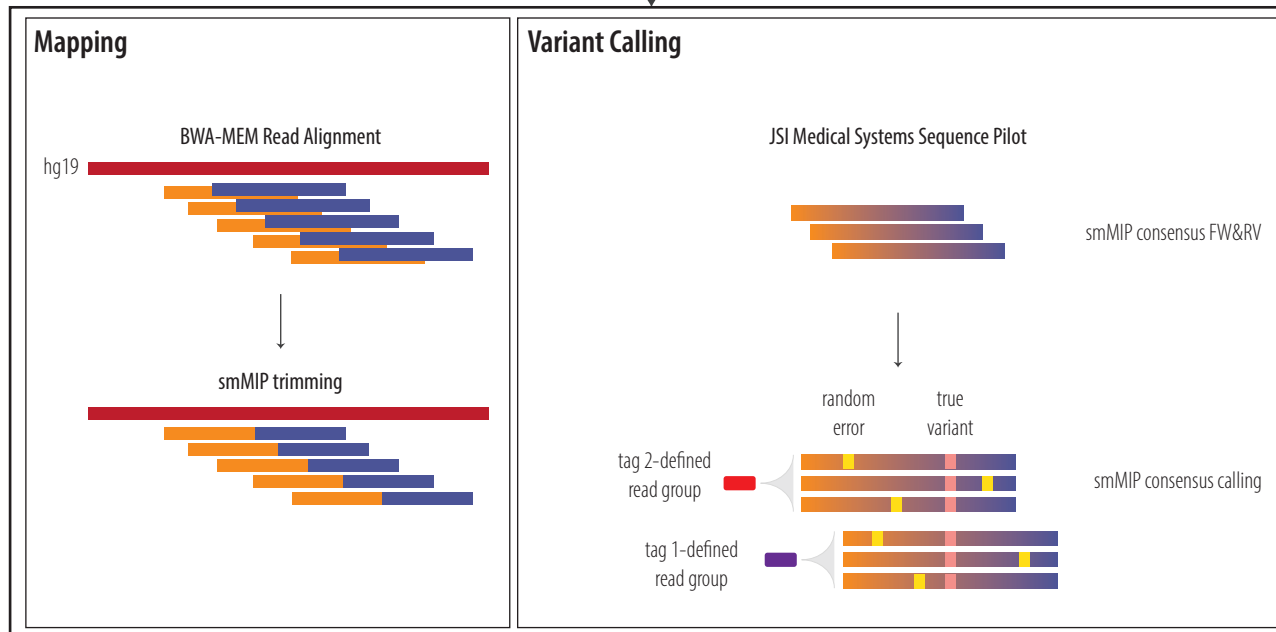
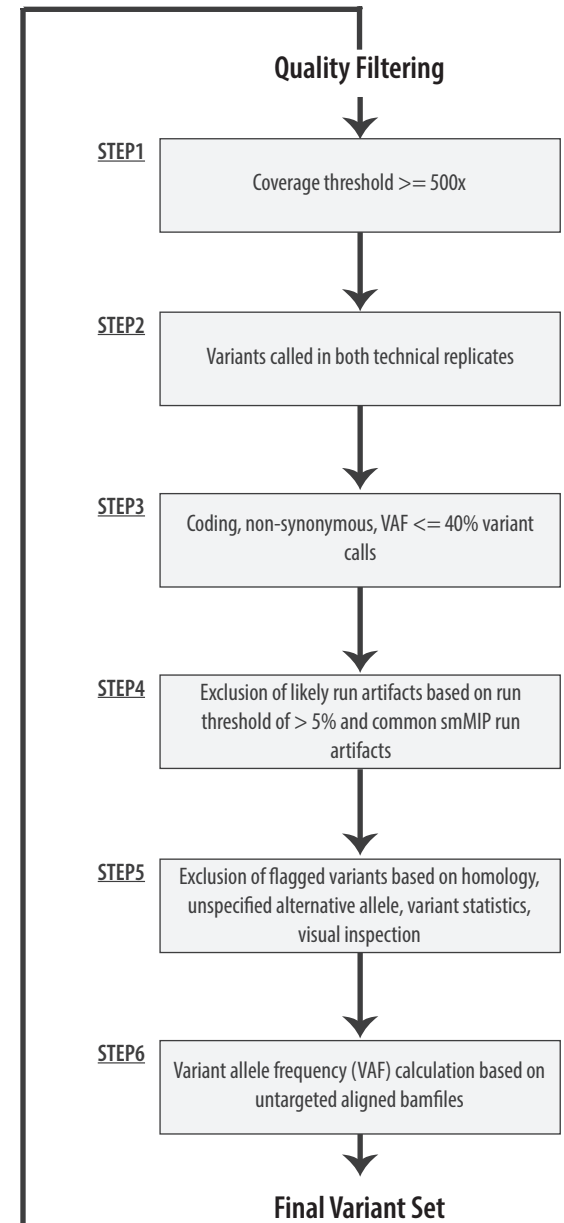


a**b****c**

File S2. smMIP workflow to further boost sensitivity for low-level somatic mutation detection. (a - adapted from Hiatt et. al. 2013) smMIP insert-sizes were shortened to 54nt to enable full forward and reverse read coverage (double sequencing of each insert), target-sequences were generally targeted by at least two independent smMIP-probes (double tiling), and each smMIP capture underwent two independently barcoded PCR-reactions (double PCR replicates). Sequencing was performed using the Illumina NextSeq 500 system. (b - adapted from Hiatt et. al. 2013) Raw sequencing data was converted to FastQ-files after which two independent data processing strategies were applied. Mapping: FastQ-file reads were aligned to the reference genome (Hg19) with BWA-MEM after which the overlap between Forward (FW) and Reverse (RV) reads was trimmed off (smMIP trimming). Variant Calling: FastQ-files were imported in JSI Medical Systems Sequence Pilot in which first a consensus between FW and RV reads is determined (smMIP consensus FW&RV), after which tag-defined read groups enable smMIP consensus calling. (c) The aligned reads and resulting variant calls are then subjected to a stringent quality filtering pipeline. First, individuals with an average sequencing depth <500x are excluded (STEP1). Second, only variants called in both technical replicates are kept (STEP2). Third, coding, non-synonymous variant calls with a Variant Allele Frequency (VAF) $\leq 40\%$ are kept (STEP3). Fourth, likely run-specific artifacts (variant calls in > 5% of samples per run) are excluded (STEP4). Fifth, exclusion of variants based on homology, unspecified alternative allele, variant statistics, and visual inspection (STEP5). And finally, for each variant position we generated mpileups based on the aligned reads to determine the final VAF for each CHDM (STEP6).