

Genome analysis

Nanopanel2 calls phased low-frequency variants in Nanopore panel sequencing data

Niko Popitsch ^{1,2,*}, Sandra Preuner^{1,*} and Thomas Lion^{1,3}

¹Children's Cancer Research Institute, Vienna 1090, Austria, ²Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna 1030, Austria and ³Department of Pediatrics, Medical University of Vienna, Vienna 1090, Austria

*To whom correspondence should be addressed.

Associate Editor: Christina Kendzierski

Received on January 21, 2021; revised on June 15, 2021; editorial decision on July 6, 2021; accepted on July 15, 2021

Abstract

Motivation: Clinical decision making is increasingly guided by accurate and recurrent determination of presence and frequency of (somatic) variants and their haplotype through panel sequencing of disease-relevant genomic regions. Haplotype calling (phasing), however, is difficult and error prone unless variants are located on the same read which limits the ability of short-read sequencing to detect, e.g. co-occurrence of drug-resistance variants. Long-read panel sequencing enables direct phasing of amplicon variants besides having multiple other benefits, however, high error rates of current technologies prevented their applicability in the past.

Results: We have developed Nanopanel2, a variant caller for Nanopore panel sequencing data. Nanopanel2 works directly on base-called FAST5 files and uses allele probability distributions and several other filters to robustly separate true from false positive (FP) calls. It effectively calls SNVs and INDELS with variant allele frequencies as low as 1% and 5%, respectively, and produces only few low-frequency false-positive calls (~1 FP call with VAF<5% per kb amplicon). Haplotype compositions are then determined by direct phasing. Nanopanel2 is the first somatic variant caller for Nanopore data, enabling accurate, fast (turnaround <48 h) and cheap (sequencing costs ~10\$/sample) diagnostic workflows.

Availability and implementation: The data for this study have been deposited at zenodo.org under DOIs accession numbers 4110691 and 4110698. Nanopanel2 is open source and available at <https://github.com/popitsch/nanopanel2>.

Contact: niko.popitsch@imba.oeaw.ac.at

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Diagnosis and treatment of cancer patients is strongly guided by knowledge about the composition of disease-relevant (sub-)clones in patient samples. This requires technologies that can accurately determine the relative frequencies of haplotypes that inform about co-occurrence of variants on the same chromosome (compound mutations) which often play a central role in drug resistance (Cao *et al.*, 2015). For example, resistance to tyrosine kinase inhibitors including the third-generation drug ponatinib is caused by multiple mutations on the same allele in the tyrosine kinase domain (TKD) of BCR-ABL1 fusions in chronic myeloid leukemia (CML) patients (Byrgazov *et al.*, 2018; Khorashad *et al.*, 2013; Zabriskie *et al.*, 2014).

High-throughput panel sequencing has become a widespread tool for the rapid and cost-effective detection of such clinical biomarkers, among other reasons, due to its superior detection limit [LOD ~1% variant allele frequency (VAF)] when compared to traditional methods (e.g. LOD of Sanger sequencing ~20% VAF).

Nanopore long-read panel sequencing, in particular, offers multiple benefits when compared to current short-read technologies: (i) direct phasing of variants within an amplicon, (ii) lower upfront and running costs, (iii) shorter interaction and turnaround times, (iv) accessibility of highly repetitive genomic regions and (v) direct detection of structural variations. The relatively high error rates of Nanopore reads, however, made it difficult to call low-frequency variants in such datasets without producing large numbers of false-positive calls (Ameur *et al.*, 2019; Kono and Arakawa, 2019; Krishnakumar *et al.*, 2018; Orsini *et al.*, 2018; Rang *et al.*, 2018).

1.1 Nanopanel2

We have developed Nanopanel2 (np2), a somatic variant calling software that effectively filters such false-positive calls. Np2 calls small somatic variants (SNVs and INDELS) from Nanopore panel sequencing data and iterates all identified haplotypes per sequenced amplicon. Briefly, np2 aligns base-calling probabilities from guppy to the specified reference (amplicon) sequence and then iterates these

alignments column-wise (pileup) to call variants. Alt-allele reads with skewed basecalling probabilities are filtered from the data. Np2 then calculates a number of statistics from aligned bases/INDELs per pileup and applies nine different algorithms to effectively filter false-positive calls (Supplementary Fig. S1, Supplementary Table S5). Our software then outputs VCF and TSV files for downstream processing that contain all evaluated variant calls (including filtered ones) and associated core statistics. Finally, np2 calculates direct haplotypes for all non-filtered (PASS) variants and produces count tables for downstream analyses as well as haplotype map PDFs for visual interpretation.

2 Materials and methods

2.1 Oncospan data

We obtained Nanopore sequencing data for nine amplicons from the cell-line-derived Horizon OncoSpan gDNA reference standard (<https://horizondiscovery.com/>, cf. Supplementary Table S3) from Oxford Nanopore Technologies Ltd. The oncospan DNA samples were sequenced on a single MinION flow-cell resulting in overall 13.5 Mio reads. From these data we created six samples with 20k reads each and one sample with 100k reads by random subsampling (Supplementary Table S2). In addition, we obtained deep Illumina WES data (500× coverage; from provider) for these amplicons, called (low-frequency) variants using the pisces (Dunn *et al.*, 2019) variant caller (v5.2.9.122, command lines are provided in Supplement) and manually inspected variant calls in a genome browser. We then considered all regions with a minimum coverage of 10× in Nanopore and Illumina data and measured np2's performance considering unfiltered pisces calls as ground truth. Note that a subset of these variants was validated by droplet digital PCR (ddPCR) in the original oncospan data.

2.2 BCR-ABL1 data

We further evaluated np2 using manually created in-vitro benchmark datasets as well as several clinical datasets in a BCR-ABL1 background with a single amplicon covering the TKD of the BCR-ABL1 fusion gene. First, we created a dilution series ('abl1_min_1', Supplementary Table S7) using ThermoFisher Topo T4 cloned plasmids containing major BCR-ABL1 constructs (wildtype, T315I single mutation, E255V single mutation and T315I+E255V compound mutations). We used either one plasmid containing both mutations (compound) and two plasmids containing one mutation each which allowed us to benchmark np2's ability to call haplotypes. Plasmids were calibrated to a BCR-ABL1 copy number of 1E+4 by RT-PCR (Gabert *et al.*, 2003). T315I single- and T315I+E255V compound plasmids were then mixed into the BCR-ABL1 WT background (VAF 80%, 50%, 10%, 5%, 3%, 1%).

We also created two replicates of a dilution series (Supplementary Table S8, 'abl1_min_2'/dilution) containing two compound variants at different concentrations (10%, 3%, 1%) to learn about np2's detection limit and about how well it could reproduce expected VAFs. In addition, we evaluated np2 with different combinations of clinically relevant variants from an international ring trial comprising cDNA samples derived from BCR-ABL positive cell lines with a BCR-ABL transcript copy number of about 10% ('abl1_min_2'/ringtrial).

To learn about the influence of read depth on np2's performance, we sequenced a MinION flow-cell with 9 clinical samples ('abl1_min_3', Supplementary Table S9) and randomly downsampled the alignments to 1k, 5k, 10k and 25k reads per sample. In addition, we sequenced one benchmark sample (plasmid mix, 'abl1_flo_1') and eight multiplexed clinical samples ('abl1_flo_2') on two 'Flongles' (Supplementary Table S10) to evaluate np2's performance on these smaller flow-cells.

The truth status of variants in clinical samples was determined/validated using various orthogonal methods including Sanger sequencing (Hochhaus *et al.*, 2020), ligation-dependent PCR (Preuner *et al.*, 2008), MiSeq panel sequencing (Machova Polakova *et al.*, 2015) and Pyrosequencing (Alikian *et al.*, 2012). Notably, we

do not have a complete ground truth for clinical samples due to the different detection limits of the used validation techniques (e.g. 20% for Sanger sequencing). Consequently, false-positive rates for the respective datasets might be inflated as we do not know whether low-frequency variants called by np2 are in fact true variants.

2.3 Nanopanel2 preprocessing

Np2 takes guppy (v3.6.1) base-called FAST5 files and a JSON configuration file (we used np2's default configuration for all evaluation runs) as input and optionally starts by demultiplexing reads using porechop (v0.2.4, Wick *et al.*, 2017) if configured. It then extracts FASTQ files and guppy probability values from the FAST5 files. In this step, np2 can optionally split reads that exceed a configurable length as we observed considerable numbers of chimeric reads in the oncospan data (possibly artefacts from the PCR reaction). Reads are then aligned to the reference sequence using the configured long-read mappers and BAM files are annotated with extracted basecalling probability values for efficient downstream processing. Optionally, np2 can randomly downsample the final alignments, e.g. for faster processing or debugging purposes.

2.4 Read mapper selection

Long read alignment is non-trivial and error prone due to the high error rates in the used technology [a current review estimates ~87–98% accuracy (Logsdon *et al.*, 2020)]. We therefore evaluated the influence of different alignment algorithms on the overall performance of np2. Our software supports three different long-read alignment algorithms: minimap2 (Li, 2018), ngmlr (Sedlazeck *et al.*, 2018) and last (Kielbasa *et al.*, 2011) and can be configured to align reads with any combination of these programs. If multiple aligners were used, np2 can calculate a (majority vote) consensus call set from all produced alignments. We measured per-mapper performance using data created from the Horizon OncoSpan reference standard (oncospan) and from a set of BCR-ABL1 TKD (abl1) panel sequencing datasets (Supplementary Table S2, Supplementary Fig. S5, Section 2). Overall, minimap2 showed the best performance with respect to false-discovery rates (FDR), and false-negative rates (FNR) while ngmlr and last resulted in higher numbers of false-negative (FN) and false-positive (FP) calls, respectively (Fig. 1a, Supplementary Fig. S6, S23). Minimap2 datasets also showed the highest correlation between expected and observed VAF (Supplementary Fig. S7). We therefore conducted our main analysis using minimap2 alignments unless stated otherwise.

2.5 Variant calling

Np2 iterates alignments column-wise and, for each position, considers all alternate alleles (alt) with a configurable minimum number of supporting reads per pileup as potential sequence variants. It then interrogates guppy probability distributions for each alt base-call in the pileup and filters reads that show increased probabilities for any different allele (including the reference allele) among those. For this, np2 first iterates the maximum 'flip-flop' probability for each possible non-alt allele and then tests for equal distribution of these values using a χ^2 test, rejecting base-calls with skewed distributions ($P \leq 0.05$) that we found to be error prone. In initial experiments, we observed that false positive SNV calls were often correlated with high overall ref base probabilities in the pileup, demonstrating that the basecaller was unsure about its decision to call the alt allele (Supplementary Fig. S2). We then use this algorithm to calculate corrected alt-allele counts and frequencies and filter potential SNV calls based on configurable thresholds for these values (filter 'AQ1'). Np2 furthermore filters calls with large fractions of rejected alt-base calls (AQ2). Supplementary Figure S21a, b show, for example, the effect of this correction algorithm around homopolymer runs.

Strand bias, meaning that the alt allele is disproportionately found on reads of a particular strand, is a strong indicator of artefacts in NGS data. We have implemented a strand-bias (SB) filter for SNV calls based on testing contingency tables created from plus and minus strand counts of ref and alt reads with a χ^2 test with Yates correction ($P \leq 0.05$). Prior to testing, contingency tables are

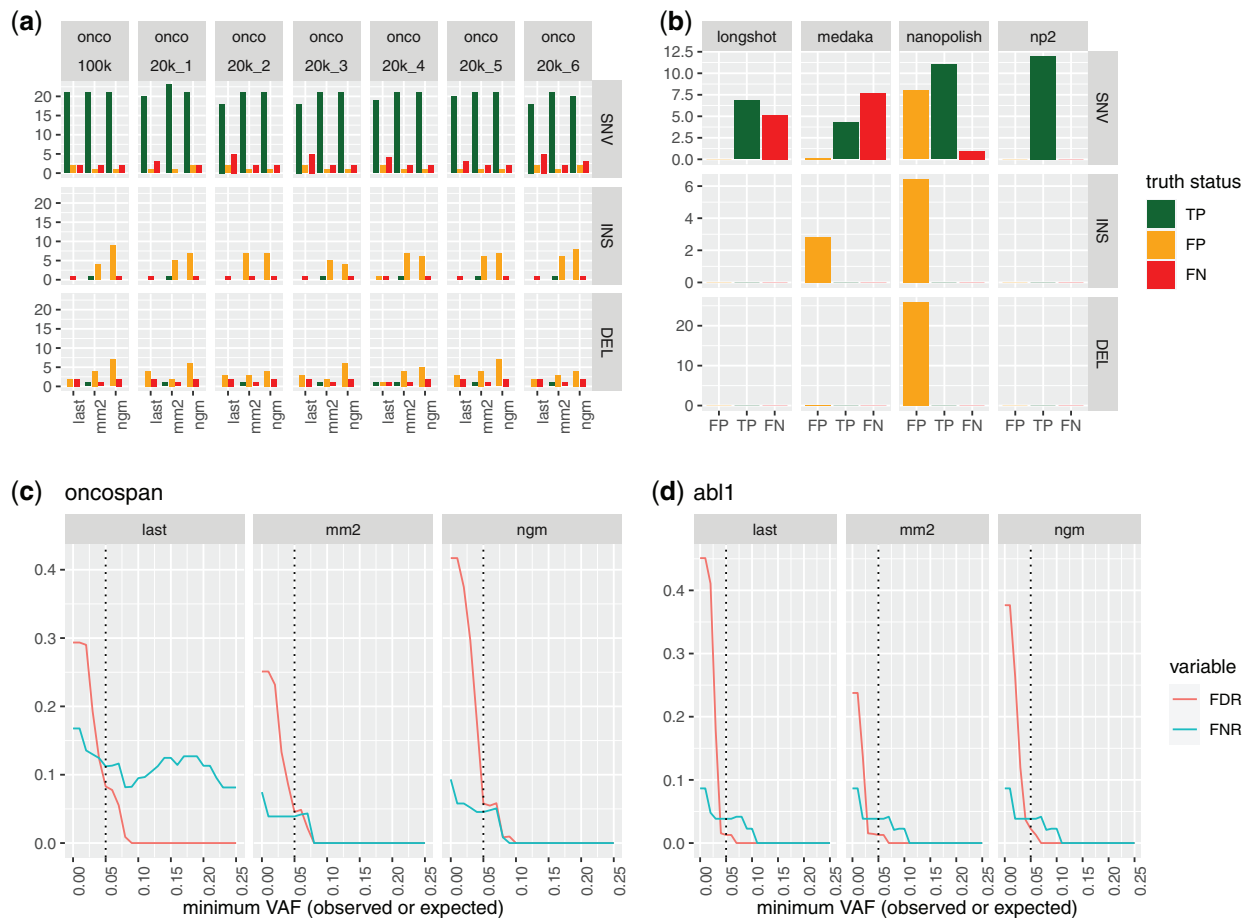


Fig. 1. Np2 performance analysis. (a) Number of true positive (TP), false positive (FP) and false negative (FN) variant calls per oncospan dataset, caller (mm2: minimap2, ngmlr) and variant type for all calls with an observed or expected (for FN) VAF > 5% (See [Supplementary Fig. S12](#) for a corresponding plot without VAF threshold). The plot shows considerable numbers of FP INDEL calls, most of which, however, have low VAF (cf. [Supplementary Fig. S8](#)). (b) Mean numbers of TP, FP, FN calls for longshot, medaka, nanopolish and nanopanel2 (bottom panel) in the 7 oncospan replicates. The evaluation was restricted to calls with expected/observed VAF > 10% (see Discussion). Np2 consistently calls all 12 expected SNVs and produces no FP calls above this VAF threshold. Longshot calls only SNVs and misses several high-VAF variants. Medaka also misses several high-VAF variants and produces some FP INDEL calls. Please note, however, that the latter might include some low frequency variants as medaka does not output allele counts/frequencies. Nanopolish misses one 25% VAF ALK SNV and produces a large number of false positive calls across variant types (SNV, INS, DEL). A screenshot depicting variant calls in the ALK amplicon is provided in [Supplementary Figure S25](#). c and d: Mean false-discovery rate [FDR = FP/(FP + TP)] and false-negative rate [FNR = FN/(FN + TP)] per mapper and minimum VAF for all oncospan (c) and abl1 (d) SNV calls. The rapid drop of these measures with increasing minimum VAF confirms that most false calls have low VAFs. Minimap2 shows the best overall performance while last and ngmlr suffer from increased numbers of FN/FP calls, respectively. Note, however, that FDR in d might still be inflated as we do not have a complete ground truth for clinical datasets

corrected for the flow-cell/sample specific strand ratio calculated from all sequenced reads (cf. [Supplementary Fig. S24](#)). Furthermore, the SB filter is applied only if alt allele carrying reads show a strong imbalance of supporting plus and minus strand reads ($aa_skew = |\log_2(\text{minus}/\text{plus})| > \text{threshold}$). INDELS are filtered only for the latter criteria.

Nanopore's current technology experiences problems in accurately calling the length of homopolymer (hp) runs ([Orsini et al., 2018](#); [Rang et al., 2018](#)), which consequently resulted in high numbers of false-positive calls at the hp run borders before filtering (data not shown). [Supplementary Fig. S21d](#) plots histograms of hp runs with minimum length 3 in the considered amplicons and shows that overall around 23% of all bases reside in such regions. To mitigate this problem, we have implemented a homopolymer filter (HP) that calculates hp lengths up- and downstream of the current pileup and filters calls if those are exceeding a given threshold (default 3 and 2 for up/downstream hp runs) and if one of these hp runs consists of the alt allele (indicating misalignment). For INDELS, we add the lengths of up- and downstream alt-allele hp runs and filter if this sum exceeds a threshold (default: 3). Although this strategy effectively reduces FP calls, it comes at the cost that some calls at hp borders will always be filtered despite their other quality measures which may result in FNs. One example of this is the oncospan FN deletion

MET_2:687GT >G ([Supplementary Fig. S22](#), [Supplementary Table S6](#)). Another, clinically relevant, example is the ABL1 variant E255K, a G > A SNV at the end of a 4-base hp G run that is followed by an A in the reference sequence. For this reason, we recommend that users take a close look at calls that were filtered only by our HP filter and revalidate the respective calls using orthogonal methods where possible. In our datasets (40 samples) we found 398 such calls (an average of 5 SNVs and 5 INDELS per sample) with a high degree of recurrence (99 distinct calls). Only one (0.25%) of these calls was a false negative and the median VAF of calls was low: 2% (SNV), 4% (INS) and 8% (DEL), respectively ([Supplementary Fig. S3](#)).

Finally, we calculate a Phred-scaled quality score [0; 100] based on different pileup statistics: for SNVs, this score is $\min(100, -\log_{10}(\text{non_alt_prob}))$ where non_alt_prob is the mean of all normalized non-alternate allele guppy probabilities. For insertions and deletions it is based on hp length, aa_skew and average base quality (see Supplement for details). Variants are filtered as 'low quality' (LQ) based on configurable, variant type specific thresholds. Further threshold-based filters include low (raw) allele frequencies (AF), low read depth (DP), low mean base qualities (BQ) and high strand-imbalance (SI) that is based on the ratio of all covering plus and minus strand reads.

2.6 Haplotype maps

Np2 will consider all unfiltered variant calls for calculating haplotype maps that may inform users about the distribution of sub-clones in a sample. First, we iterate all reads and check for presence of the respective alt/ref alleles. We then summarize these data per haplotype and use the resulting data tables to draw heatmaps for visual inspection (Fig. 2d, Supplementary Figs S17–S19). For these heatmaps, we filter the haplotypes by the following criteria: (i) minimum fraction among all considered reads $>0.1\%$, (ii) strand skew $\log_2(\text{plus_strand_reads}/\text{minus_strand_reads}) < 2$, (iii) minimum number of supportive reads >10 , (iv) a maximum of 25 haplotypes is shown. Blue squares in this heatmap indicate presence of the respective variant and colour intensity corresponds to $\log_{10}(\text{reads})$. Grey squares correspond to the reference allele, white squares indicate non (most-frequent) alt or ref alleles or missing coverage of the respective reads. Supplementary Figure S27 compares observed and expected haplotype fractions for 12 samples from *abl1_min_1* and *abl1_min_2* and confirms np2's ability to accurately call haplotypes.

2.7 Performance evaluation

Using the respective truth sets, we count FP, TP, TN and TP calls per sample and calculate false-discovery rates ($FDR = \frac{FP}{FP+TP}$), false-negative rates ($FNR = \frac{FN}{FN+TP}$), true-positive rates ($TPR = \frac{TP}{FN+TP}$) and false-positive rates ($FPR = \frac{FP}{\text{amplicon_length}[kb]}$) where appropriate.

3 Results

We first evaluated np2 with seven downsampled oncospan alignments using variant calls from corresponding Illumina whole-exome sequencing (WES) data as ground truth (Fig. 1a). Overall, we observed four recurrent FN calls (Supplementary Fig. S8, Supplementary Table S6) and a reasonably small number of FP calls, most of which were low-frequency insertions. SNV FDR and FNR dropped quickly from 25% and 9% (1% minimum VAF) to zero (10% minimum VAF) with increasing VAF threshold (Fig. 1c). For *abl1* data, we observed similar numbers with FDR and FNR rapidly dropping from $\sim 24\%$ and 9% to zero when reaching a minimum

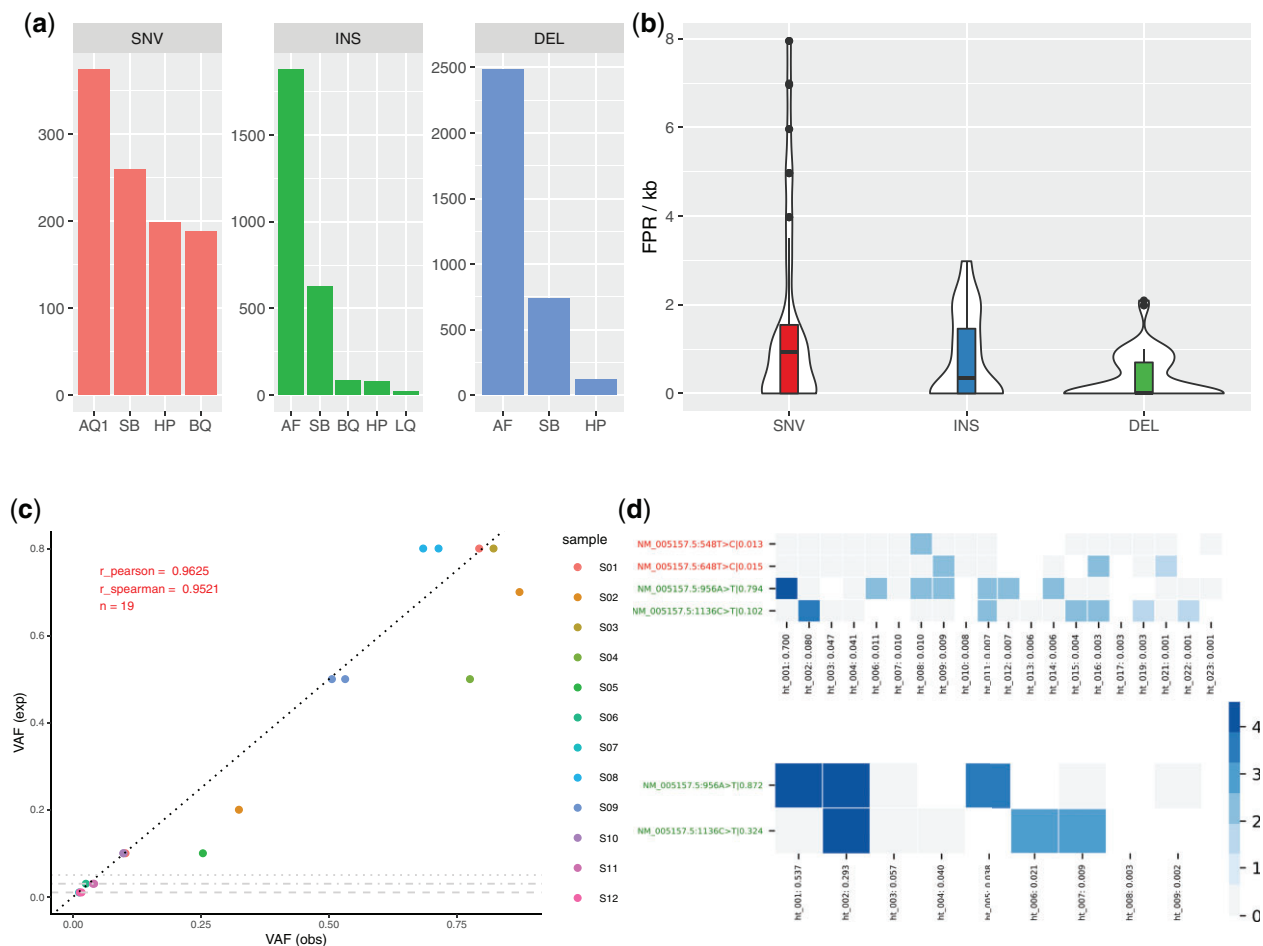


Fig. 2. (a) TN calls per variant type across 40 samples that were filtered exclusively by one single np2 filter. The plots reveal that our allele quality filter (AQ1) is the single most effective filter for suppressing FP SNV calls followed by strand bias (SB) and homopolymer (HP) filters. INDEL FPs are predominately filtered due to low allele frequency (AF), SB and HP filters. A more detailed analysis is provided in Supplementary Figure S5. (b) False-positive rates (FPR) per kb amplicon for all evaluated datasets. Overall we found medium FPR values of 0.95 and 0.7 for SNVs and INDELS. FPR per individual dataset is shown in Supplementary Figure S11. (c) Strong correlation between expected (exp) and observed (obs) VAF for 18 (single and compound) benchmark variants of *abl1_min_1*, particularly for low-frequency variants. Compound variants have equal expected and similar observed VAF (e.g. S08; all 12 samples are described in Supplementary Table S7). One variant with an expected VAF of 1% was not called by np2 and is not shown, horizontal grey lines show calling thresholds for deletions, insertions and SNVs, respectively. (d) Exemplary haplotype maps as plotted by np2 for two benchmark samples containing the same mutations (Y axis, green labels). Red labels denote low-frequency false positive calls; raw VAF as determined by np2 after the pipe symbol). The X axes list found haplotypes sorted by their relative fraction (number after the colon). Blue and grey squares denote present/absent variations, white squares indicate that the respective reads did not span the respective positions or had a deletion at this position. Colour intensity corresponds to $\log_{10}(\text{reads})$. In the upper sample, both mutations were introduced using different plasmids. Targeted VAFs were 80% E255V (ht_001), 10% T315I (ht_002) and 10% wt (ht_003 + 4), respectively. The lower sample contained a mix of 70% single E255V (ht_001), 20% compound E255V+T315I (ht_002) and 10% wt (ht_003 + 4). The haplotype maps confirm the respective linkage status of the benchmark variants although we find a low-frequency ht in the upper map carrying both mutations (ht_011; freq = 0.7%), possibly resulting from sequencing errors. Supplementary Figure S27 shows haplotype calling statistics for 12 benchmark samples from *abl1_min_1* and *abl1_min_2*

VAF threshold of 11% (Fig. 1d, Supplementary Fig. S9). Moreover, FDR of clinical *abl1* data is possibly inflated as we were unable to validate all low-frequency variants due to the detection limits of the used validation methods. Due to the low number of validated insertions and deletions in our datasets we could not conduct the same performance analysis for INDELS. Np2 did, however, call 13 out of 21 (TPR 62%) INDELS (2 deletions and one insertion with mean expected VAF 5.3%) over all oncospan replicates and did not produce any unfiltered high VAF INDEL calls (Supplementary Figs S10 and S11). Together, these results confirm that np2 can call low-frequency SNVs and INDELS with acceptable false discovery rates and reliably calls variants with VAF above 10%. Further investigating the produced FPs, we could observe a high degree of recurrence (Supplementary Fig. S12) which indicates a systematic cause and offers possibilities for further optimisations, e.g. by filtering variant calls with high recurrence across multiple independent samples.

3.1 Filtering stats

We then investigated which of np2's filters are most effective for suppressing FP calls by analyzing the filter subsets assigned to >200k true-negative (TN) calls in 40 samples (Supplementary Fig. S4). For SNVs, we found our novel allele quality filter (AQ1, Section 2) to be the single most effective filter preventing 375 false-positives that would not have been detected by any other filter. For INDELS, allele frequency was the most effective single filter which is a direct result of using respective VAF thresholds in our evaluation runs. Besides this, filtering for strand bias (SB) and homopolymer runs (HP) seemed particularly effective for all variant types (Fig. 2a).

3.2 Variant caller comparison

As np2 is the first available somatic variant caller for Nanopore data we chose to compare its performance to three state-of-the-art germline variant callers (Fig. 1b): Medaka (Oxford Nanopore Technologies Ltd, 2021), Longshot (Edge and Bansal, 2019), and Nanopolish (Loman et al., 2015). This evaluation was restricted to oncospan calls with VAF >10% (see Supplement for used command lines). Np2 called all expected variants for this dataset and did not produce any FP calls above this VAF threshold. In contrast, all evaluated germline callers produced several false positive and negative SNV and INDEL calls. While these results are reasonable because those tools were not developed as somatic variant callers they underline np2's ability to efficiently suppress high-frequency FPs. To further demonstrate this, we calculated false-positive rates per kilobase amplicon (FPR) across all datasets with np2's default VAF thresholds of 1% for SNVs, 3% for insertions and 5% for deletions, respectively, and found median FPR values of 0.95 and 0.7, respectively, for SNVs and INDELS (Fig. 2b, Supplementary Fig. S13; note that the lower FPR for INDELS results from the more stringent AF filtering) which correspond to one FP SNV/INDEL call per 1053/1429 amplicon bases, respectively.

3.3 Reproducibility and detection limits

To measure how accurately np2 can reproduce expected VAFs from a plasmid dilution series, we evaluated it using two benchmark datasets with known concentrations of variant-carrying plasmids (Section 2). We found a high correlation (r_{pearson} 0.96 and 0.98, respectively, Fig. 2c, Supplementary Fig. S7) between expected and observed VAF values. Furthermore, we observed very similar reported VAFs for pairs of variants located on the same plasmid. Np2 called 28 of 31 variants (90%) down to a VAF of 1%. All three FN SNV calls had an expected VAF of 1% and when considering only such low-frequency variants, np2 called 4/7 (FNR = 0.43).

We were interested in the minimum required read depth for np2 and randomly downsampled the data from a MinION flow-cell (*abl1_min_3*) to sets of 1k, 5k, 10k and 25k reads per sample, respectively. We then compared the resulting np2 call sets and found high correlation despite the relatively large difference in read numbers (Supplementary Fig. S15). TPR compared perfectly between the downsamples (Supplementary Fig. S16). Encouraged by these results, we sequenced one benchmark sample and eight

multiplexed clinical samples on two 'Flongle' flow-cells to evaluate the suitability of these smaller flow-cells for clinical applications. Across these 9 samples we called all 10 expected variants and again obtained only few low-frequency FPs (Supplementary Fig. S20). In addition, we observed good variant calling performance (no FN and only 3 FP over 7 replicates) on a lowly covered oncospan amplicon (*KIT_1*) with a median coverage of only 330 reads in the 20k samples (Supplementary Table S2). Together, this suggests that np2 does not require high read depth to accurately call somatic variants and indicates its possible application in targeted Nanopore sequencing (Kovaka et al., 2020; Payne et al., 2020).

4 Discussion

Nanopanel2 is the first somatic variant caller for Nanopore panel sequencing data and our analysis shows that np2 accurately calls variants above 10% VAF and reaches good performance for low-frequency variants down to its currently recommended thresholds of 1% VAF for SNVs and 5% VAF for INDELS. Reported VAFs showed very high correlation to the expected values. Furthermore, np2 provides information on the distribution of subclones in a sample by directly calling haplotypes from (amplicon-spanning) long reads. Together, this makes it a viable alternative to current amplicon sequencing methods for clinical and research applications.

Despite the promising results of our evaluation using benchmark and clinical samples, more work is required to improve np2's calling accuracy. One particular problem is the accurate calling in and around homopolymer runs, a known problem of Nanopore's current sequencing technology (Rang et al., 2018). Although our hp filter effectively removes most FPs in such regions, this also results in FN calls and for the time being, we recommend users to revalidate calls that were filtered exclusively by the HP filter (see Section 2). We do, however, expect this problem to be mitigated in the future due to recent improvements in nanopore chemistry, e.g. R10 nanopores (Van der Verren et al., 2020).

Nanopore's technology enables fast and cost-efficient on-demand sequencing which is particularly attractive for small diagnostic labs that cannot wait until larger (short-read) flow-cells have been filled with multiplexed patient samples. Cost efficiency is also key for clinical management that includes early detection and continuous monitoring of disease relevant (resistance) clones (Orsini et al., 2018). Here, we have showcased a clinical diagnostics workflow for CML patients that is cheap (e.g. 8 multiplexed samples on a single 90\$Flongle flow-cell with ~\$50 library preparation costs) and fast (library prep 2–3 h, sequencing: 12–24 h, bioinformatics ~14 h).

Besides clinical applications, np2 has other obvious usage scenarios including metagenomic and full-length virus sequencing (Oude Munnink et al., 2020; Xu et al., 2020). In an initial proof-of-concept experiment, we have applied np2 to SARS-Cov2 Nanopore direct RNA sequencing data with promising results (Supplementary Fig. S26). Our experiments with low-coverage amplicons/downsamples furthermore indicate that np2's allele quality filtering method could possibly also be applied to target-enriched (Kovaka et al., 2020; Payne et al., 2020) or even whole-genome nanopore data (Bowden et al., 2019) although this is not possible with the current np2 implementation. We furthermore speculate that our allele quality filtering method could be combined with current approaches for read polishing and germline variant calling (Ahsan et al., 2020; Vaser et al., 2017; Wick et al., 2017).

Acknowledgements

The authors thank Philipp Rescheneder and Phillip James (Oxford Nanopore Technologies Ltd.) for providing the oncospan datasets and helping with data interpretation and Thomas Ernst (University Jena, Germany) for the permission to test samples with quantified mutation loads derived from the EUTOS international control round for deep sequencing analysis of BCR-ABL mutations in CML.

Author contributions

N.P., T.L. and S.P. conceptualized the project and designed the experiments. N.P. designed and implemented the software and conducted all *in silico* experiments. S.P. prepared and sequenced all BCR-ABL1 datasets. All authors were involved in interpretation of the data, discussed the results and commented on the manuscript. N.P. wrote the manuscript with contributions from all authors.

Financial Support: none declared.

Conflict of Interest: Nanopanel2 is dual-licensed under a copyleft open-source license and a proprietary software license. N.P. will receive fees resulting from commercial software licensing.

References

- Ahsan,U. *et al.* (2020) NanoCaller for accurate detection of SNPs and indels in difficult-to-map regions from long-read sequencing by haplotype-aware deep neural networks. *bioRxiv*, doi: 10.1101/2019.12.29.890418.
- Alikian,M. *et al.* (2012) BCR-ABL1 kinase domain mutations: methodology and clinical evaluation. *Am. J. Hematol.*, **87**, 298–304.
- Ameur,A. *et al.* (2019) Single-molecule sequencing: towards clinical applications. *Trends Biotechnol.*, **37**, 72–85.
- Bowden,R. *et al.* (2019) Sequencing of human genomes with nanopore technology. *Nat. Commun.*, **10**, 1869.
- Byrgazov,K. *et al.* (2018) BCR-ABL1 compound mutants display differential and dose-dependent responses to ponatinib. *Haematologica*, **103**, e10–e12.
- Cao,H. *et al.* (2015) De novo assembly of a haplotype-resolved human genome. *Nat. Biotechnol.*, **33**, 617–622.
- Dunn,T. *et al.* (2019) Pisces: an accurate and versatile variant caller for somatic and germline next-generation sequencing data. *Bioinformatics*, **35**, 1579–1581.
- Edge,P. and Bansal,V. (2019) Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat. Commun.*, **10**, 4660.
- Gabert,J. *et al.* (2003) Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia – a Europe Against Cancer program. *Leukemia*, **17**, 2318–2357.
- Hochhaus,A. *et al.* (2020) European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia*, **34**, 966–984.
- Khorashad,J.S. *et al.* (2013) BCR-ABL1 compound mutations in tyrosine kinase inhibitor-resistant CML: frequency and clonal relationships. *Blood*, **121**, 489–498.
- Kielbasa,S.M. *et al.* (2011) Adaptive seeds tame genomic sequence comparison. *Genome Res.*, **21**, 487–493.
- Kono,N. and Arakawa,K. (2019) Nanopore sequencing: review of potential applications in functional genomics. *Dev. Growth Different.*, **61**, 316–326.
- Kovaka,S. *et al.* (2020) Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *bioRxiv*, doi: 10.1101/2020.02.03.931923.
- Krishnakumar,R. *et al.* (2018) Systematic and stochastic influences on the performance of the MinION nanopore sequencer across a range of nucleotide bias. *Sci. Rep.*, **8**, 3159.
- Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, **34**, 3094–3100.
- Logsdon,G.A. *et al.* (2020) Long-read human genome sequencing and its applications. *Nat. Rev. Genet.*, **X**, 1–18.
- Loman,N.J. *et al.* (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods*, **12**, 733–735.
- Machova Polakova,K. *et al.* (2015) Next-generation deep sequencing improves detection of BCR-ABL1 kinase domain mutations emerging under tyrosine kinase inhibitor treatment of chronic myeloid leukemia patients in chronic phase. *J. Cancer Res. Clin. Oncol.*, **141**, 887–899.
- Orsini,P. *et al.* (2018) Design and MinION testing of a nanopore targeted gene sequencing panel for chronic lymphocytic leukemia. *Sci. Rep.*, **8**, 11798.
- Oude Munnink,B.B. *et al.* (2020) Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Medicine*, **26**, 1405–1410.
- Oxford Nanopore Technologies Ltd. (2021) medaka: sequence correction provided by ONT research. <https://github.com/nanoporetech/medaka> (17 May 2021, date last accessed).
- Payne,A. *et al.* (2020) Nanopore adaptive sequencing for mixed samples, whole exome capture and targeted panels. *bioRxiv*, doi: 10.1101/2020.02.03.926956.
- Preuner,S. *et al.* (2008) Quantitative monitoring of cell clones carrying point mutations in the bcr-abl tyrosine kinase domain by ligation-dependent polymerase chain reaction (ld-pcr). *Leukemia*, **22**, 1956–1961.
- Rang,F.J. *et al.* (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.*, **19**, 90.
- Sedlazeck,F.J. *et al.* (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods*, **15**, 461–468.
- Van der Verren,S.E. *et al.* (2020) A dual-constriction biological nanopore resolves mononucleotide sequences with high fidelity. *Nat. Biotechnol.*, **38**, 1415–1420.
- Vaser,R. *et al.* (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.*, **27**, 737–746.
- Wick,R.R. *et al.* (2017) Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genomics*, **3**, e000132.
- Xu,Y. *et al.* (2020) NanoSPC: a scalable, portable, cloud compatible viral nanopore metagenomic data processing pipeline. *Nucleic Acids Res.*, **48**, W366–W371.
- Zabriskie,M.S. *et al.* (2014) BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia. *Cancer Cell*, **26**, 428–442.