Quantifying the Influence of Mutation Detection on Tumour Subclonal Reconstruction

Vinayak Bhandari[1,2,*], Lydia Y. Liu[1,2,6*], Adriana Salcedo[1,2], Shadrielle M. G. Espiritu[1], Quaid D. Morris[3,4,5,6], Paul C. Boutros[1,2,6,7,8,9,10,11,12]

[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

[8] Department of Human Genetics, University of California, Los Angeles

[9] Department of Urology, University of California, Los Angeles

[10] Institute for Precision Health, University of California, Los Angeles

[11] Jonsson Comprehensive Cancer Centre, University of California, Los Angeles

[12] Lead Contact

*These authors contributed equally to this work

Address for correspondence:

Dr. Paul C. Boutros

12-109 CHS

10833 Le Conte Avenue

Los Angeles, California

90095

Email: pboutros@mednet.ucla.edu

Phone: 310-794-7160

Abstract

Recent studies have used whole-genome sequencing to estimate subclonal populations in tumours and have linked this heterogeneity to clinical outcomes. Many algorithms have been developed to perform subclonal reconstruction but their variability and consistency is largely unknown. To quantify this effect, we evaluated six pipelines for reconstructing the evolutionary histories of 293 localized prostate cancers from single samples and 10 from multi-region sampling to probe the heterogeneity of subclonal reconstruction. We identified extensive variance across pipelines in the subclonal architectures they predict and in their assignments of CNAs and SNVs to different parts of the evolutionary tree. Individual pipelines showed consistent types of bias, with those using SomaticSniper and Battenberg preferentially predicting homogenous cancer cell populations while those using MuTect tending to predict multiple populations of cancer cells. Subclonal reconstructions using multi-region sampling showed that single-sample reconstructions systematically underestimated intra-tumoural heterogeneity, detecting on average fewer than half of the cancer cell populations identified by multi-region sequencing. These biases suggest caution in interpreting the specific architectures and subclonal variants identified, particularly from single-sample reconstructions.

Background

Understanding tumour heterogeneity and subclonal architecture is important for the elucidation of the mutational and evolutionary processes underlying tumorigenesis and treatment resistance [1–4]. Most studies of tumour heterogeneity have focused on small 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. Notably, these studies have been able to reconstruct sample trees that illustrate the spatial relationships between multiple primary and metastatic lesions using shared and private mutations [6, 11]. Many have further delved into intra-tumoural heterogeneity and constructed clone trees that demonstrate the phylogenetic relationship between cancer cell populations that are shared or unique between lesions [5, 7, 9]. The later analyses not only provide insight to the convergent and branching evolution of cancer, but also characterize cancer cell migration and highlight the subclonal complexity within individual lesions.

Several groups have applied 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 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) [12].

A variety of algorithms have been developed to reconstruct the subclonal architecture of cancers [13–18]. These algorithms broadly attempt to infer the phylogenetic relationship between cancer cell populations based on cancer cell fractions (the fraction of cancer cells in which each variant is present) and several use Markov Chain Monte Carlo methods to cluster mutations, estimate the number of cancer cell populations and infer their relationships [14–16, 19]. However, there has not been a systematic comparison of the features and consistencies 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 multi-region reconstructions, raising questions about the magnitude of underestimation present in large-cohort studies.

To address these gaps in the field, we evaluated six pipelines comprised of well-established subclonal copy number alteration (CNA) and single nucleotide variant (SNV) detection tools, linked to three independent subclonal reconstruction algorithms. The pipelines were applied to a set of 293 high-depth tumour-normal pairs [12, 20] and 10 tumours with multi-region sequencing [8, 21]. 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 six pipelines (**Figure 1**, **Supplementary Table 1**). Each patient had whole-genome sequencing of a region from the index lesion (see methods) that was macro-dissected to > 70% tumour cellularity (mean coverage ± standard deviation: 63.9 ± 16.7) and of matched blood reference tissue (mean coverage ± standard deviation: 41.2 ± 9.0), as reported previously [12]. To quantify the roles of subclonal variant detection we called CNAs using Battenberg and TITAN [7, 22] and SNVs using SomaticSniper and MuTect [23, 24]. We then used the CNA and SNV calls in different combinations as inputs for three widely used subclonal reconstruction algorithms: PhyloWGS [14], DPClust [15] and PyClone [16]. All subclonal reconstruction tools were run with default parameters and the same post-processing heuristics to minimize bias (see Methods). We further quantified the heterogeneity that arises in subclonal reconstruction from spatially sampling the same tumour, examining 10 tumours with multi-region sequencing [8, 21]. For multi-region sequencing samples, subclonal reconstruction was performed both with all regions together and with each region individually using PhyloWGS, allowing estimation of the variability of single-region reconstruction.

Subclonal reconstructions with PyClone were successfully completed for all samples for which reconstructions were attempted. Success rates were also high for reconstructions using DPClust (failure rate of 3.1%), except for instances where excessive computational memory was required (> 250 GB), samples for which mutation calling was unsuccessful and one instance for which the only tumour population found had greater than 100% cancer cell fraction, likely due to upstream errors in purity estimation. Reconstructions with PhyloWGS were unsuccessful in a number of instances due to excessive runtime (> 3 months) or failure to produce non-*polytumour* (*i.e.*, multiple independent primary tumours, see methods) 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 (below 5 for both CNAs and SNVs, see methods) for all cancer cell populations detected. The rate of reconstruction failure varied across the pipelines using PhyloWGS for single-region reconstructions (SomaticSniper-TITAN: 1.4%,

SomaticSniper-Battenberg: 2.0%, MuTect-TITAN: 4.4%, MuTect-Battenberg: 9.6%, mean failure rate ± standard deviation: 4.4% ± 3.2%). Multi-region reconstructions with PhyloWGS failed in one instance in the SomaticSniper-Battenberg-PhyloWGS pipeline due to a lack of shared SNVs and CNAs between samples from the same tumour, a requirement for PhyloWGS multi-region reconstructions. Overall, the failure rate of 9.6% for MuTect-Battenberg-PhyloWGS 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 using PhyloWGS for reconstructions of the 293 samples with single-region sequencing (**Figure 2A-B, Supplementary Table 2**). Previous work within this cohort using the SomaticSniper-TITAN-PhyloWGS pipeline identified four subtypes of patients with distinct clonal CNA profiles and three subtypes with distinct subclonal CNA profiles [12] based on PhyloWGS processed CNAs, 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 all 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, possibly reflecting a CNA caller bias. (**Figure 2B**).

We next assessed the concordance of these four PhyloWGS-based pipelines in their identification of clonal and subclonal CNAs. We calculated the proportion of 1 Mbp genomic bins 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 sample in the same bin. 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-PhyloWGS and MuTect-TITAN-PhyloWGS pipelines as compared to 75% of subclonal calls. This modest disagreement was expected because even in pipelines with the same CNA caller (and therefore the same input to

PhyloWGS), the reconstructed clonality of CNA segments can differ due to the allelic frequencies of the SNVs detected in the CNA region, and this can differ across SNV callers. By contrast, when the CNA caller was changed, 45% of clonal CNA calls were concordant between MuTect-Battenberg-PhyloWGS and MuTect-TITAN-PhyloWGS pipelines as compared to 22% of subclonal CNA calls. Thus, we observe diversity in clonal CNA profiles across pipelines, but particularly observe extensive diversity in subclonal CNA profiles across pipelines using different CNA callers.

Next, we evaluated how estimates of the timing of specific CNAs as early (*i.e.*, clonal) or late (*i.e.*, subclonal) were affected by the subclonal reconstruction pipeline. Within each pipeline we determined whether each 1 Mbp genomic bin had different proportions of gains and losses clonally and subclonally (FDR < 0.05, Pearson's χ^2 Test, trunk: loss, neutral, gain *vs.* branch: loss, neutral, gain). The number of genes in regions identified with CNAs occurring statistically more frequently early or late differed dramatically across PhyloWGS-linked pipelines (MuTect-TITAN: 5,592; SomaticSniper-TITAN: 4,719; MuTect-Battenberg: 1,510; SomaticSniper-Battenberg: 295). A consensus set of 270 genes showed 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 were preferentially altered early in tumour evolution, 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 SNV detection and classification. All four PhyloWGS-linked 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, likely detecting more false positives and fewer false negatives.

For each pair of PhyloWGS-linked 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 (mean concordance ± standard deviation: 76.9 ± 1.0%) but subclonal SNV calls were much more discordant (mean concordance ± standard deviation: 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 mutations in five genes known to be driven by recurrent somatic SNVs and eight known to be driven by recurrent somatic CNAs in prostate cancer [12, 20]. 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 (*i.e.*, in the clonal population) with 89.7 ± 11.4 % (mean across all pipelines ± standard deviation) of all driver alterations detected clonally across the PhyloWGS-linked 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 divergent was *MED12*, where there was disagreement across pipelines with the same SNV callers. This highlights how the choice of algorithm can influence downstream subclonal reconstructions: since *MED12* is located on the X chromosome and Battenberg does not generate copy number calls for regions of uncertainty and the sex chromosomes, its mutations were disregarded during subclonal reconstruction because PhyloWGS only considers SNVs with overlapping copy number calls (neutral or aberrated). The

same explanation applies to the disagreement in SNV numbers in pipelines with the same SNV callers above.

Subclonal Architecture Reconstructions from Single Samples

Given these significant differences in the attribution of individual mutations to different stages of tumour evolution, we sought to understand whether the underlying subclonal architectures were being predicted consistently across pipelines. We therefore first examined the number of subclones detected in each sample across the pipelines using PhyloWGS (**Figure 6A**). Reconstructions with SomaticSniper-Battenberg-PhyloWGS predominantly detected monoclonal tumours: only one population of cancer cells was detected in 81.5% of the samples. By contrast, the two pipelines using MuTect-based SNV calls identified <16% of samples as monoclonal. The SomaticSniper-TITAN-PhyloWGS 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-PhyloWGS and MuTect-Battenberg-PhyloWGS pipelines detecting many bi-clonal samples, and the MuTect-TITAN-PhyloWGS pipeline predicting more complex tree structures.

We directly compared the subclonal architectures predicted for each sample across the four PhyloWGS-linked pipelines. There was considerable discordance in the predicted subclonal architectures. Consistent with the large number of monoclonal structures predicted by SomaticSniper-Battenberg-PhyloWGS, their predictions poorly agreed with those from other pipelines (**Supplementary Figure 2A-C**). Predictions of subclonal architecture between pipelines using MuTect for SNV calls generally agreed in overall clonality (*i.e.*, monoclonal *vs* polyclonal), with 226/265 tumours matching (**Supplementary Figure 2D)**. However, the specific phylogeny of the evolutionary tree matched for only 84/226 due to the diversity of structures observed in polyclonal reconstructions (17/226 were monoclonal samples). The concordance was also low between pipelines using MuTect to call SNVs and the SomaticSniper-TITAN-PhyloWGS pipeline, where pipelines using MuTect called substantially more samples as polyclonal (**Supplementary Figure 2E-F**). Thus differences in mutation calling substantially affect downstream subclonal reconstructions.

Considering the importance of clonality for predicting prognosis in prostate cancer [12], we evaluated clonality predictions using two other pipelines where reconstructions were performed

using DPClust or PyClone. Reconstructions using DPClust showed strong agreement with reconstructions linked to PhyloWGS with inputs from the same CNA and SNV callers (Battenberg and MuTect, respectively): both algorithms predicted the same number of subclones for 62.2% of samples and predictions were within one subclone for 97.3% of samples (**Figure 6C**). Predictions of clonality (*i.e.*, monoclonal or polyclonal) were also in agreement for 94.3% of samples. Clonality calls from PyClone showed lower agreement with PhyloWGS while using inputs from the same CNA and SNV callers (TITAN and SomaticSniper, respectively), with both algorithms predicting the same number of subclones for only 45.7% of samples (**Figure 6D**). This was driven by the large number of monoclonal calls from PyClone, with predictions of subclone number differing by only one for 92.4% of samples. These data suggest that reconstructions from PhyloWGS and DPClust converge towards similar evolutionary histories, while those from PyClone tend to estimate fewer cancer cell populations, at least in single-sample reconstructions.

Given these differences, we also evaluated subclonal architectures predicted by PhyloWGS-linked pipelines using the union and intersection of mutation calls as inputs. As expected, MuTect had substantially more unique SNV calls compared to SomaticSniper ($p < 2.2 \times 10^{-16}$, Mann-Whitney U-test, $\text{median}_{\text{Unique calls, MuTect}} = 5{,}330$, $\text{median}_{\text{Unique calls, SomaticSniper}} = 627$, **Supplementary Figure 3A**). Unique raw CNA calls by TITAN and Battenberg were also substantially imbalanced, with a median of 50.2% and 1.2% of the covered genome having unique CNA calls across samples, respectively ($p < 2.2 \times 10^{-16}$ Mann-Whitney U-test; **Supplementary Figure 3B**). Compared to the PhyloWGS-linked pipeline with inputs from SomaticSniper and TITAN, the pipeline using the union of CNA calls and union of SNV calls linked to PhyloWGS predicted many more polyclonal trees (146/249 *vs.* 231/249, respectively; **Supplementary Figure 3C**) with clonality agreement for 59% (148/249) of samples. This was not surprising since a higher number of input mutations with ranging variant allele frequencies (VAFs) can lead to predictions of higher numbers of cancer cell populations [14]. The pipeline using the intersect of CNA calls and intersect of SNV calls had more balanced clonality predictions (42% monoclonal *vs.* 59% polyclonal; **Supplementary Figure 3D**) and these predictions agreed with the SomaticSniper-TITAN-PhyloWGS predictions for 61% (140/229) of samples. Interestingly, the PhyloWGS-linked pipeline using the union of SNV calls and the intersect of CNA calls (**Supplementary Figure 3E**) predicted clonality with similar skew to the

pipeline using the union of calls for both SNVs and CNAs (8.9% and 91.1% *vs.* 5.9% and 94.0%; monoclonal and polyclonal). Similarly, the pipeline using the intersect of SNV calls and union of CNA calls (**Supplementary Figure 3F**) predicted clonality with similar balance to the pipeline using the intersect of calls for both SNVs and CNAs (30.3% and 69.7% *vs.* 39.7% and 60.3%; monoclonal and polyclonal). This suggests that PhyloWGS-based predictions of complex polyclonal phylogenies are primarily driven by large numbers of SNV calls from MuTect, and complexity in copy number aberrations has a smaller influence on the delineation of cancer cell populations.

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 using PhyloWGS, providing an estimate of the extent to which the former underestimates true clonal complexity.

Assignments of SNVs as clonal or subclonal showed substantial deviation 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-PhyloWGS pipeline. The strongest agreement for SNV assignments between single- and multi-region reconstructions was observed for the SomaticSniper-Battenberg-PhyloWGS 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 [25], 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 4**). 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-PhyloWGS and SomaticSniper-Battenberg-PhyloWGS pipelines, respectively. One reason for the low rate of agreement for SomaticSniper-Battenberg-PhyloWGS 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 was again observed in CNA assignments between single-region and multi-region reconstructions, as indicated by the standard deviations above.

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 linked to PhyloWGS (**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. Surprisingly, multi-sample reconstructions did not detect significantly more subclones within the index lesion sample compared to the single-sample reconstruction of the index lesion ($p = 0.217$, Mann-Whitney U-Test, median number of $\text{subclones}_{\text{Index lesion, single-sample reconstruction}} = 2$, median number of $\text{subcones}_{\text{Index lesion, multi-sample reconstruction}} = 2$; **Supplementary Figure 5**). This may suggest that multi-region reconstructions detect more of the overall heterogeneity within a tumour, but do not identify additional cell populations within individual samples.

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-PhyloWGS pipeline. This was likely due to PhyloWGS only considering regions with identical copy number status across all samples of the tumour for multi-region reconstruction. Accordingly, while multi-region reconstructions can reveal more of the true intra-tumoural heterogeneity in tumours, the underestimation of subclonal diversity from

single-region reconstructions may be larger than currently appreciated. Thus as expected, multi-region reconstructions tend to find more subclones relative to those from single-regions.

Furthermore, since tumour subclones are defined by the set of mutations that are shared within the cancer cell population, we analyzed the SNVs assigned to the index-lesion during single- and multi-region reconstructions to compare the composition of the clonal populations. In almost all samples multi-region reconstruction found subclonal mutations assigned to the clonal population defined by single-region reconstruction of the index lesion. This highlights the underestimation of intra-tumoural heterogeneity from single-region analyses due to sampling bias. However, the majority of single-region clonal SNVs were indeed defined as clonal in multi-region reconstruction, except in the MuTect-TITAN-PhyloWGS pipeline (SomaticSniper-TITAN-PhyloWGS: 59.1% ± 38.5%; SomaticSniper-Battenberg-PhyloWGS: 73.2% ± 27.3%; MuTect-TITAN-PhyloWGS: 42.1% ± 29.8%; MuTect-Battenberg-PhyloWGS: 62.2% ± 33.8%; percentage of single-region clonal SNVs assigned as clonal in multi-region reconstruction, mean ± standard deviation; **Supplementary Figure 6A**). A similar analysis of subclonal SNVs showed that while most of the subclonal SNVs defined by single-region reconstructions are also identified as subclonal in multi-region reconstructions in pipelines that use MuTect, surprisingly in pipelines using SomaticSniper the majority were assigned as clonal instead (SomaticSniper-TITAN-PhyloWGS: 54.7% ± 39.6%; SomaticSniper-Battenberg-PhyloWGS: 53.4% ± 56.9%; MuTect-TITAN-PhyloWGS: 12.2% ± 14.4%; MuTect-Battenberg-PhyloWGS: 5.2% ± 5.6%; percentage of single-region subclonal SNVs assigned as clonal in multi-region reconstruction, mean ± standard deviation; **Supplementary Figure 6B**). This also highlights a potential limitation in multi-region subclonal reconstruction as SNVs in CNAs that were not shared between samples were not considered in these reconstructions. In summary, these data suggest that though multi-region reconstruction may more precisely define the evolutionary landscape of the cancer cell populations, caution should be taken in interpreting clonal and subclonal mutation compositions in both multi- and single-region studies. A consensus approach for clonal and subclonal population definition using both single- and multi- region reconstructions may be useful.

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 [26], we systematically estimated the subclonal architectures of 293 tumours using six 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 in prostate cancer.

The two subclonal CNA detection algorithms used differed modestly in their ability to define subclonal architectures, while large differences were driven by changing the SNV detection approach. Differences between these mutation callers led to major divergences in subclonal reconstruction, where pipelines using MuTect found extensive subclonal diversity. Indeed one pipeline combination, SomaticSniper-Battenberg-PhyloWGS, led to the detection of only a single population of cancer cells in the vast majority of reconstructions. Additional differences arose when the same mutational inputs were used with different reconstruction algorithms. While PhyloWGS and DPClust converged towards similar numbers of subclones, reconstructions from PyClone predicted many more monoclonal phylogenies. Future studies may benefit from applying at least one of DPClust or PhyloWGS along with PyClone when evaluating subclonal architectures, and these may provide upper- and lower-bounds on the number of cancer cell populations in single samples, respectively. The clinical impact of these differences is considerable, given findings that tumours with polyclonal architectures are at elevated risk for both relapse and distant metastasis after primary treatment [12]. Future studies also need to carefully consider the failure-rates of different reconstruction algorithms. Reconstructions relying on PyClone were successfully completed for all samples while PhyloWGS and DPClust had similar failure rates (4.4% *vs.* 3.1%, respectively). Though with the higher rates of success using MuTect and Battenberg with DPClust compared to with PhyloWGS, DPClust may be useful for those with preference for this mutation caller combination. With the size of whole-

genome sequencing datasets rapidly growing, these failures should not drastically affect statistical power in large datasets. However, these failures may be problematic for smaller studies and are likely to disproportionately affect reconstructions of tumours with a high mutational burden (*e.g.* due to excessive computational requirements).

Our evaluation of subclonal reconstruction using data from spatially distinct regions of tumours found that reconstructions relying on a single sample systematically underestimated the number of subclones in a tumour: multi-region reconstructions found twice as many subclones as single-region reconstructions, as a lower-bound estimation. This is in line with previous work in kidney cancer [6, 11] and is the first such analysis for prostate cancer. These data further agree 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 [26]. 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 provide guidance for evaluating the subclonal architecture of cancer samples.

Conclusions

We present a systematic assessment of the variability across six 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 and reconstruction algorithm in downstream solutions. 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, 20, 21, 27–30]. 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, 21].

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, 20, 21]. Briefly, raw sequencing reads from the tumour and normal samples were aligned against human reference genome build hg19 using bwa-aln (v0.5.7) [31]. 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) [32]. Tumour and normal sample-level BAMs were extracted

separately, had headers corrected with SAMtools (v0.1.9) [33] and were indexed with picard (v1.107). ContEst (v1.0.24530) [34] 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) [22]. First, Battenberg (v2.2.6) was installed with underlying ASCAT (v2.5) [35] 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) [36] pipeline for whole-genome sequence preprocessing and subclonal copy number assessment. GC and mappability files for bias correction were prepared using HMMcopy (v0.1.1) [37]and bowtie (v2.2.6) [38] on the hg19 reference genome. Heterogeneous positions in the sequence data were identified by MutationSeq (v4.3.7) [39] 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 [22] was used to obtain the cellularity, ploidy and subclonal CNAs for downstream analysis.

We used MuTect (v1.1.4) [24] and SomaticSniper (v1.0.2) [23] for the identification of somatic single nucleotide variants from whole-genome sequencing data. MuTect was run to obtain candidate SNVs with dbSNP138 [40], COSMIC (v66) [41] 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 [20]. 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 (v0.1.9) [33], 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) [42] using RefGene, subjected to blacklist filtering to remove known germline contaminants and sequencing artifacts and whitelist filtering through COSMIC (v70) [41]. This was done before downstream subclonal reconstruction.

Subclonal Reconstruction of Tumours using PhyloWGS

We used the cnv-int branch of PhyloWGS (https://github.com/morrislab/phylowgs/tree/cnvint, commit: 3b75ba9c40cfb27ef38013b08f9e089fa4efa0c0) [14] 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 segments shorter than 10 Kbp. 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 (3b75ba9) inputs for each sequence sample, by combining TITAN-SomaticSniper, TITAN-MuTect, Battenberg-SomaticSniper and Battenberg-MuTect as CNA-SNV mutation caller pairs. All default parameters were used, including limiting the number of SNVs considered to 5,000 for the

interest of runtime. The PhyloWGS (3b75ba9) 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 (3b75ba9) were used to perform subclonal reconstruction (SomaticSniper-TITAN-PhyloWGS, TSP; MuTect-TITAN-PhyloWGS, TMP; SomaticSniper-Battenberg-PhyloWGS, BSP; MuTect-Battenberg-PhyloWGS, BMP).

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 (3b75ba9) 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 (3b75ba9), 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 combinations of CNA and SNV calls as inputs to PhyloWGS (3b75ba9). The single- and multi-region reconstructions from these 10 tumours provided data for the downstream analysis of intra-tumoural heterogeneity in subclonal reconstruction.

Union and Intersection of Mutation Calling Algorithms

We obtained the union and intersection of raw SNV calls by SomaticSniper (v1.0.2) and MuTect (v1.1.4) for each tumour sample using vcf-isec of vcftools (v0.1.15). The union and intersection sets of SNVs were then annotated and filtered with the same method as described above before being used in subsequent analysis.

We determined the union and intersection of CNAs called by TITAN (v1.11.0) and Battenberg (v2.2.6), first parsed using parse_cnvs.py script of PhyloWGS (3b75ba9) for consistent formatting, on a per base-pair basis. The intersection of CNAs, based on genomic coordinates and major and minor copy number, was determined using the GenomicRanges (v1.28.6) package in R (v3.2.5). Regions with disagreeing copy number were identified using bedtools (v2.27.1) and bedr (v1.0.6). A region is defined to have an algorithm-unique CNA if one algorithm identified a copy number aberration for the region while the other identified it as copy number neutral (major and minor copy number of 1). Regions were both algorithms identified different copy number aberrations were classified as disagreements. The union set of CNAs thus contained the intersection of CNAs and CNAs unique to either TITAN (v1.11.0) or Battenberg (v2.2.6), and disagreement regions were excluded as there was no obvious way to resolve the discrepancies. In contrast to TITAN, when a region is determined to have a subclonal aberration, Battenberg (v2.2.6) produces two entries, a clonal and subclonal copy number for the same genomic region. These regions were labelled Battenberg-unique for its clear delineation of subclonal CNAs. However, the TITAN (v.1.11.0) copy number aberration result for the region (if any) is used in the union of CNAs to avoid conflicting CNA calls in the same region, as one cannot combine the clonal Battenberg (v2.2.6) results with the TITAN (v1.11.0) aberrations. The union and intersection sets of CNAs were further filtered to remove any segments under 10 Kbp. These were used in the create_phylowgs_inputs.py script to generate inputs for PhyloWGS

(3b75ba9) in different combinations of CNA and SNV mutation sets for subsequent subclonal reconstruction.

Subclonal Reconstruction of Tumours using PyClone and DPClust

We used PyClone (v0.12.9) [16] for the subclonal reconstruction of 293 tumours with single samples, using somatic SNVs obtained from SomaticSniper (v.1.0.2) and CNAs obtained from TITAN (v1.11.0), processed as described above. A mutation input file was created for each sample by obtaining the tumour reference and variant read counts for each SNV detected in the VCF and annotating them with the TITAN-determined (v1.11.0) major and minor copy number for the position. SNVs in regions without copy number information were discarded, and the normal copy number was set to 2 for autosomes and 1 for chromosomes X and Y. The mutation input file, along with tumour content as predicted by TITAN (v1.11.0) were used as inputs for the run_analysis_pipeline to launch PyClone (v0.12.9) [16], using 12345 as the seed and otherwise default parameters. PyClone (v0.12.9) identified subclonal populations were pruned using the same heuristic as that for PhyloWGS (3b75ba9). Specifically, for each tumour sample, a mutation cluster was pruned if it had fewer than five supporting SNVs or a cancer cell fraction below 0.02, unless it was the only mutation cluster detected. Pruned clusters were merged with their nearest neighbors in cancer cell fraction. Moreover, two clusters were merged if they differed in cancer cell fraction by less than 0.02.

Similarly, we used DPClust (v2.2.5) [15] for subclonal reconstruction, using somatic SNVs obtained from MuTect (v1.1.4) and CNAs obtained from Battenberg (v2.2.6), processed as described above. DPClust (v2.2.5) was run using the dpc.R pipeline available *via* the DPClust SMC-HET Docker (https://github.com/Wedge-Oxford/dpclust_smchet_docker, commit a1ef254), using also dpclust3p (v1.0.6). The inputs for each tumour sample were the MuTect (v1.1.4) VCF, subclonal copy number, and cellularity, ploidy, and psi as predicted by Battenberg (v2.2.6), using 12345 as the seed and otherwise default parameters. The results in the subchallenge1C.txt output file were taken as the mutation clustering solution to obtain the number of subclones predicted by DPClust (v2.2.5) [15].

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 child, 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 (3b75ba9) 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 sample and created a separate profile for CNAs obtained from each of the four PhyloWGS-based 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 two. 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]. Within each previously identified sample subtype, we calculated the average clonal and subclonal CNA profiles using PhyloWGS-processed clonal and subclonal CNAs of the samples, respectively. Samples that were assigned a “branch” (subclonal) subtype in the

SomaticSniper-TITAN pipeline [12] but had no subclonal populations detected in another pipeline were excluded from subclonal subtype analysis for that pipeline. Samples that had no subclonal populations detected in the SomaticSniper-TITAN pipeline and were therefore never assigned to a "branch" (subclonal) subtype were not considered in any subclonal subtype analysis. 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 PhyloWGS-based 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 PhyloWGS-based 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 PhyloWGS-based pipelines as biased to have a CNA clonally or subclonally. Using all default parameters of gprofileR (v0.6.1) in R (v3.2.5) [43], 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 [44]. Cytoscape (v3.4.0) was used to visualize significant pathways [45]. Since all genes identified as significantly differentially altered were biased to be altered clonally, 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 PhyloWGS-based 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, 20]. 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 PhyloWGS-based pipelines.

Driver SNVs and CNAs of each sample were categorized by the number of PhyloWGS-based pipelines they were called in. Since four PhyloWGS-based 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 based on PhyloWGS. 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) [46] and BPG (v5.3.4) [47] packages. All boxplots show the median (center line), upper and lower quartiles (box limits), and whiskers extend to the minimum and maximum values within 1.5 times the interquartile range (Tukey boxplots). 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.

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 [27]

WGS Data – Berger *et al*., 2011: dbGaP, phs000330.v1.p1 [28]

WGS Data – CPC-GENE Fraser *et al.*, 2017: EGA, EGAS00001000900; GEO: GSE84043 [20]

WGS Data – The Cancer Genome Atlas Research Network, 2015: https://portal.gdc.cancer.gov/projects/TCGA-PRAD [29]

WGS Data - Weischenfeldt *et al.*, 2013: EGA, EGAS00001000400 [30]

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

WGS Data – CPC-GENE Taylor *et al.*, 2017: EGA, EGAS00001001615; EGA, EGAS000010000258 [21]

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

Code availability

Code is available upon request.

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 Dr. Jüri Reimand for guidance and support. We also 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;545:446–51. doi:10.1038/nature22364.
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;376:2109–21. doi:10.1056/NEJMoa1616288.
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;173:581–594.e12. doi:10.1016/j.cell.2018.03.057.
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;173:595–610.e11. doi:10.1016/j.cell.2018.03.043.
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;520:353–7. doi:10.1038/nature14347.
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;366:883–92. doi:10.1056/NEJMoa1113205.
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. 2012;149:994–1007. doi:10.1016/j.cell.2012.04.023.
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. 2015;47:736–45. doi:10.1038/ng.3315.
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. 2015;47:367–72. doi:10.1038/ng.3221.
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;173:611–623.e17. doi:10.1016/j.cell.2018.02.020.

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. 2014;46:225–33. doi:10.1038/ng.2891.
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. 2018;173:1003–13. doi:10.1016/j.cell.2018.03.029.
13. Jiao W, Vembu S, Deshwar AG, Stein L, Morris Q. Inferring clonal evolution of tumors from single nucleotide somatic mutations. BMC Bioinformatics. 2014;15:35. doi:10.1186/1471-2105-15-35.
14. 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;16:35. doi:10.1186/s13059-015-0602-8.
15. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat Commun. 2014;5:2997. doi:10.1038/ncomms3997.
16. 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. 2014;11:396–8. doi:10.1038/nmeth.2883.
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. 2014;10:e1003703. doi:10.1371/journal.pcbi.1003703.
18. Oesper L, Mahmoody A, Raphael BJ. THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome Biol. 2013;14:R80. doi:10.1186/gb-2013-14-7-r80.
19. Dentro SC, Leshchiner I, Haase K, Tarabichi M, Wintersinger J, Deshwar AG, et al. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. bioRxiv. 2018;:312041. doi:10.1101/312041.
20. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017;541:359–64. doi:10.1038/nature20788.
21. 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;8:13671. doi:10.1038/ncomms13671.

22. 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;24:1881–93. doi:10.1101/gr.180281.114.
23. 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;28:311–7. doi:10.1093/bioinformatics/btr665.
24. 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;31:213–9.
25. 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;49:1015–24. doi:10.1038/ng.3891.
26. Salcedo A, Tarabichi M, Espiritu SMG, Deshwar AG, David M, Wilson NM, et al. Creating Standards for Evaluating Tumour Subclonal Reconstruction. bioRxiv. 2018;:310425.
27. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated Evolution of Prostate Cancer Genomes. Cell. 2013;153:666–77. doi:10.1016/j.cell.2013.03.021.
28. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. Nature. 2011;470:214–20. doi:10.1038/nature09744.
29. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163:1011–25. doi:10.1016/j.cell.2015.10.025.
30. 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;23:159–70. doi:10.1016/j.ccr.2013.01.002.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60. doi:10.1093/bioinformatics/btp324.
32. 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;20:1297–303. doi:10.1101/gr.107524.110.

33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. doi:10.1093/bioinformatics/btp352.
34. 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;27:2601–2. doi:10.1093/bioinformatics/btr446.
35. 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;107:16910–5. doi:10.1073/pnas.1009843107.
36. 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;6:1–10. doi:10.1093/gigascience/gix042.
37. Lai D, Ha G, Shah S. HMMcopy: Copy number prediction with correction for GC and mappability bias for HTS data. 2016. http://bioconductor.org/packages/release/bioc/html/HMMcopy.html.
38. 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:R25. doi:10.1186/gb-2009-10-3-r25.
39. 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;28:167–75. doi:10.1093/bioinformatics/btr629.
40. 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;29:308–11.
41. 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;43:D805–11.
42. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164–e164. doi:10.1093/nar/gkq603.

43. 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. 2007;35 Web Server issue:W193-200. doi:10.1093/nar/gkm226.

44. 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;466:368–72. doi:10.1038/nature09146.

45. 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;13:2498–504. doi:10.1101/gr.1239303.

46. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35. doi:10.1186/1471-2105-12-35.

47. 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. 2017;: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 PyClone, DPClust and four pipelines linked to PhyloWGS, each with different combinations of SNVs and CNAs. For tumours with samples from multiple regions, reconstructions of subclonal architectures were performed by considering samples from all regions using each of the four PhyloWGS-linked 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 of the same tumour, 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 linked to PhyloWGS. **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 1 Mbp genomic bins 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. TSP,

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

Figure 3 – Differentially Altered CNAs

a) Venn diagram of the genes with CNAs that are significantly biased to be altered clonally (*i.e.*, early during tumour evolution) or subclonally (*i.e.*, late during tumour evolution) during tumour evolution based on reconstructions from four pipelines linked to PhyloWGS. Numbers in overlapping areas correspond to the genes with biased timing based on multiple pipelines with the same directionality (*i.e.*, consistent bias for the same gene towards early or late alterations). CNAs in a core set of 270 genes show bias timing based on all four pipelines. All 270 of these genes are enriched for alterations early during tumour evolution. **b**) A pathway enrichment map based on the 270 genes that are consistently 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 linked to PhyloWGS. **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-PhyloWGS pipeline. Pipelines that use MuTect detected vastly more subclonal SNVs compared to pipelines that use SomaticSniper. **b**) The proportion of disagreement for SNV calls in genes, where one pipeline called a SNV in a gene while the other called the gene wildtype, was 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 linked to PhyloWGS and a consensus. **a**) Sub-panels show data for five genes known to be recurrently altered by SNVs in localized prostate cancer. Each column within each sub-panel represent one tumour sample and clonal and subclonal SNVs detected by each pipeline are indicated. **b**) Sub-panels show data for eight genes known to be recurrently altered by CNAs in localized prostate cancer. Clonal and subclonal CNAs identified by each pipeline are shown along with a consensus. SNVs (**a**) and CNAs (**b**) in known prostate cancer driver genes were typically detected clonally. The consensus for clonal driver CNA and SNV 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 linked to PhyloWGS, one linked to DPClust and one linked to PyClone **a-b**) The proportion of samples identified to be monoclonal or polyclonal by each of the four pipelines linked to PhyloWGS (**a**) and the breakdown of samples into tree structures (**b**). The SomaticSniper-Battenberg pipeline linked to PhyloWGS called drastically more monoclonal phylogenies than the other PhyloWGS-linked pipelines. PhyloWGS-linked pipelines with MuTect as the SNV caller reconstructed more complex phylogenies with up to five subclones. **c**) Comparisons of the number of predicted subclones by PhyloWGS and DPClust using SNV inputs from MuTect and CNA inputs from Battenberg. **d**) Comparisons of the number of predicted subclones by PhyloWGS and PyClone using SNV inputs from SomaticSniper and CNA inputs from TITAN. For **c-d**, each marker shows the number of subclones called for one sample. Background shading highlights samples where the two reconstructions predicted different numbers of 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 using PhyloWGS. **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-PhyloWGS pipeline achieved the highest proportion of clonality assignment concordance. In the mismatches between single-region and multi-region

reconstructions, 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 PhyloWGS-linked 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, because the X chromosome was not considered by pipelines using Battenberg.

Supplementary Figure 2 – Comparing Clonality Between Pipelines Linked to PhyloWGS

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 – Comparing Clonality Between Pipelines with the Union and Intersection of Mutational Calls linked to PhyloWGS

a) Unique and intersecting SNV calls by MuTect and SomaticSniper. MuTect calls substantially more unique SNVs compared to SomaticSniper. **b**) Unique and intersecting CNA calls by TITAN and Battenberg. **c-f**) Comparing predictions of clonality from the SomaticSniper-TITAN pipeline linked to PhyloWGS to other PhyloWGS-linked pipelines that had SNV and CNA inputs based on the union or intersection of mutational calls. PhyloWGS-based pipelines using the union of SNV calls tend to have clonality predictions skewed towards polyclonal reconstructions.

Supplementary Figure 4 – Discrepancies Between Single-Region and Multi-Region CNA Assignments

The disagreement in clonal and subclonal CNA assignments between single- and multi-region subclonal reconstructions based on PhyloWGS. CNAs are compared by 1Mbp genomic bins between single-region reconstructions and their corresponding multi-region reconstructions. The SomaticSniper-Battenberg-PhyloWGS 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 5 – 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 based on PhyloWGS. 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 through 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 Figure 6 – Index Lesion Clonality in Single- and Multi- Region Reconstructions

The clonal and subclonal populations defined by index lesion single-region reconstructions based on SNV assignments as compared to multi-region reconstructions. Proportion of SNVs that were assigned as clonal in multi-region reconstructions that were assigned as clonal **a**) and subclonal **b**) in single-region reconstructions of the index lesion, by each pipeline.

Supplementary Tables

Supplementary Table 1 – Single-Region Subclonal Reconstruction Data

Subclonal reconstruction results for 293 tumours with single-region sequencing based on PhyloWGS. 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 linked to PhyloWGS 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.

Figure 1

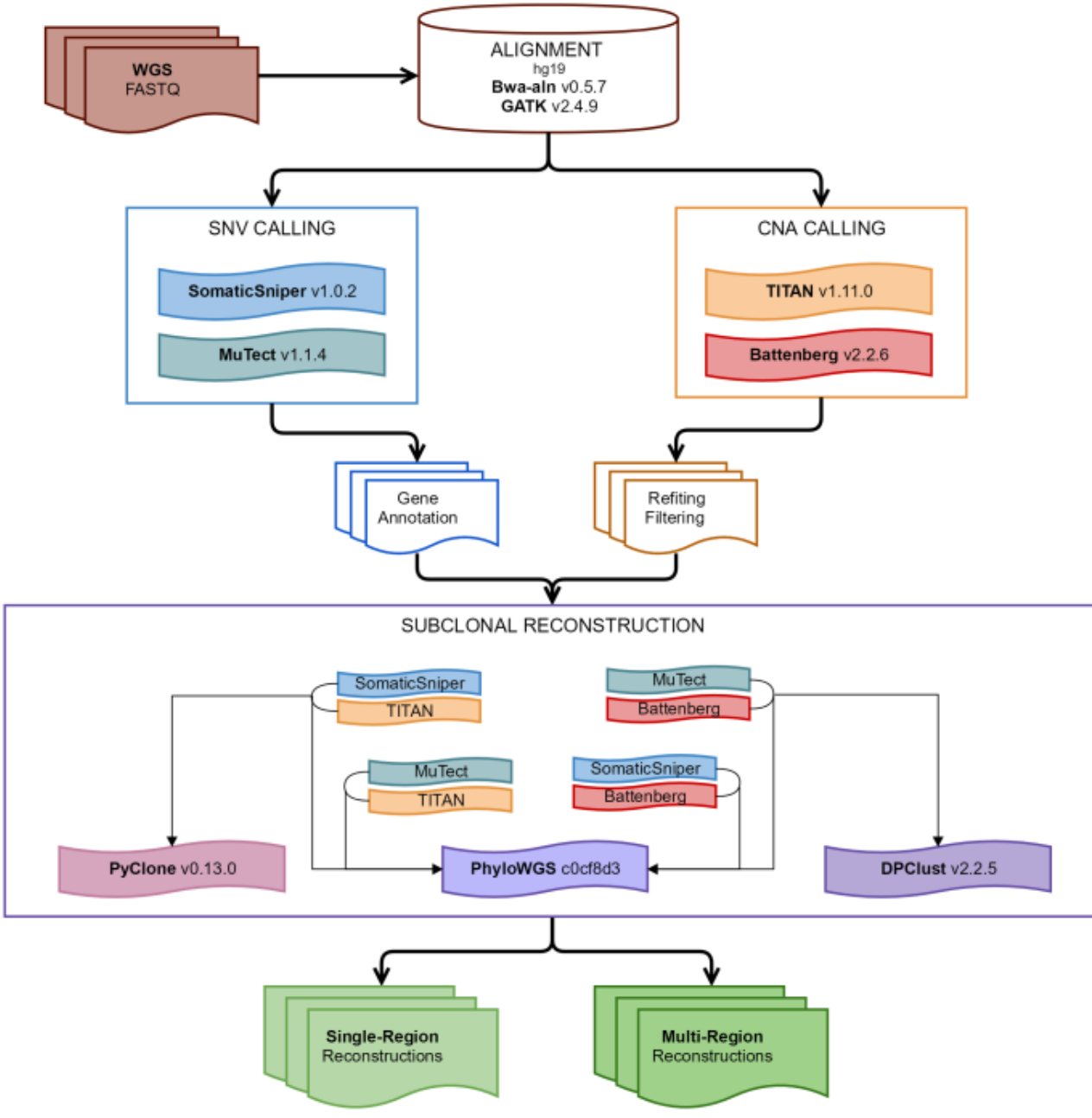

WGS
FASTQ
ALIGNMENT
hg19
Bwa-aln v0.5.7
GATK v2.4.9
SNV CALLING
SomaticSniper v1.0.2
MuTect v1.1.4
CNA CALLING
TITAN v1.11.0
Battenberg v2.2.6
Gene
Annotation
Refiting
Filtering
SUBCLONAL RECONSTRUCTION
SomaticSniper
TITAN
MuTect
Battenberg
MuTect
TITAN
SomaticSniper
Battenberg
PyClone v0.13.0
PhyloWGS c0cf8d3
DPClust v2.2.5
Single-Region
Reconstructions
Multi-Region
Reconstructions

Figure 2

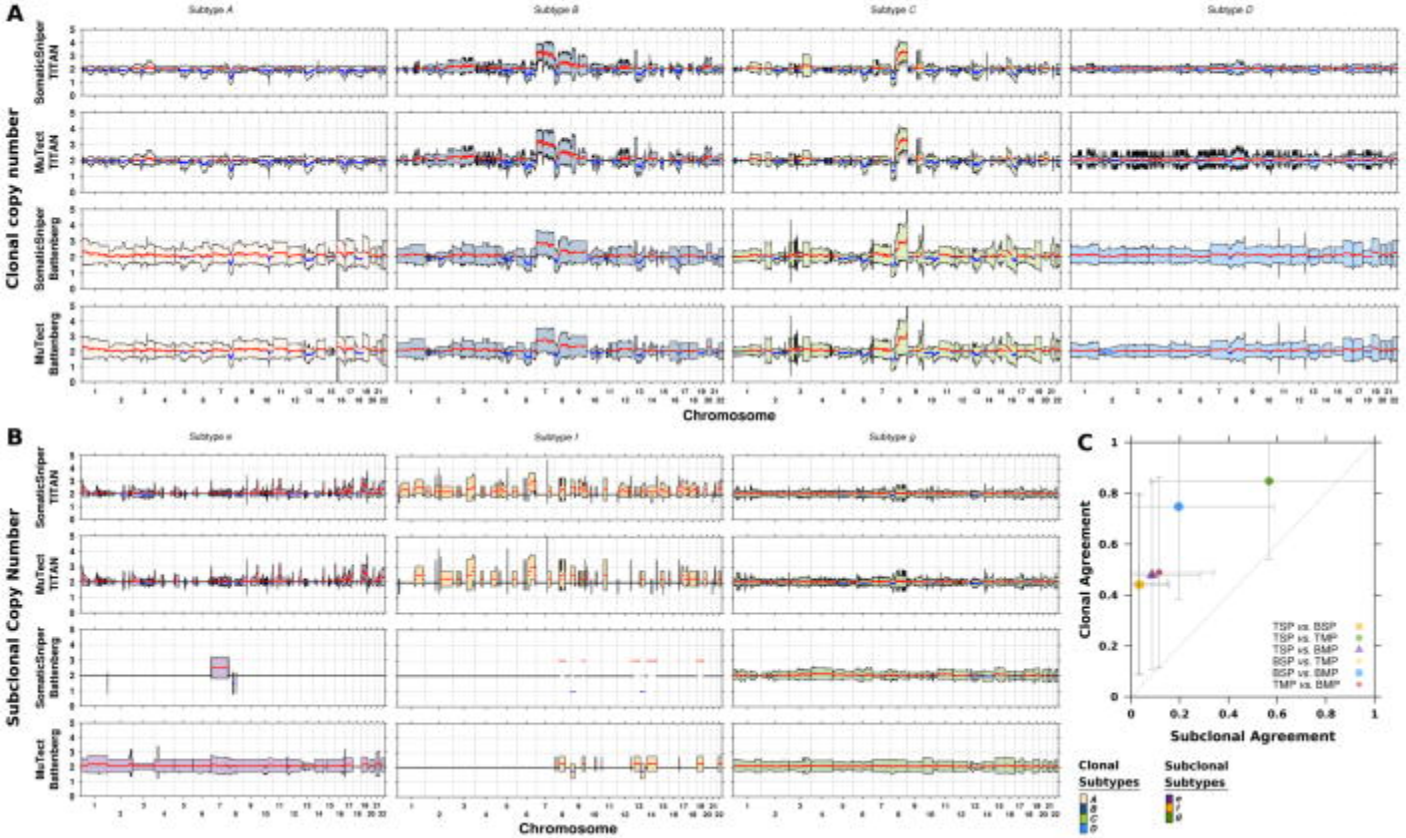
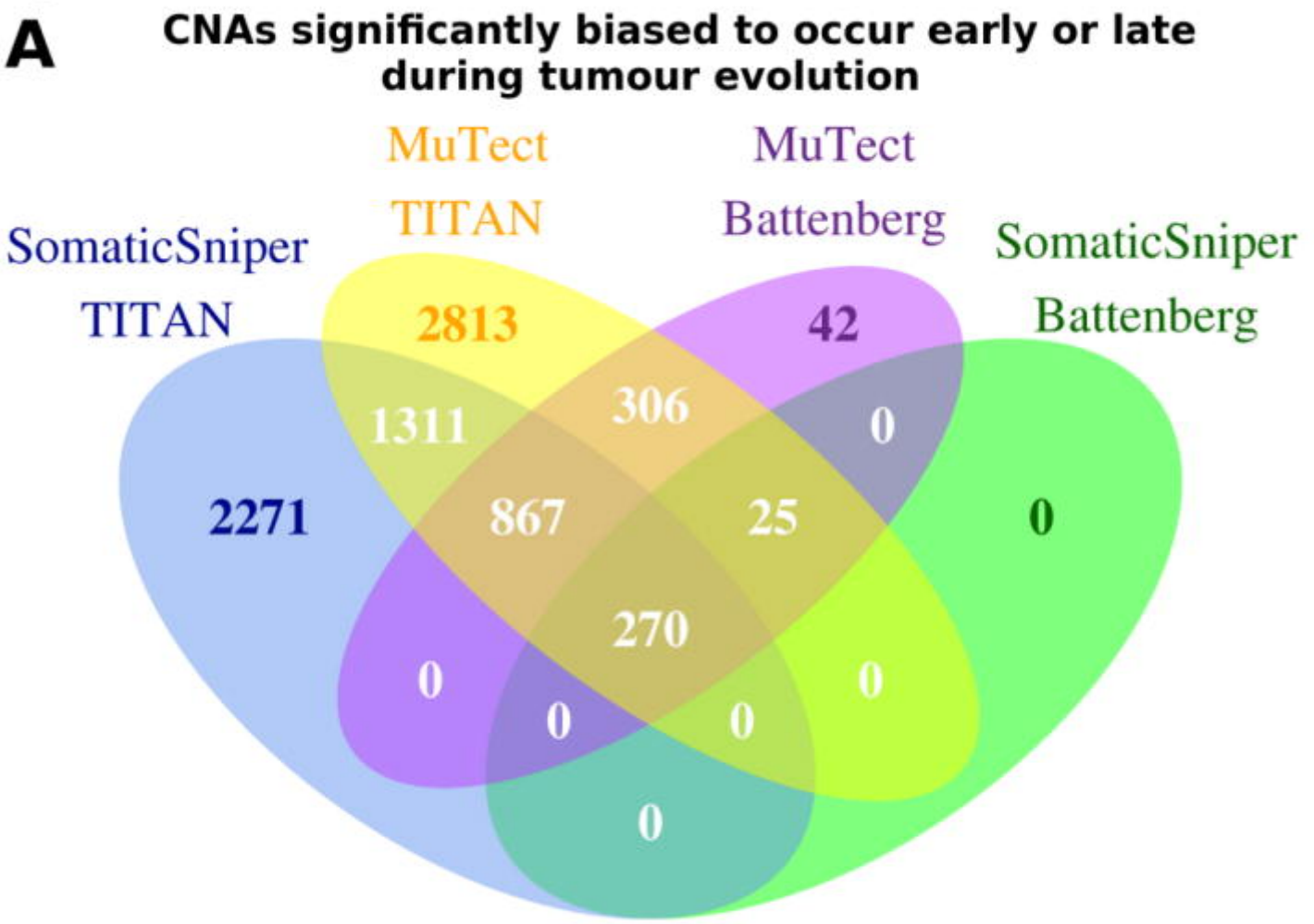

A
Subtype A
Subtype B
Subtype C
Subtype D
Clonal copy number
SomaticSniper TITAN
MuTect TITAN
SomaticSniper Battenberg
MuTect Battenberg
Chromosome
B
Subtype e
Subtype f
Subtype g
Subclonal Copy Number
SomaticSniper TITAN
MuTect TITAN
SomaticSniper Battenberg
MuTect Battenberg
Chromosome
C
Clonal Agreement
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
CNAs significantly biased to occur early or late during tumour evolution
MuTect TITAN
MuTect Battenberg
SomaticSniper TITAN
SomaticSniper Battenberg
2813
42
1311
306
0
2271
867
25
0
270
0
0
0
0
0

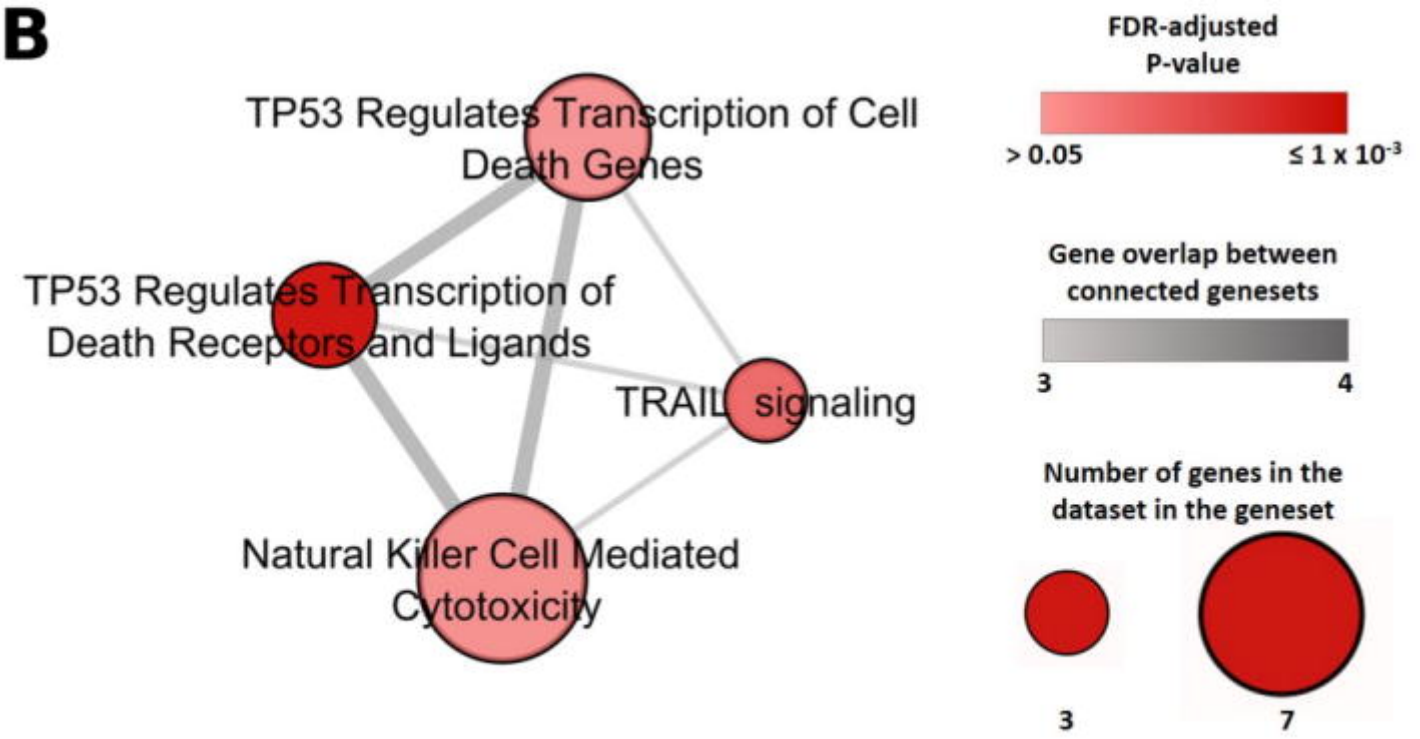

B
TP53 Regulates Transcription of Cell Death Genes
TP53 Regulates Transcription of Death Receptors and Ligands
TRAIL signaling
Natural Killer Cell Mediated Cytotoxicity
FDR-adjusted P-value
> 0.05
≤ 1 x 10-3
Gene overlap between connected genesets
3
4
Number of genes in the dataset in the geneset
3
7

Figure 4

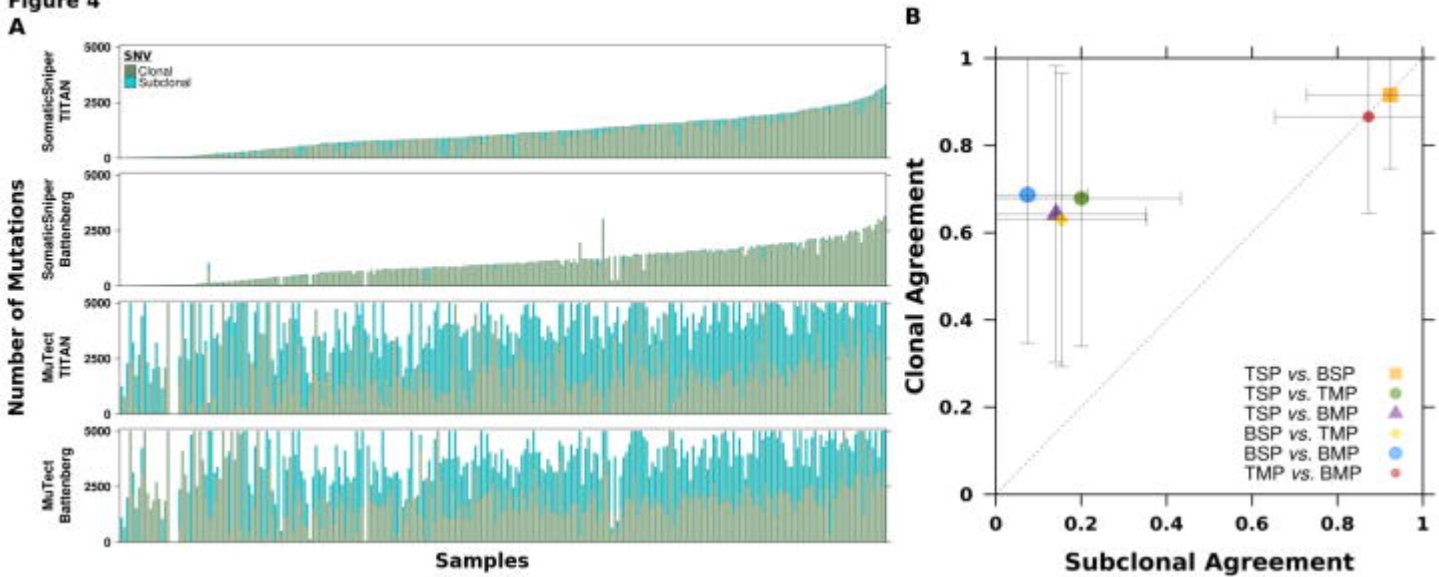

A
SNV
Clonal
Subclonal
SomaticSniper TITAN
SomaticSniper Battenberg
MuTect TITAN
MuTect Battenberg
Number of Mutations
5000
2500
0
Samples
B
Clonal Agreement
Subclonal Agreement
TSP vs. BSP
TSP vs. TMP
TSP vs. BMP
BSP vs. TMP
BSP vs. BMP
TMP vs. BMP

Figure 5

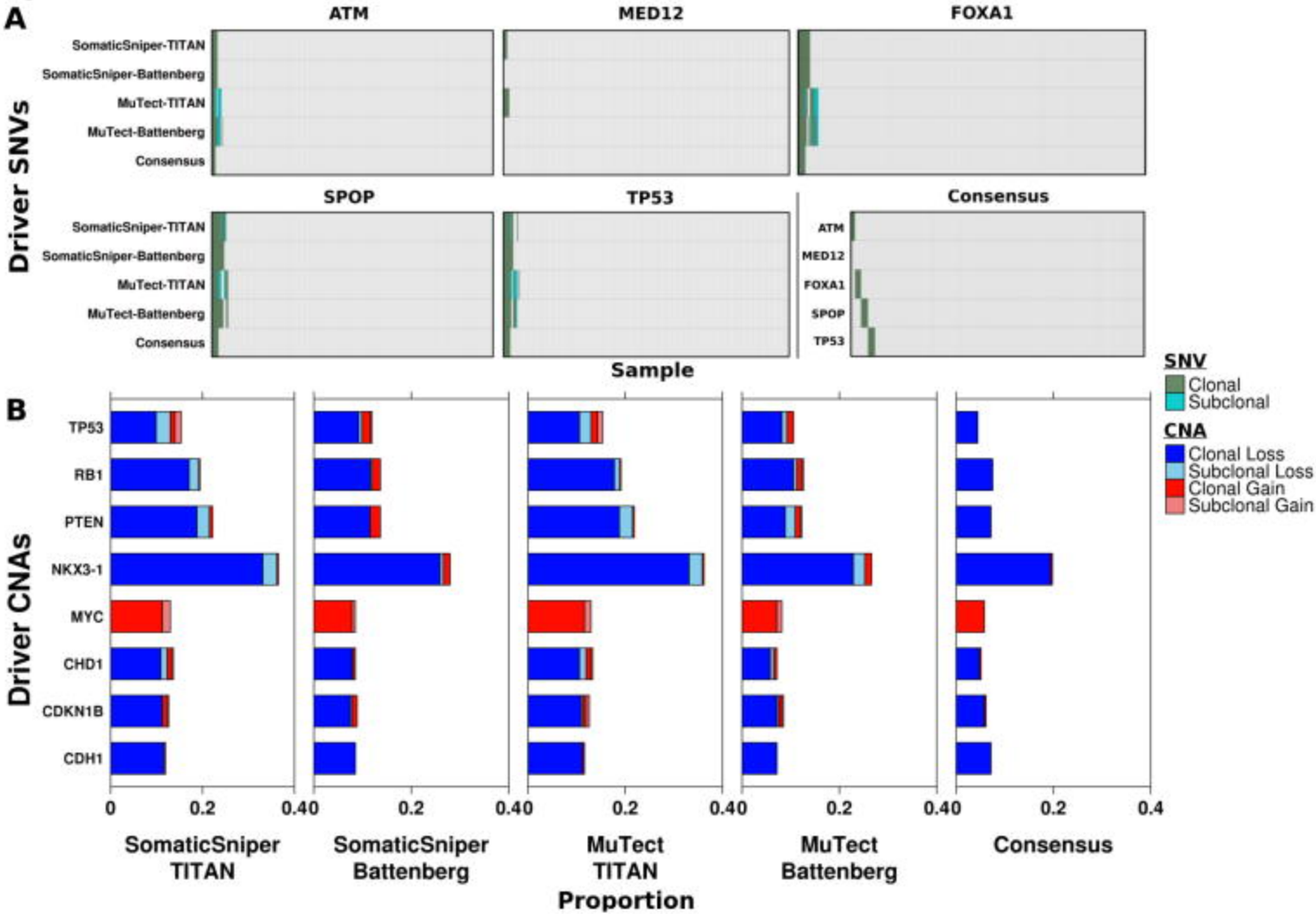

A
ATM
MED12
FOXA1
SPOP
TP53
Consensus
Driver SNVs
SomaticSniper-TITAN
SomaticSniper-Battenberg
MuTect-TITAN
MuTect-Battenberg
Consensus
ATM
MED12
FOXA1
SPOP
TP53
Sample
SNV
Clonal
Subclonal
CNA
Clonal Loss
Subclonal Loss
Clonal Gain
Subclonal Gain
B
Driver CNAs
TP53
RB1
PTEN
NKX3-1
MYC
CHD1
CDKN1B
CDH1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

Figure 6

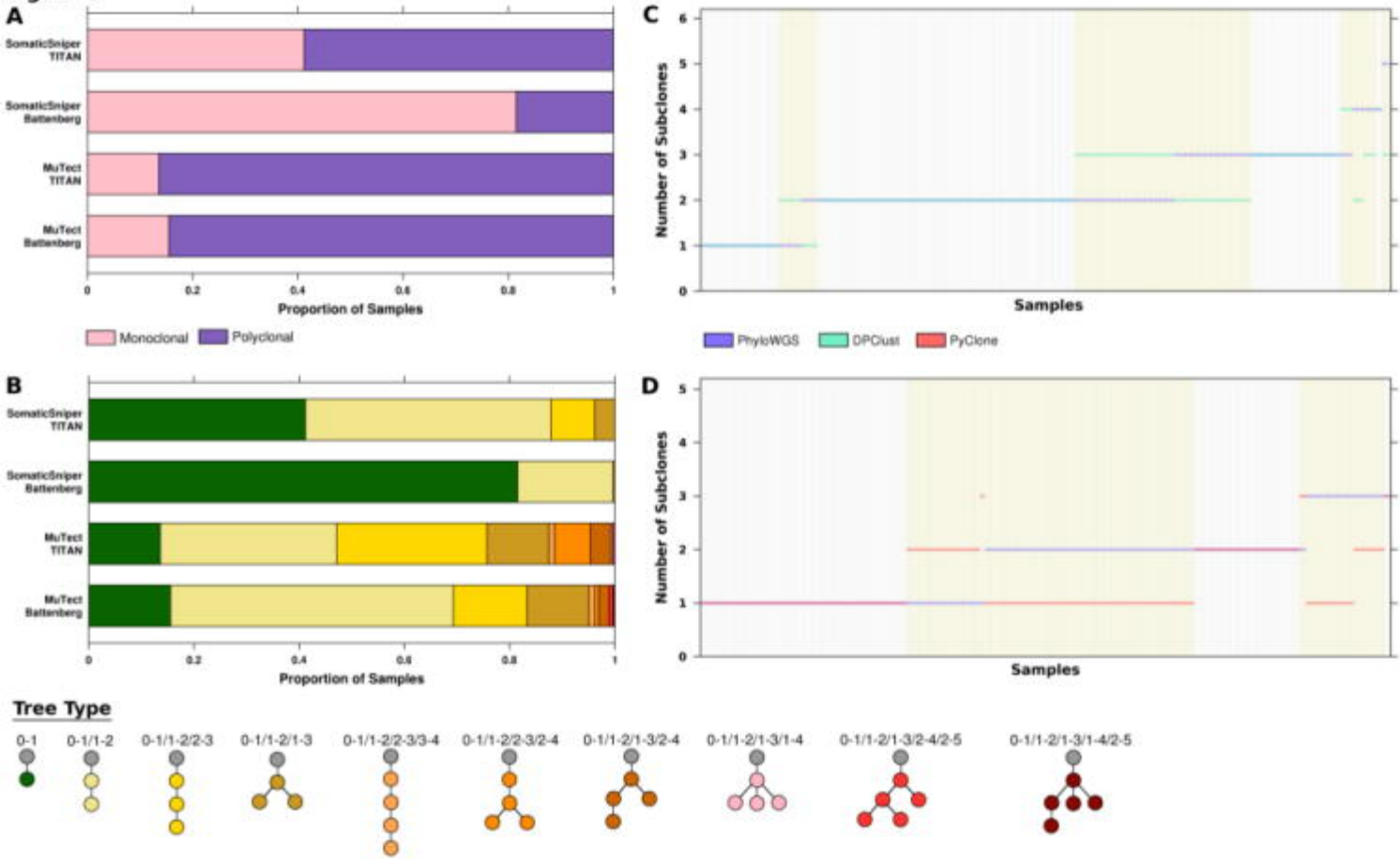
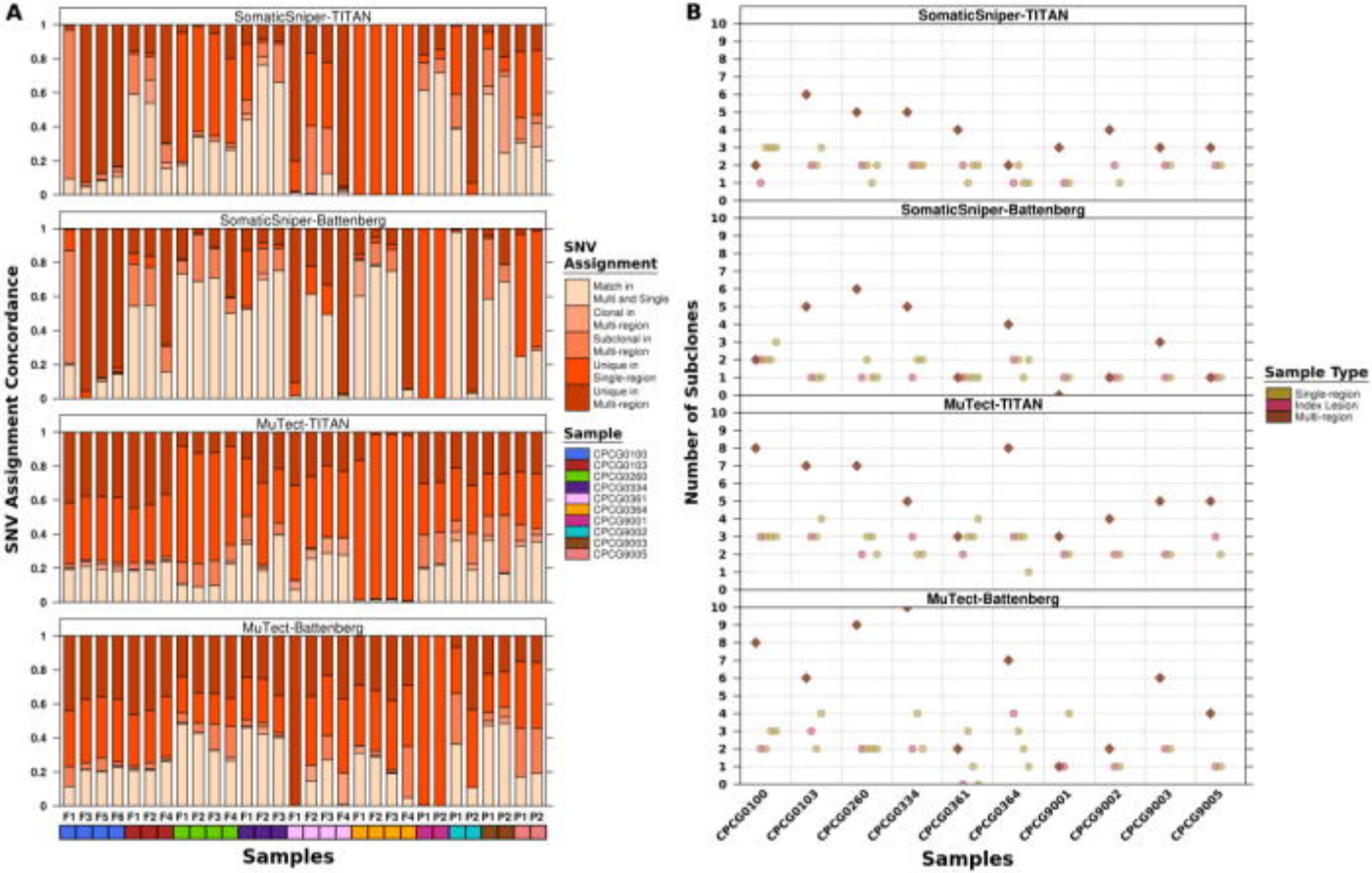

A
SomaticSniper TITAN
SomaticSniper Battenberg
MuTect TITAN
MuTect Battenberg
0
0.2
0.4
0.6
0.8
1
Proportion of Samples
Monoclonal
Polyclonal
B
SomaticSniper TITAN
SomaticSniper Battenberg
MuTect TITAN
MuTect Battenberg
0
0.2
0.4
0.6
0.8
1
Proportion of Samples
C
Number of Subclones
0
1
2
3
4
5
6
Samples
PhyloWGS
DPClust
PyClone
D
Number of Subclones
0
1
2
3
4
5
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 Assignment Concordance
SNV Assignment
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
Samples
B
SomaticSniper-TITAN
SomaticSniper-Battenberg
MuTect-TITAN
MuTect-Battenberg
Number of Subclones
Sample Type
Single-region
Index Lesion
Multi-region
CPCG0100
CPCG0103
CPCG0260
CPCG0334
CPCG0361
CPCG0364
CPCG9001
CPCG9002
CPCG9003
CPCG9005
Samples