The Inter and Intra-Tumoural Heterogeneity of Subclonal Reconstruction

Vinayak Bhandari[1,2,*], Lydia Y. Liu[1,*], Adriana Salcedo[1,2], Shadrielle M. G. Espiritu[1], Quaid D. Morris[3,4,5,6], Paul C. Boutros[1,2,6,7]

[1] Ontario Institute for Cancer Research, Toronto, Canada

[2] Department of Medical Biophysics, University of Toronto, Toronto, Canada

[3] Department of Computer Science, University of Toronto, Toronto, Canada

[4] Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada

[5] Donnelly Centre for Cellular and Biomolecular Research, Toronto, Canada

[6] Vector Institute for Artificial Intelligence, Toronto, ON M5G 1M1, Canada

[7] Department of Pharmacology and Toxicology, University of Toronto, Toronto, Canada

*These authors contributed equally to this work

Addresses for correspondence:

Dr. Paul C. Boutros

661 University Avenue, Suite 510

MaRS Centre

Toronto, Ontario, Canada,

M5G 0A3

Email: Paul.Boutros@oicr.on.ca

Phone: 416-673-8564

Abstract

Background. Each tumour is a community of cells related through a common ancestry. Each cell in this community faces a slightly different set of selective pressures, and evolves at least partially independently, leading to development of subclonal populations. Recent studies have used whole-genome sequencing data to estimate the origins of this clonal heterogeneity, and have linked it to clinical outcomes like response to curative treatment. A wide variety of algorithms have been developed to perform this type of subclonal reconstruction, but their variability and consistency is largely unknown. To address this issue, we evaluated four pipelines to reconstruct the evolutionary histories of 293 localized prostate cancers from single samples. We then applied these pipelines to 10 tumours with multi-region sampling to probe the intra-tumoural heterogeneity of subclonal reconstruction.

Results. We identify extensive variance across modern subclonal reconstruction pipelines in the subclonal architectures they predict and in their assignments of CNAs and SNVs to different parts of the evolutionary tree. Further, pipelines showed consistent types of bias, with those using SomaticSniper and Battenberg preferentially predicting monoclonal architectures while those using MuTect tending to predict multiple populations of cancer cells. In general, there was greater concordance in the assignment of clonal variants than in that of subclonal ones. Subclonal reconstructions using multi-region sampling showed that single-sample reconstructions systematically underestimate intra-tumoural heterogeneity, detecting on average fewer than half of the cancer cell populations identified by multi-region sequencing.

Conclusions. Analysis of a large cohort of single- and multi-region tumour whole genomes shows that selection of variant detection and subclonal reconstruction algorithms significantly biases the results of subclonal reconstruction. These biases suggest caution in interpreting the specific architectures and subclonal variants identified, particularly from single-sample reconstructions.

Keywords

Tumour heterogeneity, subclonal reconstruction, cancer genomics

Background

Understanding tumour heterogeneity and subclonal architecture is important for the elucidation of the mutational processes underlying tumorigenesis and treatment resistance (1–4). Most studies of tumour heterogeneity have focused on small number patient cohorts with multi-region sequencing (5–11). Despite their small sample sizes, these studies have provided remarkable insight, demonstrating multiple subclones within a single tumour, clonal relationships between primary and metastatic tumours and evidence for multiple primary tumours within a single patient.

More recently, several groups have begun applying these techniques to large cohorts of single-sample tumour whole genomes. For example, we reconstructed the subclonal architectures of 293 localized prostate cancers using whole-genome sequencing of a single-region of the index lesion (12). The larger sample sizes made feasible by single-region studies allow quantitative identification of specific mutational events that are biased to occur early during tumour development. Further, patients with less subclonal diversity (*e.g.* with only a single detectable population of cancer cells; termed *monoclonal*) tend to have superior clinical outcomes to those with more subclonal diversity (*e.g.* those with highly *polyclonal* tumours).

A broad variety of algorithms have been developed to reconstruct the subclonal architecture of cancers (13–18), but there has not been a systematic comparison of the features and consistency of their reconstructions on a large dataset. It is thus unclear to what extent these pipelines agree on large cohorts of real data, and whether specific pipelines are biased towards certain types of reconstructions. It is further unclear to what extent single-sample reconstructions differ from the multi-region reconstructions, raising questions about the magnitude of underestimation present in large-cohort studies.

To address these gaps in the field, we applied four pipelines comprised of well-established subclonal copy-number alteration (CNA) and single-nucleotide variant (SNV) detection, linked to a single reconstruction algorithm, to a set of 293 high-depth tumour-normal pairs (12,19) and 10 tumours with multi-region sequencing (8,20). We quantify differences in variant detection and predictions of clonality, and overall subclonal architecture, generating useful guidance for the community and a resource for improving existing methods and benchmarking new ones.

Results

To investigate the inter- and intra-tumoural heterogeneity of subclonal reconstruction we reconstructed the subclonal architectures of 293 primary localized prostate tumours using four pipelines (**Figure 1**, **Supplementary Table 1**). Each patient had whole-genome sequencing of a region of the index lesion that was macro-dissected to > 70% tumour cellularity (mean coverage 63.9 ± 16.7) and of matched blood reference tissue (mean coverage 41.2 ± 9.0), as reported previously (12). We focused on quantifying the roles of subclonal variant detection, and kept the reconstruction approach consistent, using PhyloWGS (13), while varying subclonal CNA detection to use either TITAN or Battenberg (7,21), and subclonal SNV detection to use either SomaticSniper or MuTect (22,23). To quantify the heterogeneity that arises in subclonal reconstruction from spatially sampling the same tumour, we also examined 10 tumours with multi-region sequencing (8,20). For multi-region sequencing, subclonal reconstruction was performed both with all regions together and with each region individually, allowing estimation of the variability of single-region reconstruction.

Subclonal reconstructions were unsuccessful in a number of instances due to excessive runtime (> 3 months) or failure to produce non-*polytumour* (*i.e.*, consisting of multiple independent primary tumours) phylogeny solutions after rerunning with 12 different random number generator seeds. Additional reasons for reconstruction failure include producing solutions where the cellular prevalence of the only tumour cell population detected was below 10%, or lack of supporting CNAs and SNVs for all cancer cell populations detected. The rate of reconstruction failure varied across the pipelines for single-region reconstructions (SomaticSniper-TITAN: 1.4%, SomaticSniper-Battenberg: 2.0%, MuTect-TITAN: 4.4%, MuTect-Battenberg: 9.6%, mean failure rate 4.4% ± 3.2). Multi-region reconstructions failed in one instance in the SomaticSniper-Battenberg pipeline due to a lack of shared SNVs or CNAs between samples. Overall, the failure rate of 9.6% for MuTect-Battenberg led to decreased power for all downstream analyses and use of this pipeline must be carefully considered for small-scale projects.

Clonal and Subclonal CNAs

We first assessed the clonal and subclonal copy number profiles of localized prostate cancers based on the four pipelines in the 293 samples with single-region sequencing (**Figure 2A-B,**

Supplementary Table 2). Previous work within this cohort using the SomaticSniper-TITAN pipeline identified four subtypes of patients with distinct clonal CNA profiles and three subtypes with distinct subclonal CNA profiles (12), so we first examined the average profiles within these subtypes. Hallmark prostate cancer copy-number events such as deletions in 8p and gains in 8q are seen clonally across the pipelines, particularly in subtype *C* (**Figure 2A**). CNA profiles in subtype *A* appear to be quieter for pipelines using TITAN to call CNAs and the MuTect-TITAN pipeline seems to call a large number of small CNAs for patients in subtype *D*. Interestingly, the average subclonal copy number is almost always above two across the genome in subclonal subtype *f* in pipelines using TITAN to call CNAs. (**Figure 2B**).

We next assessed the concordance of these pipelines in their identification of clonal and subclonal CNAs. We calculated the proportion of CNA calls where one pipeline called a CNA and the other pipeline called either the opposite CNA (*i.e.*, gain *vs.* loss) or neutral for the same 1 Mbp genomic bin in the same sample. We found significantly greater agreement for clonal CNA calls compared to subclonal CNA calls (all FDR $< 2.0 \times 10^{-8}$, Mann-Whitney U-test; **Figure 2C**). Similarly, pipelines using the same CNA caller tended to agree, although even here divergence was common. For example, 87% of clonal CNA calls were concordant between SomaticSniper-TITAN and MuTect-TITAN pipelines as compared to 75% of subclonal calls. By contrast, when the CNA caller was changed, 45% of clonal CNA calls were concordant between MuTect-Battenberg and MuTect-TITAN pipelines as compared to 22% of subclonal CNA calls. Thus we observe diversity in clonal CNA profiles across pipelines, but much larger diversity for subclonal CNA profiles.

Next, we evaluated whether estimates of the timing of specific CNAs as early (*i.e.*, clonal) or late (*i.e.*, subclonal) were affected by the subclonal reconstruction pipeline. The raw number of genes in which CNAs were identified as being statistically more frequent (FDR < 0.05, Fisher's exact test) early or late differed dramatically across pipelines, from a high of 5,592 for MuTect-TITAN to a low of 295 for SomaticSniper-Battenberg. A consensus set of 270 genes show a bias in timing in all pipelines in a consistent direction (**Figure 3A**, **Supplementary Table 3**). As expected from the discordance in subclonal CNA calls across the pipelines, all of these 270 genes preferentially occur clonally, and were enriched in *TP53*-based regulation of death receptors, *TRAIL* signaling and natural killer cell mediated cytotoxicity (FDR < 0.05; **Figure 3B**,

Supplementary Table 4). These data suggest that execution of multiple pipelines may be a valuable way to identify a low false-positive set of core predictions.

Clonal and Subclonal SNVs

We next evaluated the influence of pipelines on clonal and subclonal SNVs detection and classification. All four pipelines identified similar numbers of clonal SNVs (SomaticSniper-TITAN: 1007 ± 751, SomaticSniper-Battenberg: 1018 ± 724, MuTect-TITAN: 1782 ± 1214, MuTect-Battenberg: 1972 ± 1130; mean ± standard deviation). By contrast, pipelines using MuTect detected an order of magnitude more subclonal SNVs than those using SomaticSniper (SomaticSniper-TITAN: 118 ± 206, SomaticSniper-Battenberg: 29 ± 110, MuTect-TITAN: 2123 ± 1341, MuTect-Battenberg: 1763 ± 1197; mean ± standard deviation; **Figure 4A**). This suggests that pipelines employing MuTect are more likely to detect subclonal SNVs, and may lead to more false positives and fewer false negatives.

For each pair of pipelines, we calculated the proportion of genes where one pipeline called a SNV in a gene while the other called the gene wildtype (**Figure 4B, Supplementary Table 5**). Calls for clonal SNVs were in high agreement across pipelines (average concordance = 76.9 ± 1.0%) but subclonal SNV calls were much more discordant (average concordance = 39.0 ± 37.8%), particularly between pipelines using different SNV callers. For example, 91% of subclonal SNV calls by SomaticSniper-Battenberg and SomaticSniper-TITAN were identified in both pipelines. By contrast, 91% of subclonal SNVs from SomaticSniper-Battenberg and MuTect-Battenberg were identified by only one of the pipelines. These data match the trends seen with CNAs, above: there are substantial discrepancies in the subclonal mutational landscape across pipelines, but clonal landscapes are generally consistent.

To understand the influence of these differences on the identification of specific driver genes we examined the clonality of five genes known to be driven by recurrent somatic SNVs and eight known to be driven by recurrent somatic CNAs in prostate cancer (12,19). The driver SNVs we examined were *ATM*, *FOXA1*, *MED12*, *SPOP* and *TP53* (**Figure 5A**), and the driver CNAs were *CDH1*, *CDKN1B*, *CHD1*, *MYC*, *NKX3-1*, *PTEN*, *RB1* and *TP53* (**Figure 5B**). These driver events were overwhelmingly detected early during tumour evolution with 89.7 ± 11.4 % (mean across all pipelines ± standard deviation) of all driver alterations detected clonally across the pipelines. There was also broad consensus in these calls (*e.g.*, all four pipelines calling a clonal

mutation in a driver gene in the same sample). For example, of the samples where a clonal gain of *MYC* was called by any of the pipelines, all four pipelines called the alteration in 74% of cases (**Supplementary Figure 1**). One outlier where calls were very divergent was *MED12*, where all four pipelines never called a mutation with the same clonality.

Subclonal Architecture Reconstructions from Single Samples

Given these significant differences in the attribution of individual mutations to different stages of the tumour phylogeny, we sought to understand whether the underlying subclonal architecture was being predicted consistently across pipelines. We therefore examined the number of subclones detected in each sample across each pipeline (**Figure 6A**). Reconstructions with SomaticSniper-Battenberg predominantly detected monoclonal tumours: only one population of cancer cells for 81.5% of the samples. By contrast, the two pipelines using MuTect-based SNV calls identified <16% of samples as monoclonal. The SomaticSniper-TITAN analyses were intermediate to these two extremes, predicting monoclonal architecture for 41.2% of samples. These data can be further broken down by the tree structure (**Figure 6B**) where we observe the SomaticSniper-TITAN and MuTect-Battenberg pipelines detecting many bi-clonal samples.

We next directly compared the subclonal architectures predicted for each sample across the four pipelines. Overall, we observed considerable variance in the predicted subclonal architectures. Consistent with the large number of monoclonal structures predicted by SomaticSniper-Battenberg, their predictions were highly discordant with other pipelines (**Supplementary Figure 2A-C**). Predictions of subclonal architecture between the two pipelines using MuTect for SNV calls were highly concordant with clonality predictions (*i.e.*, monoclonal *vs* polyclonal) in agreement for 226/265 tumours (**Supplementary Figure 2D)**. However, of these 226 tumours the specific phylogeny of the evolutionary tree matched for only 84 tumours due to the diversity of structures observed in polyclonal reconstructions. The concordance was also low between pipelines using MuTect to call SNVs and the SomaticSniper-TITAN pipeline, where the pipelines using MuTect called substantially more samples as polyclonal (**Supplementary Figure 2E-F**). Overall, the aforementioned differences in mutation calling are highly visible in downstream subclonal reconstructions.

Subclonal Reconstruction from Single and Multi-Region Samples

Our analyses of a large cohort of single-sample reconstructions highlight large inter-pipeline differences in the identification and clonal-assignment of CNAs and SNVs, and in the reconstruction of phylogenetic trees. To better understand how these results compare to ground-truth, we next focused on a set of ten localized prostate cancers where tumour samples from multiple regions of the tumour were available (30 genomes in total, ranging from 2-4 genomes per tumour). These data allowed us to directly compare single-region reconstructions to multi-region reconstructions, providing an estimate of the extent to which the former underestimates true clonal complexity.

Assignments of SNVs as clonal or subclonal showed substantial variance between single-region and multi-region reconstructions (**Figure 7A**). For example, only 19.9 ± 11.2% of SNV assignments agreed between single- and multi-region reconstructions for the MuTect-TITAN pipeline. The strongest agreement for SNV assignments between single- and multi-region reconstructions were observed for the SomaticSniper-Battenberg pipeline, although considerable inter-tumoural variance was observed (41.5 ± 30.5% of SNV assignments matched, mean ± standard deviation). This high rate of agreement is likely at least in part the result of the preference of this pipeline towards monoclonal reconstructions (*i.e.,* the reconstruction only predicted a single population of cancer cells and all SNVs were assigned to that population). In the mismatches between single-region and multi-region reconstructions, SNVs tend to be classified as clonal in single-region reconstructions while subclonal in multi-region reconstructions. Consistent with simulations (24), this suggests that multi-region reconstructions are able to better define subclonal populations of cells, and subclonal reconstruction on single-regions alone is prone to overlooking such complexities.

We also examined the agreement between single-region and multi-region reconstructions in defining the clonality of CNAs (**Supplementary Figure 3**). Agreements between single- and multi-region CNA assignments were also highly variable, with only 40.1 ± 32.6% and 26.1 ± 28.2% (mean ± standard deviation) of CNAs matching in assignment between the single- and multi-region reconstructions for MuTect-TITAN and SomaticSniper-Battenberg pipelines, respectively. One reason for the low rate of agreement for SomaticSniper-Battenberg may be again be the high rate of monoclonal calls in single-region assessments as compared to the scarcity of monoclonal multi-region reconstructions. However, extensive variance is again

observed in CNA assignments between single-region and multi-region reconstructions, as indicated by the standard deviations of agreements above.

Finally, we quantified the differences in the number of subclones that were detected from single-region and multi-region reconstructions of the 10 tumours using all four pipelines (**Figure 7B, Supplementary Table 6**). Across the pipelines, multi-region reconstructions detected more subclones than single-region reconstructions: 4.5 ± 2.4 subclones were detected with multi-region reconstructions while 2.0 ± 0.85 subclones were detected with single-region reconstructions (mean ± standard deviation). Thus the typical single-sample reconstruction identified somewhat less than half of the subclones present in the full tumour. Indeed, multi-sample reconstructions detect significantly more subclones even within the index lesion sample compared to the single-sample reconstruction of the index lesion ($p = 2.50 \times 10^{-3}$, Mann-Whitney U-Test, median number of $\text{subclones}_{\text{Index lesion, single-sample reconstruction}} = 2$, median number of $\text{subcones}_{\text{Index lesion, multi-sample reconstruction}} = 3$; **Supplementary Figure 4**). However, multi-region reconstructions did not always detect more subclones. For example, pipelines using SomaticSniper detected more subclones in single-region samples for CPCG100 compared to multi-region reconstructions for the same tumour. This was also the case for two reconstructions from the MuTect-Battenberg pipeline. Overall, these data suggest that multi-region reconstructions tend to find more subclones compared to reconstructions from single-regions and can more precisely define the mutational landscape of the clonal population.

Discussion

It is difficult to benchmark the accuracy of subclonal methodologies since a gold-standard experimental dataset is not yet available. Simulation frameworks are of great value, but may not fully recapitulate the error-profiles and signal-biases of real data. Here, to complement ongoing efforts to benchmark subclonal reconstruction using simulated data (25), we systematically estimated the subclonal architecture of 293 samples with four pipelines. These data provide the first experimental lower-bound on the algorithmic variability of tumour subclonal reconstruction in a large high-depth whole-genome sequencing cohort. We complement these data by assessing the intra-tumoural heterogeneity of subclonal reconstruction pipelines across a set of 10 multi-region tumours, providing an estimate of the degree to which single-sample reconstructions underestimate clonal complexity.

The two subclonal CNA detection algorithms here differ modestly in their features, while large differences are driven by changing the SNV detection approach. Differences between these mutation callers lead to major divergences in subclonal reconstruction, where pipelines using MuTect tend to find extensive subclonal diversity. Indeed one pipeline combination, SomaticSniper-Battenberg, led to the vast majority of reconstructions indicating the presence of only a single population of cancer cells. The clinical impact of these differences is considerable, given findings that tumours with polyclonal architecture are at elevated risk for both relapse and distant metastasis after primary treatment (12).

Our evaluation of subclonal reconstruction using data from spatially distinct regions of tumours found that reconstructions relying on a single sample systematically underestimate the number of subclones in a tumour: multi-region reconstructions tend to find twice as many subclones as single-region reconstructions. This agrees with previous work showing the distinct mutational profiles of prostate cancer samples from spatially distinct regions of the same tumour (8) and reinforces the hypothesis that sufficient sampling will uncover multiple subclones in nearly all prostate cancers. This may also suggest that mutation callers that are considered highly sensitive still cannot fully overcome the inherent spatial heterogeneity in tumours.

Larger datasets are necessary to better evaluate the performance of subclonal reconstruction methodologies. Efforts towards this are ongoing *via* crowd-sourced benchmarking efforts with

simulated data, such as the DREAM SMC-Het benchmarking challenge (25). One can anticipate that large single-cell sequencing datasets will arrive to further push the boundaries of accuracy for subclonal reconstruction algorithms. In the meantime, this work involving a large clinical cohort will aid in refining subclonal reconstruction methods and evaluating the subclonal architecture of cancer samples.

Conclusions

We present a systematic assessment of the variability across four subclonal reconstruction pipelines for a large, well-characterized dataset. We observe considerable inter-tumoural heterogeneity in reconstruction solutions for tumours with single-region sequencing, reflecting the choice of SNV and CNA callers in downstream reconstructions. Further, we show that subclonal reconstructions relying on a single-sample systematically underestimate the intra-tumoural heterogeneity of tumours irrespective of the mutational callers utilized. This work sets the stage for future systematic assessments of subclonal reconstruction algorithms and their development to better reconstruct the histories of cancers.

Methods

Patient Cohort

We aggregated a retrospective cohort of localized prostate tumours with patient consent and Research Ethics Board approval from published datasets, with whole-genome sequencing of tumour samples and matched blood-based normal samples (12,19,20,26–29). The cohort includes 293 patients with tumour samples from the index lesion and 10 patients with multiple samples from intraductal carcinoma and juxtaposed adjacent invasive carcinoma. For patients receiving radiotherapy, the index tumour was identified on transrectal ultrasound and sampled by needle biopsies (TRUS-Bx) and was deemed the largest focus of disease that was confirmed pathologically. A fresh-frozen needle core ultrasound-guided biopsy to this index lesion was obtained for macro-dissection. For patients receiving surgery, the index tumour was identified macroscopically by a GU expert pathologist at the point of surgery and later sampled and biobanked. A fresh-frozen tissue specimen from the index lesion was then obtained from macro-dissection. Details of the patient cohort have been described previously (12,20).

We focused on patients with clinical intermediate-risk disease as defined by NCCN, with intermediate risk factors (T2b or T2c disease, ISUP Grade Group 2 or 3 or pre-treatment prostate specific antigen (PSA) serum levels between 10-20 ng/mL). All patients received either precision image-guided radiotherapy or radical prostatectomy with no randomization or classification and were hormone naive at time of therapy. Four patients in the multi-region sequencing cohort carried germline BRCA2 mutations and had formalin-fixed paraffin-embedded tissues instead of fresh-frozen. Sample regions suitable for micro-dissection (tumour cellularity > 70%) were marked by genitourinary pathologists and manually macro-dissected, followed by DNA extraction and sequencing.

Whole-genome sequencing data analysis

Protocols for whole-genome sequencing data generation and processing have been previously described (12,19,20). Briefly, raw sequencing reads from the tumour and normal samples were aligned against human reference genome build hg19 using bwa-aln (v0.5.7) (30). Lane-level BAMs from the same library were merged and duplicates were marked using picard (v1.92). Local realignment and base quality recalibration were performed together for tumour/normal pairs using GATK (v.2.4.9) (31). Tumour and normal sample-level BAMs were extracted

separately, had headers corrected with SAMtools (v0.1.9) (32) and were indexed with picard (v1.107). ContEst (v1.0.24530) (33) was used to estimate lane-level and sample-level sample mix-up and lane-level cross-individual contamination on all sequences, with no significant contaminated detected.

Tumour Somatic Mutation Assessment

We identified subclonal copy number aberrations from whole-genome sequencing data using Battenberg (v2.2.6) (7) and TITAN (v1.11.0) (21). First, Battenberg (v2.2.6) was installed with underlying ASCAT (v2.5) (34) using the installation and running wrapper cgpBattenberg (v3.1.0). Required reference files were downloaded as instructed in https://github.com/Wedge-Oxford/battenberg and further required data files were generated as instructed in https://github.com/cancerit/cgpBattenberg. An ignore file was created for the genome assembly hg19 to exclude all chromosomes not in 1-22 and X. Battenberg (v2.2.6) was run with -gender of xy for male patients and -t of 14 to run using 14 threads, and otherwise default parameters. The resulting primary solution was subjected to manual refitting in situations meeting the following criteria: 1) the solution involved a high copy number segment with high BAF and low logR, indicating an unrecognized homozygous loss event, 2) nearly all copy number segments were subclonal, 3) there were unreasonably high copy numbers up to infinity. Refitting was performed until the concerns for refitting were resolved or for three attempts after which the original solution was accepted. The CNAs obtained from the primary solution, along with tumour cellularity and ploidy were used for further analysis. We have described subclonal copy number analysis using TITAN (v1.11.0) previously in detail (12). Briefly, TITAN (v1.11.0) was run through the Kronos (v1.12.0) (35) pipeline for whole-genome sequence preprocessing and subclonal copy number assessment. GC and mappability files for bias correction were prepared using HMMcopy (v.0.1.1) (36)and bowtie (v2.2.6) (37) on the hg19 reference genome. Heterogeneous positions in the sequence data were identified by MutationSeq (v4.3.7) (38) using known dbSNP sites from GATK (v2.4.9). For each whole-genome sequence, TITAN (v1.11.0) made predictions of the existence of one to five subclones based on the given input -numClusters and the solution with the lowest S_Dbw validity index (21) was used to obtain the cellularity, ploidy and subclonal CNAs for downstream analysis.

We used MuTect (v1.1.4) (23) and SomaticSniper (v1.0.2) (22) for the identification of somatic single nucleotide variants from whole-genome sequencing data. MuTect was run to obtain candidate SNVs with dbSNP138 (39), COSMIC (v66) (40) and default parameters except the -tumor_lod option (tumor limit of detection). The -tumor_lod option was set to 10 to increase the stringency of detection. Outputs that contained REJECT were filtered out and the remaining SNV calls were used for downstream analysis. Details for SomaticSniper (v1.0.2) variant calling have been described previously (19). In short, SomaticSniper (v1.0.2) was used to identify candidate SNVs with default parameters except the -q option (mapping quality threshold), which was set to 1 as per developer recommendation. Candidate SNVs were filtered through standard and LOH filtering using a pileup indel file generated on the sequence data using SAMtools (v0.1.6) (32), bam-readcount filtering and false positive filtering. Only high confidence somatic SNVs obtained from the high confidence filter using default parameters were used for further analysis. We performed annotation and filtering on all SNVs, with full details given previously (12). In brief, SNVs obtained by MuTect (v1.1.4) and SomaticSniper (v1.0.2) were annotated with associated genes and functions by ANNOVAR (v2015-06-17) (41) using RefGene, subjected to blacklist filtering to remove known germline contaminants and sequencing artifacts and whitelist filtering through COSMIC (v70) (40). This was done before downstream subclonal reconstruction.

Subclonal Reconstruction of Tumours

We used the cnv-int branch of PhyloWGS (https://github.com/morrislab/phylowgs/tree/cnvint, commit: 3b75ba9c40cfb27ef38013b08f9e089fa4efa0c0) (13) for the reconstruction of tumour phylogenies, as described previously (12). Briefly, subclonal CNA segments and cellularities predicted by Battenberg (v2.2.6) and TITAN (v1.11.0) were parsed using the provided parse_cnvs.py script and were filtered to remove any 1-bp segments. Somatic SNVs obtained using MuTect (v1.1.4) and SomaticSniper (v1.0.2) were also subjected to filtering to remove mutations not at callable bases (where callable bases are those with ≥ 17x coverage for the tumour and ≥ 10x coverage for the normal). We then used the create_phylowgs_inputs.py script to generate four sets of PhyloWGS inputs for each sequence sample, by combining TITAN-SomaticSniper, TITAN-MuTect, Battenberg-SomaticSniper and Battenberg-MuTect as CNA-SNV mutation caller pairs. The PhyloWGS script evolve.py reconstructed phylogenies using default parameters for each input set. Thus, four pipelines by combination of CNA and SNV

mutation callers followed by PhyloWGS (SomaticSniper-TITAN-PhyloWGS, TSP; MuTect-TITAN-PhyloWGS, TMP; SomaticSniper-Battenberg-PhyloWGS, BSP; MuTect-Battenberg-PhyloWGS, BMP) were used to perform subclonal reconstruction.

The best tree structure for each input set and the CNAs and SNVs associated with each subclone in that structure were determined by parsing the output JSON files for the consensus tree with the largest log likelihood value. Since subclones in PhyloWGS trees are numbered based on cellular prevalence instead of evolutionary relationship, trees were transformed to consistent representations to allow comparison across cohorts following two rules: 1) trees are left-heavy, 2) all nodes at a particular tree depth must have numbers greater than that of nodes at lower tree depths, with the root node (normal cell population) starting at 0. Further, pruning of nodes was performed following the heuristic that each node must have at least 5 SNVs or 5 CNAs, and at least a cellular prevalence of 2% if it is a subclonal node or a cellular prevalence of 10% if it is the clonal node, creating a subclonal diversity lower bound for each tumour (12). A pruned node was merged with its sibling if a sibling was present, merged with its parent node if a sibling node was not present or eliminated if it was the only direct child of the normal node, with its children becoming the direct children of the normal node. In situations where PhyloWGS produced a polytumour solution for the best consensus tree, the algorithm was re-run up to 12 times with different random number generator seeds after which the final polytumour solution was accepted. The seeds were applied in the following order: 12345, 123456, 1234567, 12345678, 123456789, 246810, 493620, 987240, 1974480, 3948960, 7897920 and 15795840. In the event where PhyloWGS failed to produce a solution due to reconstruction failures or excessive runtime (> 3 months), the sample was excluded from analysis for that pipeline.

Subclonal reconstruction was run on the cohort of 293 tumours with index lesion sequencing for single-region subclonal reconstruction analysis. Single-region analysis provided data for the downstream analysis of inter-tumoural heterogeneity in subclonal reconstruction. For the 10 tumours with multi-region sequencing, each individual sequencing sample was first subjected to single-region subclonal reconstruction as outlined above. Further, multi-region subclonal reconstruction was performed using PhyloWGS, by providing all regions belonging to the same tumour as input for the reconstruction. The procedure was otherwise identical to the single-region reconstructions, including the usage of four algorithm combination pipelines. The single-

and multi-region reconstructions from these 10 tumours provided data for the downstream analysis of intra-tumoural heterogeneity in subclonal reconstruction.

Phylogenetic Tree Classification

We classified the best consensus trees as monoclonal or polyclonal based on the number of subclones they encompass. Trees where only one subclone was detected were termed *monoclonal*. In monoclonal trees, the only subclone detected is then termed the clonal node or the trunk. Trees where more than one subclone was detected were termed *polyclonal*. In polyclonal trees, the subclone that was the only direct child of the normal root node was clonal (or trunk), and all of its descendants were subclonal (or branch). In situations where the normal root node had more than one direct children, the tree was termed *polytumour*, suggestive of multiple independent primary tumours. These were excluded from downstream analysis because the reconstruction of these phylogenies, especially from single sequencing samples, is challenging (12).

Mutation Classification

CNA and SNV mutations were classified as clonal or subclonal based on their node assignment in the best PhyloWGS consensus tree. The mutations that occurred between the root node (normal) and its first child node were classified as clonal (or trunk) mutations, while all others were classified as subclonal (or branch) mutations.

Subclonal Analysis of Copy Number Aberrations

We further filtered the CNAs identified by PhyloWGS using OncoScan data for samples with the data available, removing the TITAN (v1.11.0) or Battenberg (v2.2.6) predicted CNAs that did not overlap any OncoScan CNAs. For samples without OncoScan data, CNAs outputted by PhyloWGS were filtered to retain only those across genomic locations with recurrence of CNAs in previously mentioned samples, with 10 being the established empirical recurrence threshold (12). We used bins of 1Mbp across the genome to characterize the copy number profiles for each sequence sample and created a separate profile for CNAs obtained from each of the four pipelines. Genomic bins were assigned the copy number of overlapping genomic segments, either neutral or aberrated, produced by PhyloWGS. Regions not considered by PhyloWGS due to lack of information were assumed to have the normal copy number of 2. Profiles were also created separately for clonal and subclonal CNAs, where clonal profiles only included

PhyloWGS CNAs that were assigned to the clonal node and subclonal profiles only included subclonal CNAs. We further used previously identified clonal and subclonal subtypes to cluster samples (12). Average clonal and subclonal CNA profiles were generated within each subtype, for clonal or subclonal CNAs, respectively. Samples that were assigned a "branch" (subclonal) subtype in the SomaticSniper-TITAN pipeline (12) but were monoclonal for another pipeline were excluded from the subclonal subtype sets of that pipeline. Samples that were monoclonal in the SomaticSniper-TITAN pipeline and hence with no "branch" (subclonal) subtype but were polyclonal for another pipeline were not considered for any subclonal subtypes in that pipeline. For each pipeline, we used the copy number profiles of all samples with available data to generate average subtype-specific clonal and subclonal CNA profiles of localized prostate cancer, with standard deviation.

We compared the CNA profiles generated from the four pipelines by assessing the difference in clonal and subclonal CNA calls between pipeline pairs. CNA profiles generated by each pipeline were compared to profiles from other pipelines within each 1Mbp bin for each sample. For each sample, a discordance was noted if one pipeline in the pair called a CNA while the other produced a neutral or opposite call (*i.e.* gain *vs.* loss) in the same bin. The number of discordances were summed for each sample and divided by the total number of bins covered by CNAs called by both callers in that sample to assess the proportion of 1Mbp bins where the pipelines were discordant. Clonal and subclonal CNAs were compared separately for each pipeline pair, for all samples with CNA data from both pipelines in the pair. The concordance was calculated as 1 – discordance, and error bars show one standard deviation of the mean proportion of concordance between all samples compared in a pipeline pair.

We identified CNAs that were differentially altered clonally and subclonally. Using 1Mbp bins across the genome, we aggregated the number of samples with and without a CNA overlapping each 1Mbp stretch. Clonal and subclonal CNAs were annotated separately, and only samples with *polyclonal* phylogenies were considered, since they have both clonal and subclonal components. Pearson's χ^2 test was used with multiple testing correction (FDR $\leq$ 0.05) to define the bins that were significantly enriched for clonal or subclonal CNAs. CNAs in these bins were thus considered significantly differentially altered, with a predisposition to occur clonally or subclonally based on their observed frequencies. The analysis was performed separately for

CNAs obtained from the four pipelines. Genes affected by differentially altered CNAs were annotated using RefSeq, and the lists of genes considered to have CNA biases by the four pipelines were compared for overlap.

We performed pathway enrichment analysis on the genes that were identified by all four pipelines as biased to have a CNA clonally or subclonally. Using all default parameters of gprofileR (v0.6.1) in R (v3.2.5) (42), statistically significant pathways were obtained from the data sources Gene Ontology (Biological Process, Molecular Function and Cellular Component), KEGG and Reactome. We discarded pathways that involved > 500 or < 3 genes (43). Cytoscape (v3.4.0) was used to visualize significant pathways (44). Since all genes identified as significantly differentially affected were affected by clonally biased CNAs, we defined these pathways as differentially altered clonally.

Subclonal Analysis of Single Nucleotide Variants

We defined functional SNVs as those that are nonsynonymous, stop-loss, stop-gain or splice-site, based on RefGene annotations. For each gene with a functional SNV mutation in a sample, the cancer cell fraction of the SNV was recorded. In the event that a gene was mutated multiple times in a sequence sample, the SNV with the highest functional priority was used for such annotation.

We compared the four pipelines for their inference of clonal and subclonal SNVs. In each pairwise comparison, for each sample we noted the number of times one pipeline called a functional SNV in a gene while the other did not call a mutation in that gene. The number of discordances was divided by the total number of functional SNVs called by both pipelines as annotated at the gene level in each sample to obtain the proportion of discordance. Concordance was further calculated as 1 – discordance, and the error bars show one standard deviation from the mean of all sample concordances in a pipeline pair. The analysis was performed separately for clonal and subclonal mutations. Instances where one caller called a clonal mutation for a gene while the other caller called a subclonal mutation were not counted as disagreements towards either analysis.

Driver Mutation Analysis

We gathered a list of known prostate cancer driver genes based on previous large sequencing studies (12,19). The known CNA-affected driver genes considered were *MYC*, *TP53*, *NKX3-1*, *RB1*, *CDKN1B*, *CHD1*, *PTEN* and *CDH1*. The known SNV-affected driver genes considered were *ATM*, *MED12*, *FOXA1*, *SPOP* and *TP53*. PhyloWGS inferred CNAs overlapping regions of CNA-affected driver genes and annotated SNVs that occurred in SNV-affected driver genes were defined to be driver CNAs and driver SNVs, respectively. A sample was considered to have a consensus driver mutation if the mutation was called with the same clonality by all four pipelines.

We calculated the proportion of samples harboring driver CNAs separately for each of the four pipelines. Driver CNAs were divided into clonal and subclonal CNAs, and further by gain or loss. We also calculated the proportion of samples with consensus driver CNA mutations.

Driver SNVs and CNAs of each sample were categorized by the number of pipelines they were called in. Since four pipelines were used, in each sample driver SNVs and CNAs could be called in all four pipelines, three pipelines, two pipelines or one pipeline. Proportions of each category were calculated by dividing the number of samples in that category by the sum of samples assigned to all categories for the driver SNV or CNA. The analysis was done separately for clonal and subclonal calls, such that the category of the driver SNVs or CNAs in a sample was defined by the most frequent call of the clonality. For example, if a driver SNV in a sample was called clonal by two pipelines, subclonal by one pipeline and wildtype by the last pipeline, it would be counted in both category two for the clonal analysis and in category one for the subclonal analysis.

Multi-region Subclonal Reconstruction Analysis

For each of the 10 patients with multi-region sequencing data available, we compared the subclonal reconstruction solutions from each single-region with the solutions obtained from subclonal reconstruction using all tumour regions. Further, we compared the reconstructions of each index lesion from only its own single-region reconstruction and its reconstruction as part of the corresponding multi-region reconstruction. More specifically, the number of subclones assigned to the index lesion from the multi-region reconstruction was compared to the number of subclones detected in the index lesion single-region reconstruction.

In addition to number of subclones detected, we compared SNV and CNA clonality assignments between single- and multi-region reconstructions. For all SNVs that were detected in a single-region and its corresponding multi-region reconstruction, we calculated the proportion of SNVs in each of the following categories:

1. Match in multi- and single-region: clonality of SNV is the same in single- and multi-region reconstructions.
2. Clonal in multi-region: SNV was detected in both single- and multi-region reconstructions, but the clonality of the SNV was assigned to be clonal in the multi-region reconstruction, and subclonal in the single-region reconstruction.
3. Subclonal in multi-region: SNV was detected in both single- and multi-region reconstructions, but the clonality of the SNV was assigned to be subclonal in the multi-region reconstruction, and clonal in the single-region reconstruction.
4. Unique in single-region: SNV was only present in the single-region reconstruction.
5. Unique in multi-region: SNV was only present in the multi-region reconstruction.

Similarly, all CNAs that were detected in a single-region reconstruction and its matching multi-region reconstruction were assigned to categories defined in a similar fashion. Additional separation was added for CNAs defined to have the same clonality in the single-region and multi-region reconstructions, distinguishing between clonal and subclonal assignments. The distinction was also made for CNAs unique to single-region or multi-region reconstructions, separating the two categories further into clonal and subclonal as well.

Data Visualization and Reporting

Data was visualized using the R statistical environment (v3.2.5), and performed using the lattice (v0.20-34), latticeExtra (v0.6-28), VennDiagram (v1.6.21) (45) and BPG (v5.3.4) (46) packages. Figures were compiled using Inkscape (v0.91). Standard deviation of the population mean was reported for point estimates for the mean. All statistical tests were two-sided. **Additional File 1** shows the visualization of all phylogenies.

List of abbreviations

CNAs - Copy Number Aberrations

SNVs - Single Nucleotide Variants

TSP - phylogenetic reconstruction pipeline TITAN-SomaticSniper-PhyloWGS

TMP - phylogenetic reconstruction pipeline TITAN-MuTect-PhyloWGS

BSP - phylogenetic reconstruction pipeline Battenberg-SomaticSniper-PhyloWGS

BMP - phylogenetic reconstruction pipeline Battenberg-MuTect-PhyloWGS

Declarations

Ethics approval and consent to participate

All prostate tumour samples involved in this study were obtained with patient informed consent, with institutional Research Ethics Board approval and following International Cancer Genome Consortium guidelines.

Consent for publication

Not applicable.

Availability of data and material

The datasets supporting the conclusions of this article are available as listed below:

WGS Data - Baca *et al*., 2013: dbGaP, phs000447.v1.p1 (26)

WGS Data – Berger *et al*., 2011: dbGaP, phs000330.v1.p1 (27)

WGS Data – CPC-GENE Fraser *et al.*, 2017: EGA, EGAS00001000900; GEO: GSE84043 (19)

WGS Data – The Cancer Genome Atlas Research Network, 2015:

https://portal.gdc.cancer.gov/projects/TCGA-PRAD (28)

WGS Data - Weischenfeldt *et al.*, 2013: EGA, EGAS00001000400 (29)

WGS Data – CPC-GENE Espiritu *et al.*, 2018: EGA, EGAS00001000900 (12)

WGS Data – CPC-GENE Taylor *et al.*, 2017: EGA, EGAS00001001615; EGA, EGAS000010000258 (20)

Variant Data - CPC-GENE Espiritu *et al.*, 2018: EGA, EGAS00001000900 (12)

Data supporting the conclusions of this article is included within the article and its additional files. Additional data is also publicly available online at the EGA: EGAS00001000900

Competing interests

All authors declare that they have no conflicts of interest.

Funding

This study was conducted with the support of Movember funds through Prostate Cancer Canada and with the additional support of the Ontario Institute for Cancer Research, funded by the Government of Ontario. This work was supported by Prostate Cancer Canada and is proudly funded by the Movember Foundation - Grant #RS2014-01 to PCB. PCB was supported by a

Terry Fox Research Institute New Investigator Award and a CIHR New Investigator Award. This project was supported by Genome Canada through a Large-Scale Applied Research Project contract to PCB and Drs. Ryan Morin and Sohrab P. Shah. This work was supported by NSERC Discovery Grants to QDM and PCB. This research is funded by the Canadian Cancer Society (grant #705649) and by a Project Grant from the Canadian Institutes for Health Research. This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-125). This work has been funded by a Fellowship from the Canadian Institutes of Health Research to Vinayak Bhandari. The results described here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

Authors' contributions

Data Analyses: VB, LYL, SMGE

Data Visualization: VB, LYL, AS

Supervised Research: QDM, PCB

Initiated the Project: VB, LYL, QDM, PCB

Wrote the First Draft of the Manuscript: VB, LYL, PCB

Approved the Manuscript: All Authors

Acknowledgements

The authors thank all members of the Boutros lab for helpful suggestions and technical support.

References

1. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017 Apr 26;545(7655):446–51.
2. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non–Small-Cell Lung Cancer. N Engl J Med. 2017 Jun 1;376(22):2109–21.
3. Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JI, et al. Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. Cell. 2018 Apr 19;173(3):581–594.e12.
4. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. Cell. 2018 Apr 19;173(3):595–610.e11.
5. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. Nature. 2015 Apr 1;520(7547):353–7.
6. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. N Engl J Med. 2012 Mar 8;366(10):883–92.
7. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. Elsevier; 2012 May 25;149(5):994–1007.
8. Boutros PC, Fraser M, Harding NJ, de Borja R, Trudel D, Lalonde E, et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. Nat Genet. Nature Publishing Group; 2015 May 25;47(7):736–45.
9. Cooper CS, Eeles R, Wedge DC, Van Loo P, Gundem G, Alexandrov LB, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. Nat Genet. Nature Publishing Group; 2015 Apr 2;47(4):367–72.
10. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell. 2018 Apr 19;173(3):611–623.e17.

11. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet. Nature Publishing Group; 2014 Mar 2;46(3):225–33.
12. Espiritu SMG, Liu LY, Rubanova Y, Bhandari V, Holgersen EM, Szyca LM, et al. The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. Cell. Elsevier; 2018 Apr;0(0).
13. Deshwar AG, Vembu S, Yung CK, Jang G, Stein L, Morris Q. PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. Genome Biol. 2015 Feb 13;16(1):35.
14. Jiao W, Vembu S, Deshwar AG, Stein L, Morris Q. Inferring clonal evolution of tumors from single nucleotide somatic mutations. BMC Bioinformatics. BioMed Central; 2014 Feb 1;15:35.
15. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods. PMC Canada manuscript submission; 2014 Apr;11(4):396–8.
16. Fischer A, Vázquez-García I, Illingworth CJR, Mustonen V. High-Definition Reconstruction of Clonal Composition in Cancer. Cell Rep. 2014 Jun 12;7(5):1740–52.
17. Zare H, Wang J, Hu A, Weber K, Smith J, Nickerson D, et al. Inferring clonal composition from multiple sections of a breast cancer. PLoS Comput Biol. Public Library of Science; 2014 Jul;10(7):e1003703.
18. Oesper L, Mahmoody A, Raphael BJ. THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome Biol. BioMed Central; 2013 Jul 29;14(7):R80.
19. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017 Jan 9;541(7637):359–64.
20. Taylor RA, Fraser M, Livingstone J, Espiritu SMG, Thorne H, Huang V, et al. Germline BRCA2 mutations drive prostate cancers with distinct evolutionary trajectories. Nat Commun. 2017 Jan 9;8:13671.
21. Ha G, Roth A, Khattra J, Ho J, Yap D, Prentice LM, et al. TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data.

Genome Res. 2014 Nov;24(11):1881–93.

22. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics. 2012 Feb 1;28(3):311–7.
23. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013 Mar;31(3):213–9.
24. Sun R, Hu Z, Sottoriva A, Graham TA, Harpak A, Ma Z, et al. Between-region genetic divergence reflects the mode and tempo of tumor evolution. Nat Genet. 2017 Jun 5;49(7):1015–24.
25. Salcedo A, Tarabichi M, Espiritu SMG, Deshwar AG, David M, Wilson NM, et al. Creating Standards for Evaluating Tumour Subclonal Reconstruction. bioRxiv. Cold Spring Harbor Laboratory; 2018 Jul;310425.
26. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated Evolution of Prostate Cancer Genomes. Cell. 2013 Apr 25;153(3):666–77.
27. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. Nature. 2011 Feb 10;470(7333):214–20.
28. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015 Nov 5;163(4):1011–25.
29. Weischenfeldt J, Simon R, Feuerbach L, Schlangen K, Weichenhan D, Minner S, et al. Integrative Genomic Analyses Reveal an Androgen-Driven Somatic Alteration Landscape in Early-Onset Prostate Cancer. Cancer Cell. 2013 Feb 11;23(2):159–70.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.
31. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep 1;20(9):1297–303.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.
33. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating

cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011 Sep 15;27(18):2601–2.

34. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci. 2010 Sep 28;107(39):16910–5.
35. Taghiyar MJ, Rosner J, Grewal D, Grande BM, Aniba R, Grewal J, et al. Kronos: a workflow assembler for genome analytics and informatics. Gigascience. 2017 Jul 1;6(7):1–10.
36. Lai D, Ha G, Shah S. HMMcopy: Copy number prediction with correction for GC and mappability bias for HTS data. 2016.
37. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
38. Ding J, Bashashati A, Roth A, Oloumi A, Tse K, Zeng T, et al. Feature-based classifiers for somatic mutation detection in tumour–normal paired sequencing data. Bioinformatics. 2012 Jan 15;28(2):167–75.
39. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan;29(1):308–11.
40. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015 Jan;43(D1):D805–11.
41. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep 1;38(16):e164–e164.
42. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. Oxford University Press; 2007 Jul;35(Web Server issue):W193-200.
43. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010 Jul 15;466(7304):368–72.
44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003 Nov 1;13(11):2498–504.

45. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. BioMed Central Ltd; 2011;12(1):35.
46. P'ng C, Green J, Chong LC, Waggott D, Prokopec SD, Shamsi M, et al. BPG: Seamless, Automated and Interactive Visualization of Scientific Data. bioRxiv. Cold Spring Harbor Laboratory; 2017 Jun;156067.

Figure Legends

Figure 1 – Reconstruction Workflow and Experimental Design

Raw sequencing data from the tumour and normal samples were aligned against the hg19 build of the human genome using bwa-aln (v0.5.7) and GATK (v2.4.9). Somatic SNVs were called using SomaticSniper (v1.0.2) and MuTect (v1.1.4) and annotated for function. Somatic CNAs were called using TITAN (v1.11.0) and Battenberg (v2.2.6) and filtered. All single-region tumour samples had their subclonal architectures reconstructed using PhyloWGS based on four different algorithm combination pipelines for SNV and CNA calling. For tumours with samples from multiple regions, reconstructions of subclonal architecture were performed by considering samples from all regions using each of the four pipelines. Further downstream analyses compared clonality, CNAs and SNVs for each single-region sample across the pipelines, and between single-region and multi-region reconstructions in the same sample, to elucidate the inter- and intra-tumoural heterogeneity of subclonal reconstruction.

Figure 2 – Clonal and Subclonal CNAs

We assessed the average clonal and subclonal CNA profiles of localized prostate cancers based on four different subclonal reconstruction pipelines. **a**-**b**) Chromosomes are shown along the x-axis and copy number on the y-axis, while each horizontal panel represents the average clonal and subclonal CNA profiles from one pipeline, based on previously identified clonal and subclonal subtypes (12). Clonal CNA subtype average profiles are shown in **a**) and subclonal CNA subtypes are shown in **b**). Red indicates that the average is above the neutral copy number of two while blue indicates that the average is below two. The shaded regions indicate the standard deviation. Characteristic prostate cancer CNAs such as losses in 8p and gains in 8q are seen clonally across callers, prominently in clonal subtypes *B* and *C*. Average subclonal copy number is almost always above two across the genome in subtype *g*. Subclonal CNA subtype profiles show little concordance across callers except for those that use the same CNA caller. **c**) Each marker represents the comparison between a pair of callers. Clonal and subclonal disagreement represents the proportion of CNA calls where one pipeline called a CNA and for the same bin in the same sample the other pipeline called either neutral or the opposite CNA type (*i.e.* gain *vs.* loss). Overall, there is greater agreement for clonal CNA calls (all disagreement ≤ 62%) and calls between algorithms that use the same CNA caller largely agree. TSP, TITAN-

SomaticSniper-PhyloWGS; TMP, TITAN-MuTect-PhyloWGS; BSP, Battenberg-SomaticSniper-PhyloWGS; BMP, Battenberg-MuTect-PhyloWGS.

Figure 3 – Differentially Altered CNAs

CNAs that are enriched to occur clonally or subclonally based on subclonal reconstruction from different pipelines. **a**) From subclonal reconstructions based on SomaticSniper and Battenberg, CNAs in 295 genes are enriched to occur clonally and none were biased to occur subclonally. CNAs in all 270 of these genes were also considered to be enriched clonally based on reconstructions from the other three pipelines. **b**) A pathway enrichment map based on the 270 genes that are mutually considered to be differentially altered clonally. Genes that are preferentially altered by CNAs early during tumour development are significantly (FDR < 0.05) associated with *TP53*-based regulation of death receptors, *TRAIL* signaling, and natural killer cell mediated cytotoxicity.

Figure 4 – Clonal and Subclonal SNVs

We assessed the clonal and subclonal SNV landscape of localized prostate cancers using four different pipelines. **a**) The total number of clonal and subclonal SNVs based on four pipelines. Each stacked bar represents one tumour and tumours are ordered based on the total number of SNVs called by the SomaticSniper-TITAN pipeline. Pipelines that use MuTect detected vastly more subclonal SNVs compared to pipelines that use SomaticSniper. **b**) The proportion of disagreement in SNV calls, where one pipeline called a SNV while the other called wildtype, were assessed for clonal and subclonal SNVs. Each marker represents a pipeline pair that is compared, and the x- and y- axis show proportion of subclonal and clonal disagreements, respectively. Clonal SNVs generally show strong agreement between different pipelines while subclonal SNVs show less agreement between pipelines that use different SNV callers. TSP, TITAN-SomaticSniper-PhyloWGS; TMP, TITAN-MuTect-PhyloWGS; BSP, Battenberg-SomaticSniper-PhyloWGS; BMP, Battenberg-MuTect-PhyloWGS.

Figure 5 – Driver CNAs and SNVs in Localized Prostate Cancer

The clonal and subclonal distribution of localized prostate cancer driver mutations based on four pipelines and a consensus. SNVs (**a**) and CNAs (**b**) in known prostate cancer driver genes were typically detected clonally. Each column in **a**) represent a tumour. The consensus for clonal

driver mutation calls (*i.e.,* cases where all four pipelines called a clonal mutation in a driver gene) was high while there was less agreement for subclonal driver mutations.

Figure 6 – Comparing Subclonal Architecture between Pipelines

We assessed the subclonal architecture of localized prostate cancers using four pipelines. **a-b**) The proportion of samples identified to be monoclonal or polyclonal by each of the four pipelines (**a**) and the breakdown of samples into tree structures (**b**). The SomaticSniper-Battenberg pipeline called drastically more monoclonal phylogenies than the other pipelines. Pipelines with MuTect as the SNV caller reconstructed more complex phylogenies with up to five subclones.

Figure 7 - Subclonal Reconstruction from Single- and Multi-Region Samples

Subclonal reconstruction concordance between solutions based on a single-region or multiple regions from the same tumour. **a**) Clonality of SNVs identified in both single-region and multi-region reconstructions. Variants were compared at position level. Each single-region reconstruction is compared to the multi-region reconstruction of the same tumour. The SomaticSniper-Battenberg pipeline achieved the highest proportion of clonality assignment concordance. In the mismatches between single-region and multi-region reconstruction, SNVs tend to be classified to be subclonal in multi-region reconstructions while being classified as clonal in single-region reconstructions. **b**) Comparison of the number of subclones in the reconstructed phylogenies for each tumour. Multi-region reconstructions typically result in phylogenies with higher numbers of subclones compared to reconstruction from single-regions. Reconstructions of only the single index lesion typically capture half of the complexity that is uncovered by the multi-region reconstructions. Zero subclones indicate a failed reconstruction.

Supplementary Figures

Supplementary Figure 1 – Comparing the Divergence of Callers for Driver CNAs and SNVs

a-b) The four pipelines showed broad agreement in calling driver CNAs and SNVs. Of all the samples where any pipeline called a clonal CNA in *MYC*, all four pipelines called a clonal CNA in 74% of cases. This is not surprising considering the hallmark gain of chromosome 8q in prostate cancer. SNV calls in *MED12* were an outlier, where the four pipelines never called a clonal SNV in the same sample.

Supplementary Figure 2 – Comparing Clonality between Pipelines

a-f) Between pipeline pairs, the concordance of clonality for shared samples are compared. The highest disagreement is observed for pipeline pairs with the same SNV caller but different CNA callers.

Supplementary Figure 3 – Discrepancies between Single-Region and Multi-Region CNA Assignments

The disagreement in clonal and subclonal CNA assignments between single- and multi-region subclonal reconstructions. CNAs are compared by 1Mbp genomic bins between single-region reconstructions and their corresponding multi-region reconstructions. The SomaticSniper-Battenberg pipeline achieved the lowest proportion of clonality assignment concordance and mismatches were typically due to unique subclonal CNAs assigned to the multi-region reconstruction and unique clonal CNAs assigned to the single-region reconstructions.

Supplementary Figure 4 – Index Lesion Clonality in Single-Region and Multi-Region Reconstructions

We assessed the number of subclones detected specifically in the index lesion in the single-region and multi-region reconstructions. Single-region reconstructions indicate the number of subclones detected in the index lesion using only the sample from the index lesion. Multi-region reconstructions indicate the number of subclones detected in the index lesion via multi-region reconstruction. Overall, multi-region reconstructions detect more subclones within the index lesion sample compared to the single-region reconstruction of the index lesion.

Supplementary Tables

Supplementary Table 1 – Single-Region Subclonal Reconstruction Data

Subclonal reconstruction results for 293 tumours with single-region sequencing. Information is displayed separately for all four pipelines used for analysis. Phylogenetic tree variables include tree type, tree lineage type, number of subclones, tumour cellular prevalence, tree root width, tree depth, clonality and Shannon index score for subclonal diversity. Mutational variables include proportion of clonal SNVs, proportion of clonal CNAs, number of SNVs, number of CNAs, overall percent genome altered (PGA) and PGA due to clonal CNAs. Additional information includes the random number generator seed used for PhyloWGS reconstruction, coverage of sequencing sample, cellular prevalence obtained from the CNA caller used and CNA subtype assignment.

Supplementary Table 2 – CNA Data

CNA data for 293 patients in 1 Mbp genomic bins. Each number represents the copy number of the CNA overlapping the genomic bin. Positive indicates gain while negative indicates loss. Tables are separated for clonal and subclonal CNAs and for each of the four pipelines.

Supplementary Table 3 – Differentially Altered Genes

Genes tested for bias towards clonal or subclonal CNAs. All genes are ordered by genomic position and the number of samples affected by clonal and subclonal CNAs are recorded. Pearson's χ^2 test was used to test for bias, with p-values adjusted using FDR. Information is presented separately for each pipeline.

Supplementary Table 4 – Pathway analysis

Pathways significantly (FDR < 0.05) associated with the 270 genes where CNAs are biased to occur clonally based on all four pipelines.

Supplementary Table 5 – SNV Data

SNV-affected gene by patient matrix for 293 patients. Each number represents the cancer cell fraction of the subclone to which the mutation was assigned. SNVs with a cancer cell fraction of 1 are clonal and all others with an absolute value of cancer cell fraction below 1 are subclonal. Information is presented separately for the four pipelines.

Supplementary Table 6 - Multi-Region Subclonal Reconstruction Data

Subclonal reconstruction results for 10 patients with multi-region sequencing. Results are included for reconstructions of each individual region and results from the multi-region reconstructions. Information is displayed separately for all four pipelines. Phylogenetic tree variables include tree type, tree lineage type, number of subclones, tumour cellular prevalence, tree root width and clonality. Mutational variables include proportion of clonal SNVs, proportion of clonal CNAs, number of SNVs, number of CNAs, overall percent genome altered (PGA) and PGA due to clonal CNAs. Additional information includes the random number generator seed used for PhyloWGS reconstruction.

Additional Files

Additional File 1 – Per-Sample Subclonal Architecture Reports

Subclonal reconstruction solutions from four pipelines for 293 samples with single-region sequencing, followed by 10 samples with multi-region sequencing. The first page contains a legend explaining the components of single- and multi-region subclonal reconstruction figures. Subsequent pages have details for all single-region samples followed by details for single- and multi-region reconstructions for all multi-region samples. Reconstructions from all four pipelines are presented.

Figure 1

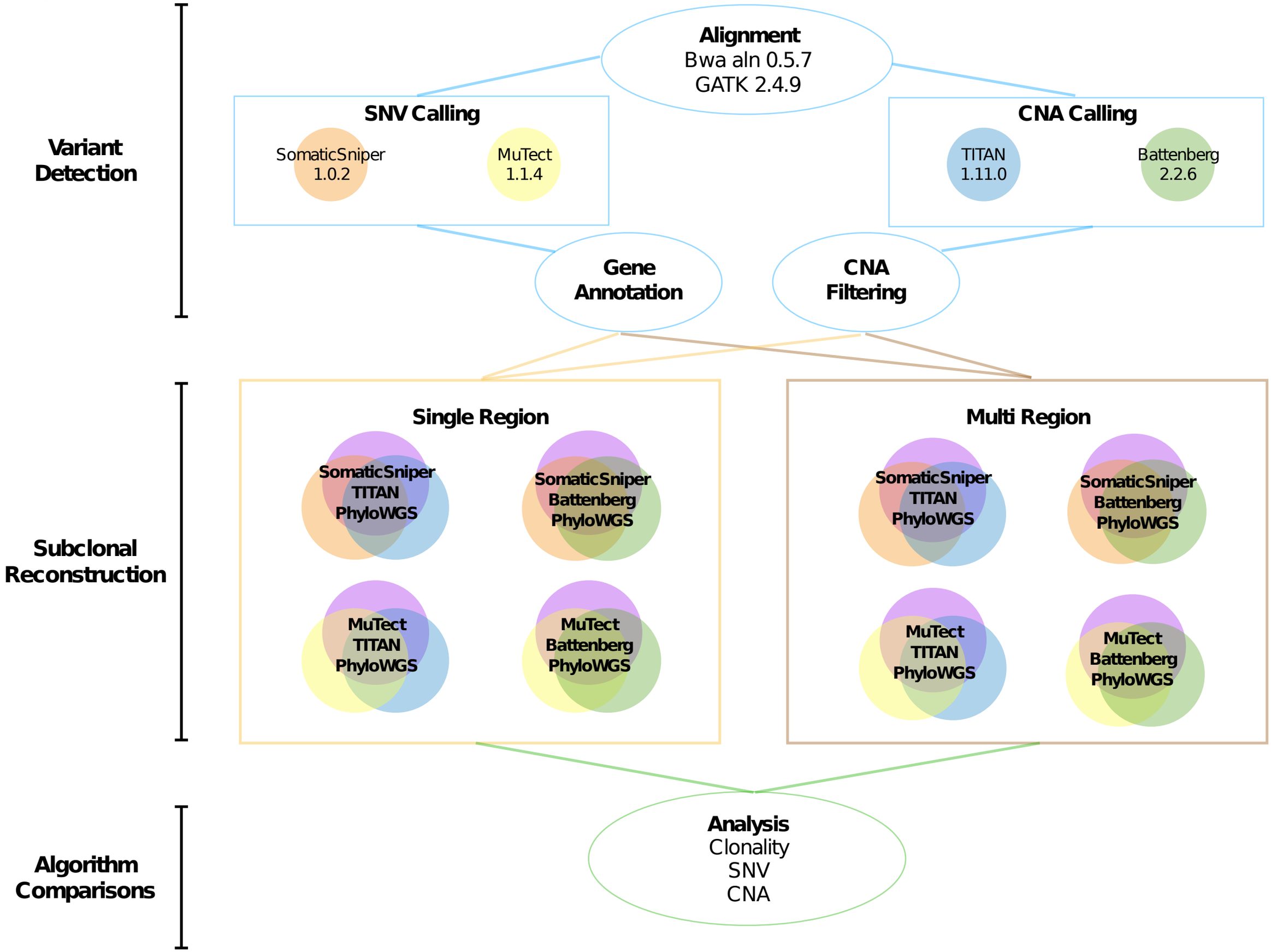

Variant Detection
Alignment
Bwa aln 0.5.7
GATK 2.4.9
SNV Calling
SomaticSniper 1.0.2
MuTect 1.1.4
CNA Calling
TITAN 1.11.0
Battenberg 2.2.6
Gene Annotation
CNA Filtering
Subclonal Reconstruction
Single Region
SomaticSniper TITAN PhyloWGS
SomaticSniper Battenberg PhyloWGS
MuTect TITAN PhyloWGS
MuTect Battenberg PhyloWGS
Multi Region
SomaticSniper TITAN PhyloWGS
SomaticSniper Battenberg PhyloWGS
MuTect TITAN PhyloWGS
MuTect Battenberg PhyloWGS
Algorithm Comparisons
Analysis
Clonality
SNV
CNA

Figure 2

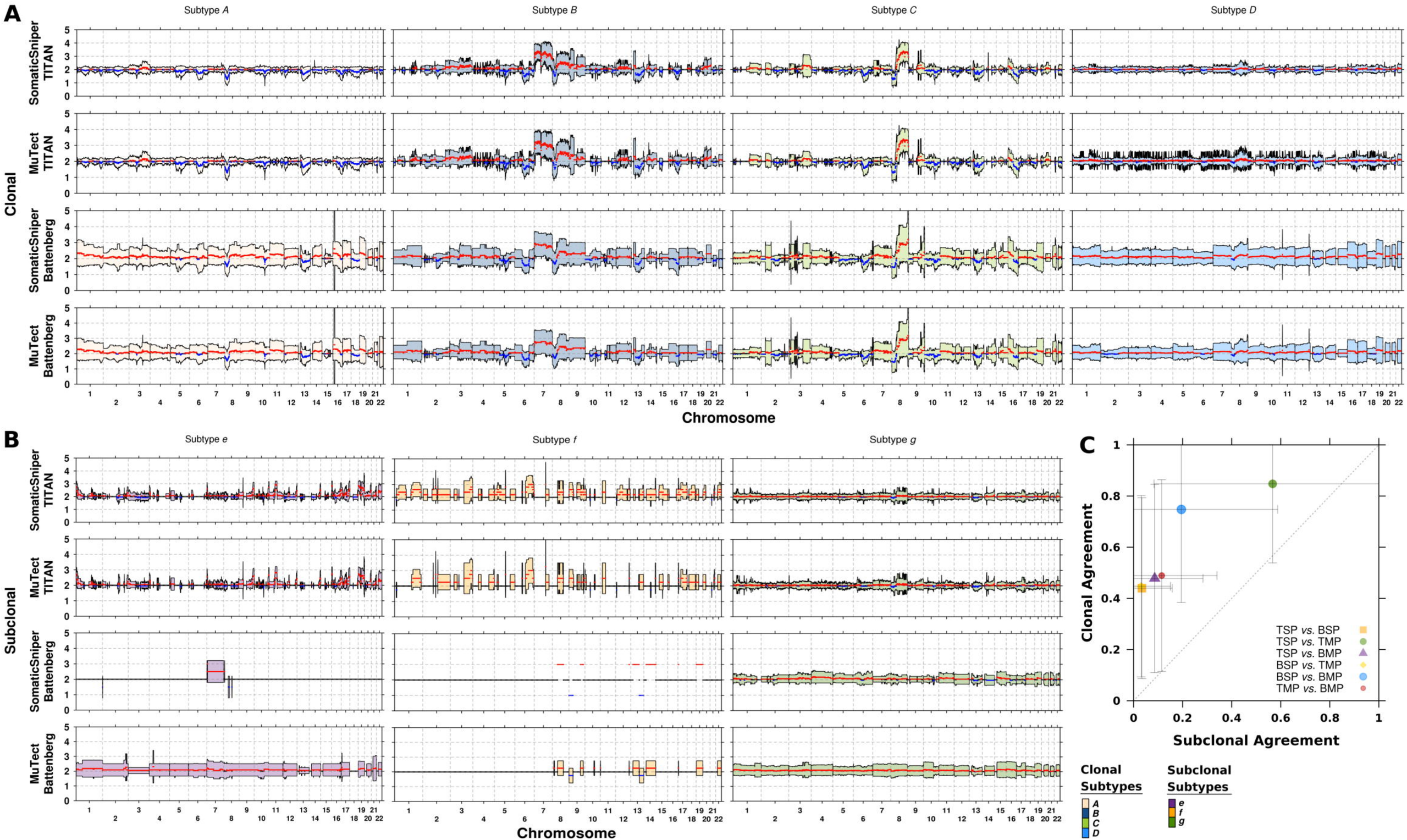
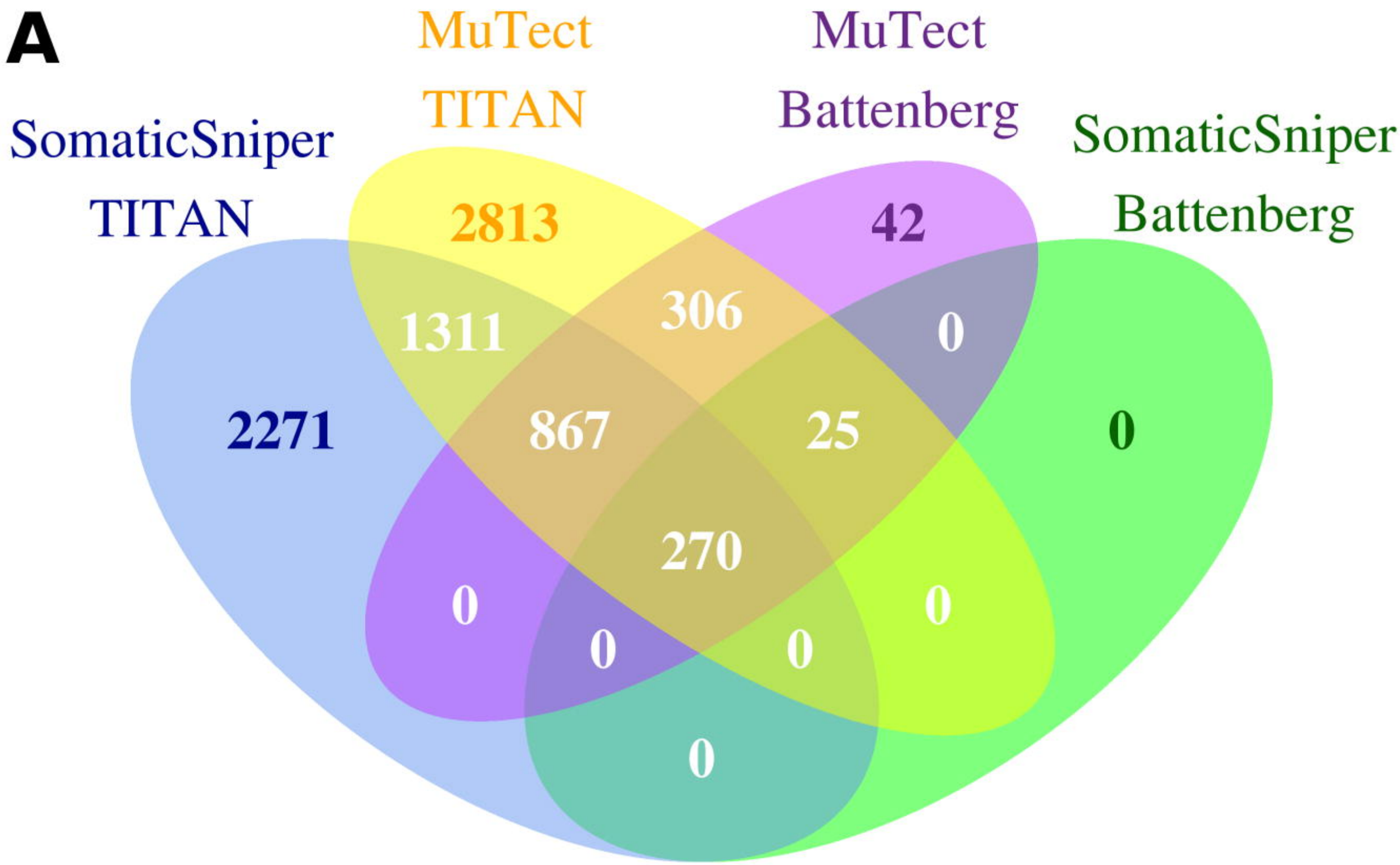

A
Subtype A
Subtype B
Subtype C
Subtype D
Clonal
SomaticSniper TITAN
MuTect TITAN
SomaticSniper Battenberg
MuTect Battenberg
Chromosome
B
Subtype e
Subtype f
Subtype g
Subclonal
Chromosome
C
Clonal Agreement
Subclonal Agreement
TSP vs. BSP
TSP vs. TMP
TSP vs. BMP
BSP vs. TMP
BSP vs. BMP
TMP vs. BMP
Clonal Subtypes
A
B
C
D
Subclonal Subtypes
e
f
g

Figure 3

A


SomaticSniper
TITAN
MuTect
TITAN
MuTect
Battenberg
SomaticSniper
Battenberg
2813
42
1311
306
0
2271
867
25
0
270
0
0
0
0
0

B

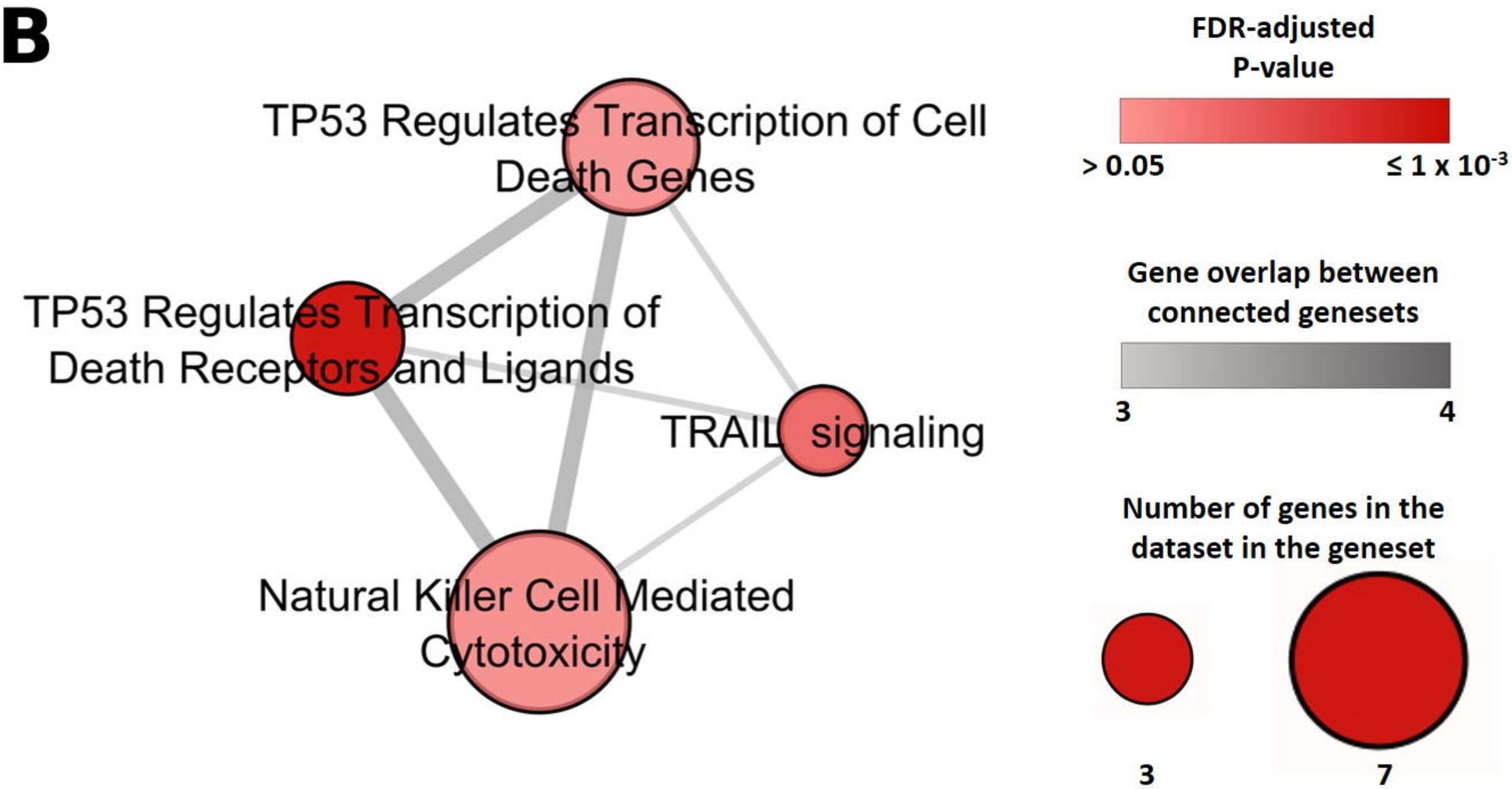

TP53 Regulates Transcription of Cell Death Genes
TP53 Regulates Transcription of Death Receptors and Ligands
TRAIL signaling
Natural Killer Cell Mediated Cytotoxicity
FDR-adjusted P-value
> 0.05
≤ 1 x 10-3
Gene overlap between connected genesets
3
4
Number of genes in the dataset in the geneset
3
7

Figure 4

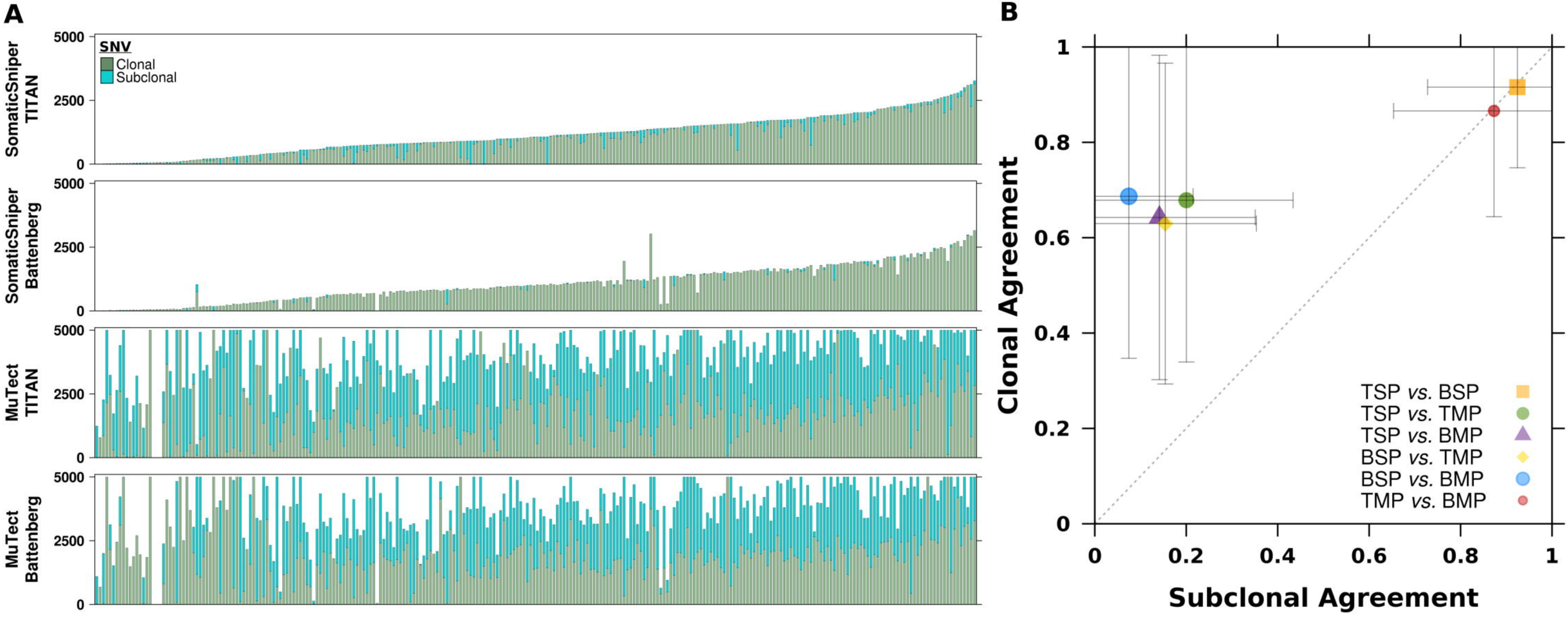

A
SNV
Clonal
Subclonal
SomaticSniper TITAN
SomaticSniper Battenberg
MuTect TITAN
MuTect Battenberg
5000
2500
0
B
Clonal Agreement
Subclonal Agreement
0
0.2
0.4
0.6
0.8
1
TSP vs. BSP
TSP vs. TMP
TSP vs. BMP
BSP vs. TMP
BSP vs. BMP
TMP vs. BMP

Figure 5

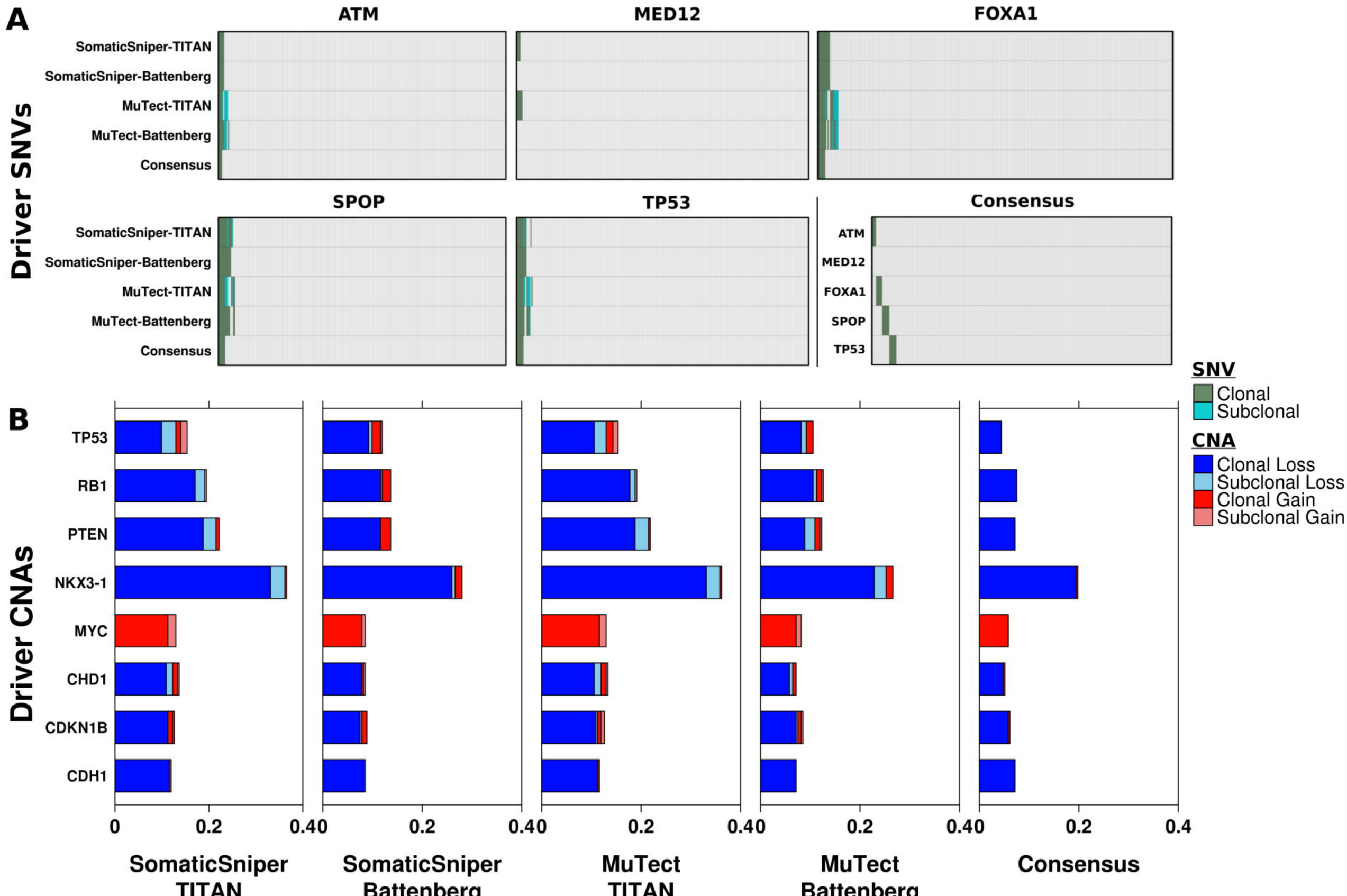

A
Driver SNVs
ATM
MED12
FOXA1
SPOP
TP53
Consensus
SomaticSniper-TITAN
SomaticSniper-Battenberg
MuTect-TITAN
MuTect-Battenberg
Consensus
ATM
MED12
FOXA1
SPOP
TP53
B
Driver CNAs
TP53
RB1
PTEN
NKX3-1
MYC
CHD1
CDKN1B
CDH1
0
0.2
0.40
0.4
SomaticSniper TITAN
SomaticSniper Battenberg
MuTect TITAN
MuTect Battenberg
Consensus
SNV
Clonal
Subclonal
CNA
Clonal Loss
Subclonal Loss
Clonal Gain
Subclonal Gain

Figure 6

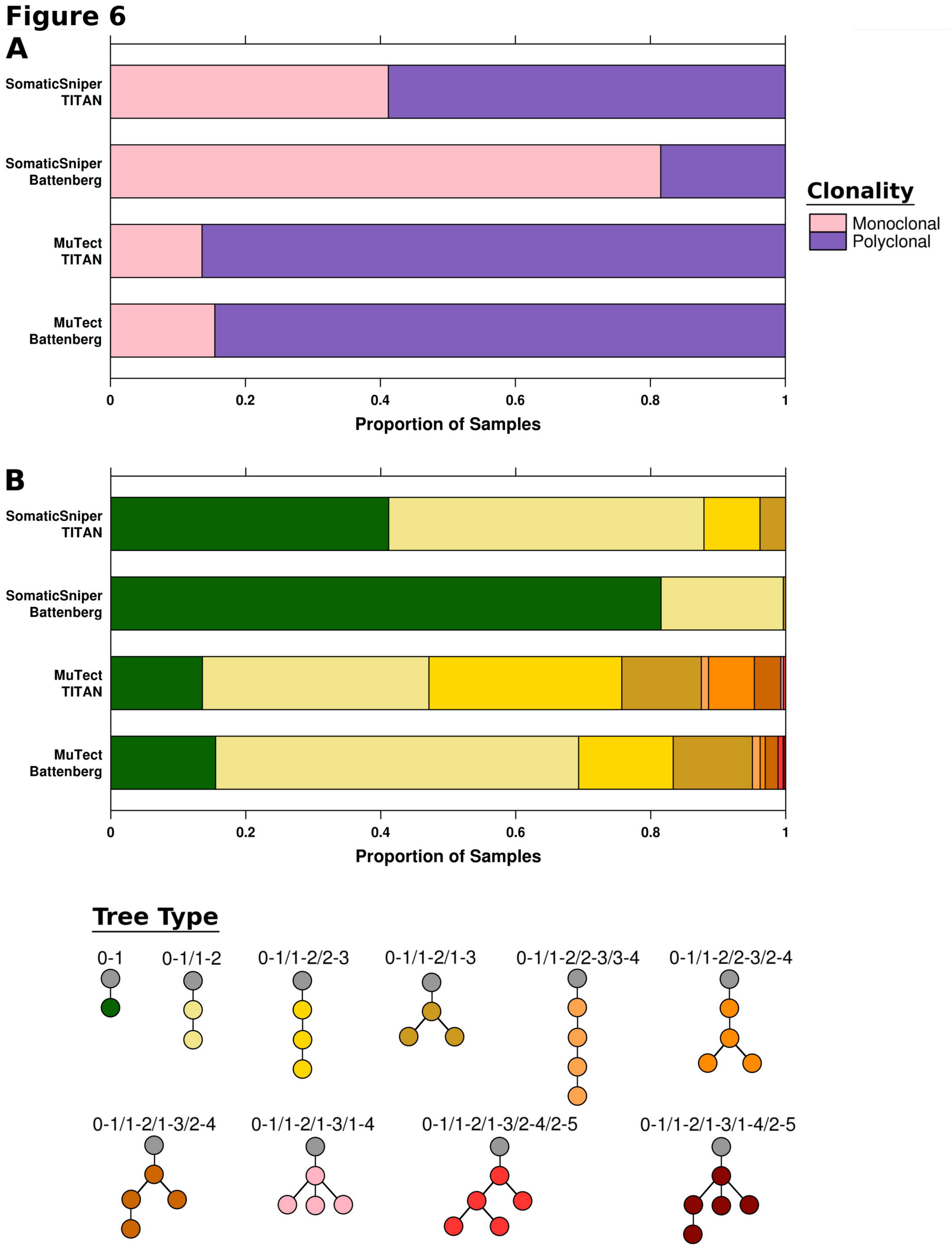
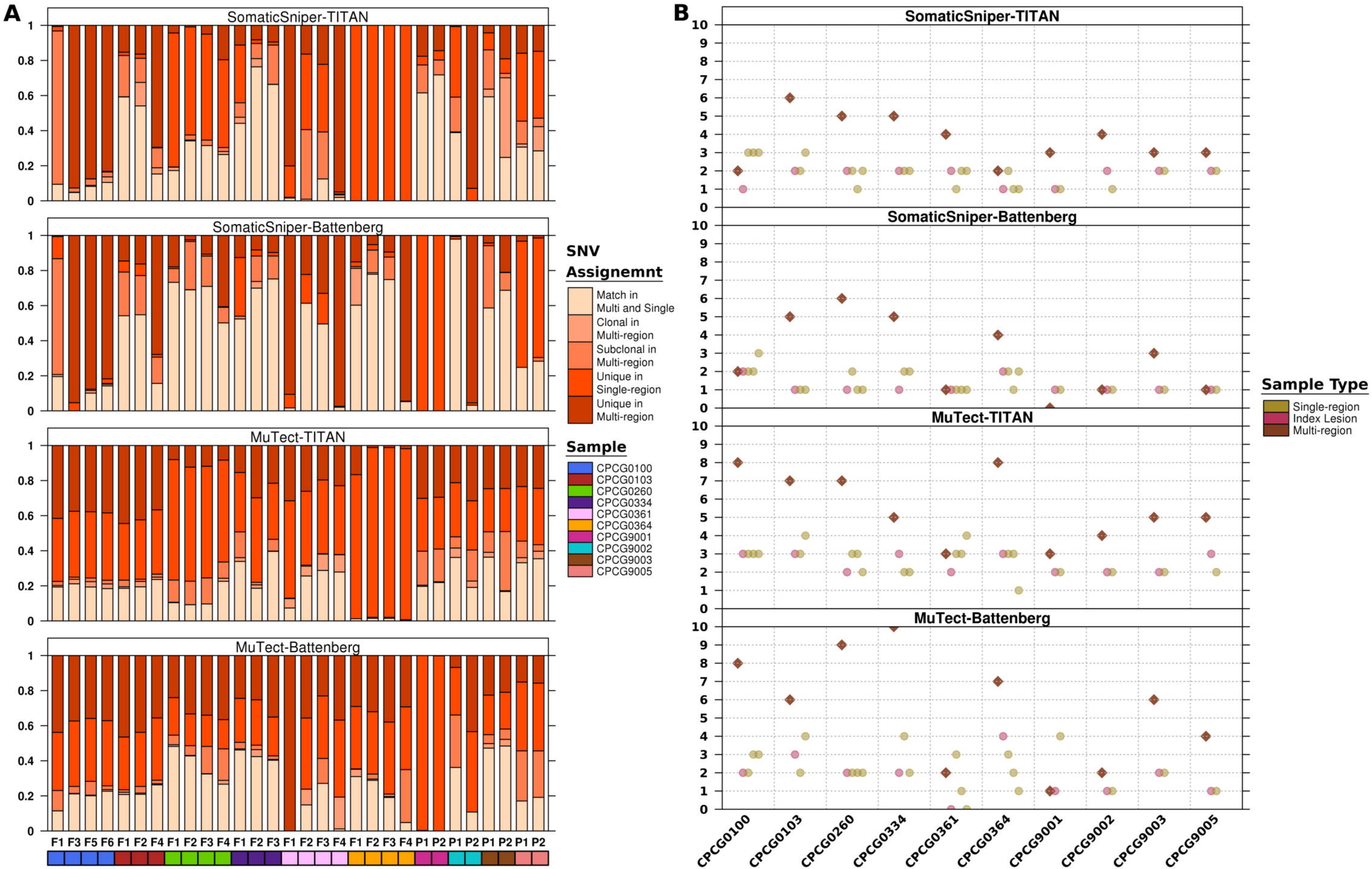

A
SomaticSniper TITAN
SomaticSniper Battenberg
MuTect TITAN
MuTect Battenberg
0
0.2
0.4
0.6
0.8
1
Proportion of Samples
Clonality
Monoclonal
Polyclonal
B
SomaticSniper TITAN
SomaticSniper Battenberg
MuTect TITAN
MuTect Battenberg
0
0.2
0.4
0.6
0.8
1
Proportion of Samples
Tree Type
0-1
0-1/1-2
0-1/1-2/2-3
0-1/1-2/1-3
0-1/1-2/2-3/3-4
0-1/1-2/2-3/2-4
0-1/1-2/1-3/2-4
0-1/1-2/1-3/1-4
0-1/1-2/1-3/2-4/2-5
0-1/1-2/1-3/1-4/2-5

Figure 7


A
SomaticSniper-TITAN
SomaticSniper-Battenberg
MuTect-TITAN
MuTect-Battenberg
SNV Assignemnt
Match in Multi and Single
Clonal in Multi-region
Subclonal in Multi-region
Unique in Single-region
Unique in Multi-region
Sample
CPCG0100
CPCG0103
CPCG0260
CPCG0334
CPCG0361
CPCG0364
CPCG9001
CPCG9002
CPCG9003
CPCG9005
F1 F3 F5 F6 F1 F2 F4 F1 F2 F3 F4 F1 F2 F3 F1 F2 F3 F4 F1 F2 F3 F4 P1 P2 P1 P2 P1 P2 P1 P2
B
SomaticSniper-TITAN
SomaticSniper-Battenberg
MuTect-TITAN
MuTect-Battenberg
Sample Type
Single-region
Index Lesion
Multi-region
CPCG0100
CPCG0103
CPCG0260
CPCG0334
CPCG0361
CPCG0364
CPCG9001
CPCG9002
CPCG9003
CPCG9005