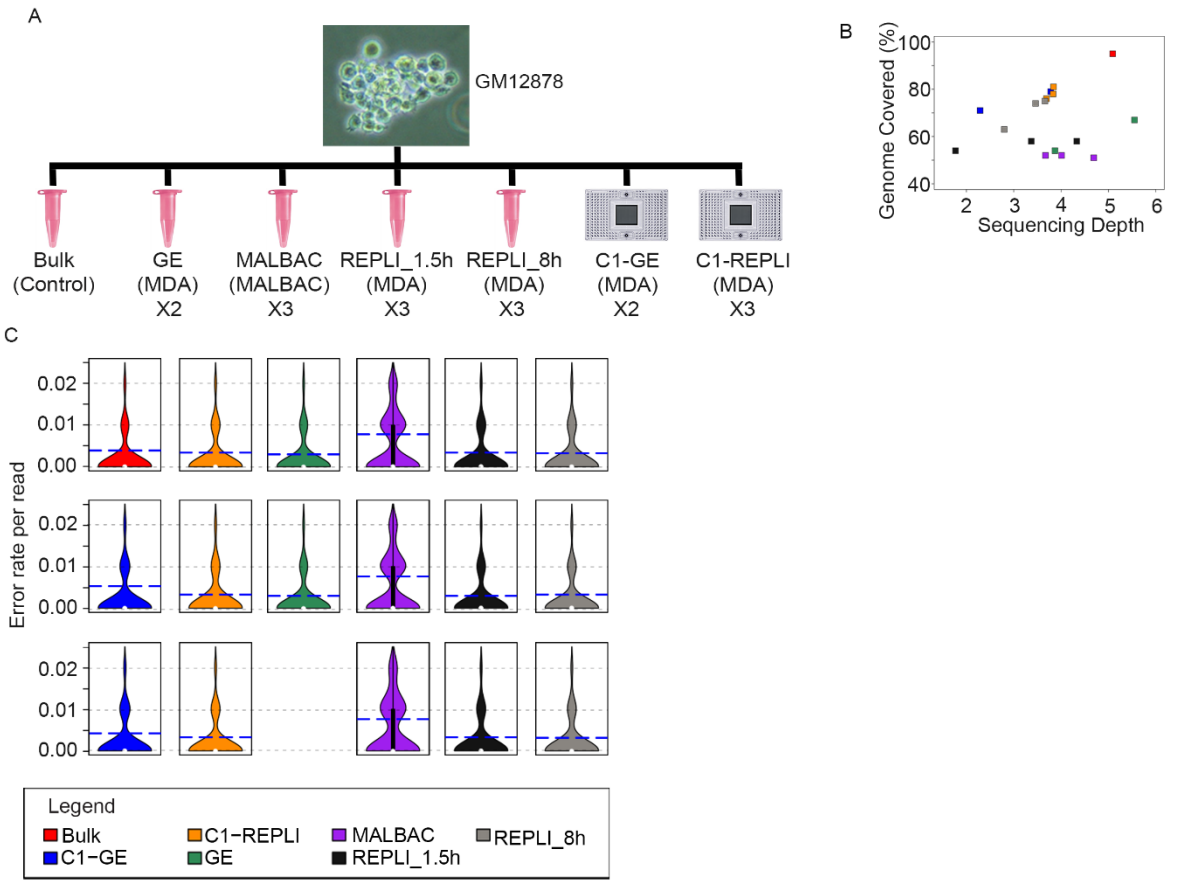
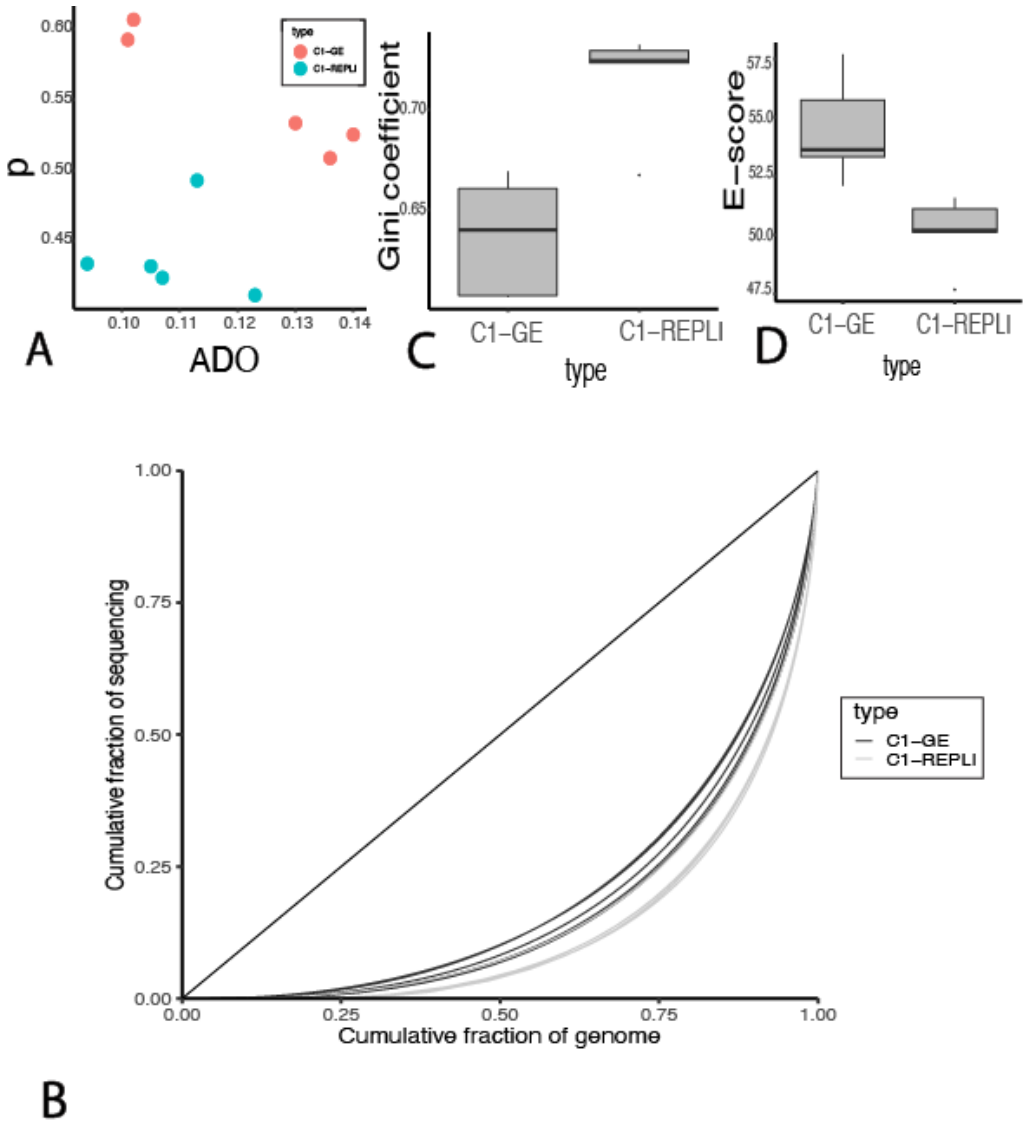
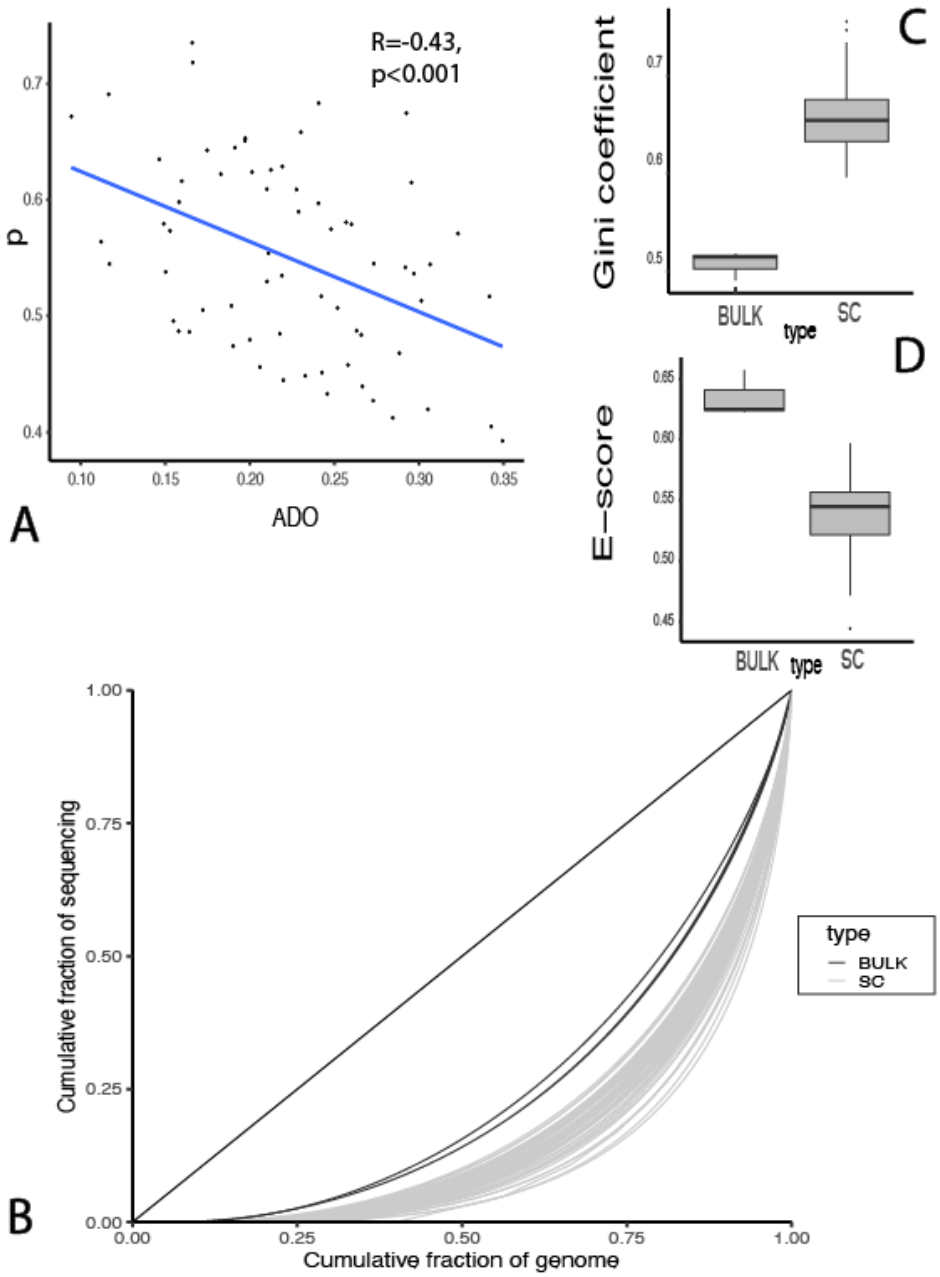


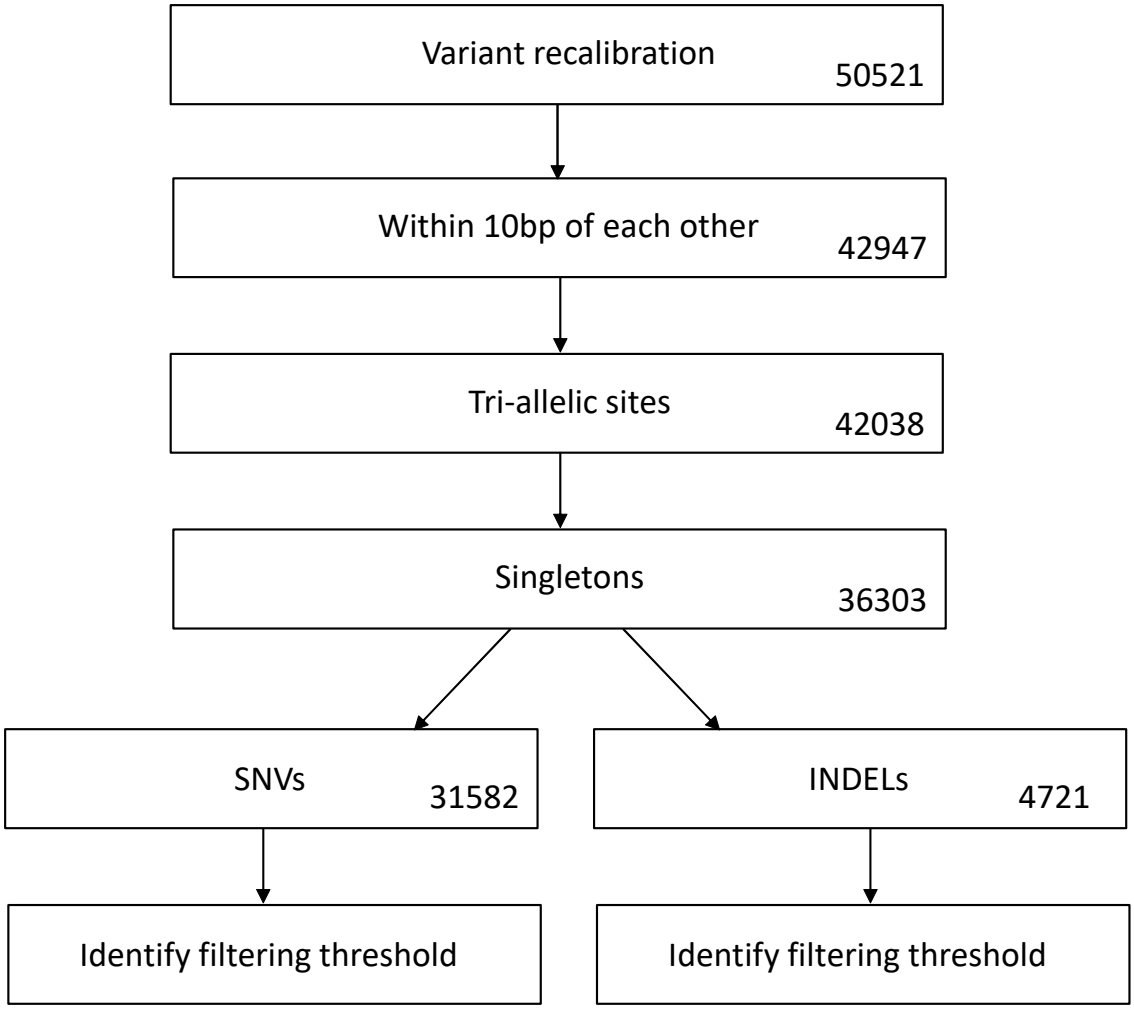
Supplementary Figure 1. Comparison of different metrics for GM12878 single cells amplified by different WGA methods. **(A)** Summary of experimental setup for comparison of different WGA methods. A total of four tube based amplification methods and two microfluidics based amplification methods were used. Unamplified DNA of a bulk sample was used as the control for this experiment. GE, Healthcare illustra GenomiPhi V2 DNA Amplification Kit (2 cells); MALBAC, MALBAC protocol based on Yikon Genomics MALBAC Single Cell WGA kit (3 cells); Repli_1.5h, Qiagen Repli-g single cell kit with 1.5 hours of amplification reaction (3 cells); Repli_8h, Qiagen Repli-g single cell kit with 8 hours of amplification reaction (3 cells); C1-GE, Healthcare illustra GenomiPhi V2 DNA Amplification Kit on C1 Autoprep System (2 cells); C1-Repli, Qiagen Repli-g single cell kit on C1 Autoprep System (3 cells); pink tubes represent manual tube-based protocols, grey arrays represent microfluidics-based protocols. **(B)** Plot of genome covered against sequencing depth. Color code of data points matches sample description in C. **(C)** Violin plot showing the error rate per read. The y-axis shows the error rate per read, the x-axis represents the number of reads. MDA methods tend to have lower mean error rates (indicated by the blue lines) compared to MALBAC.



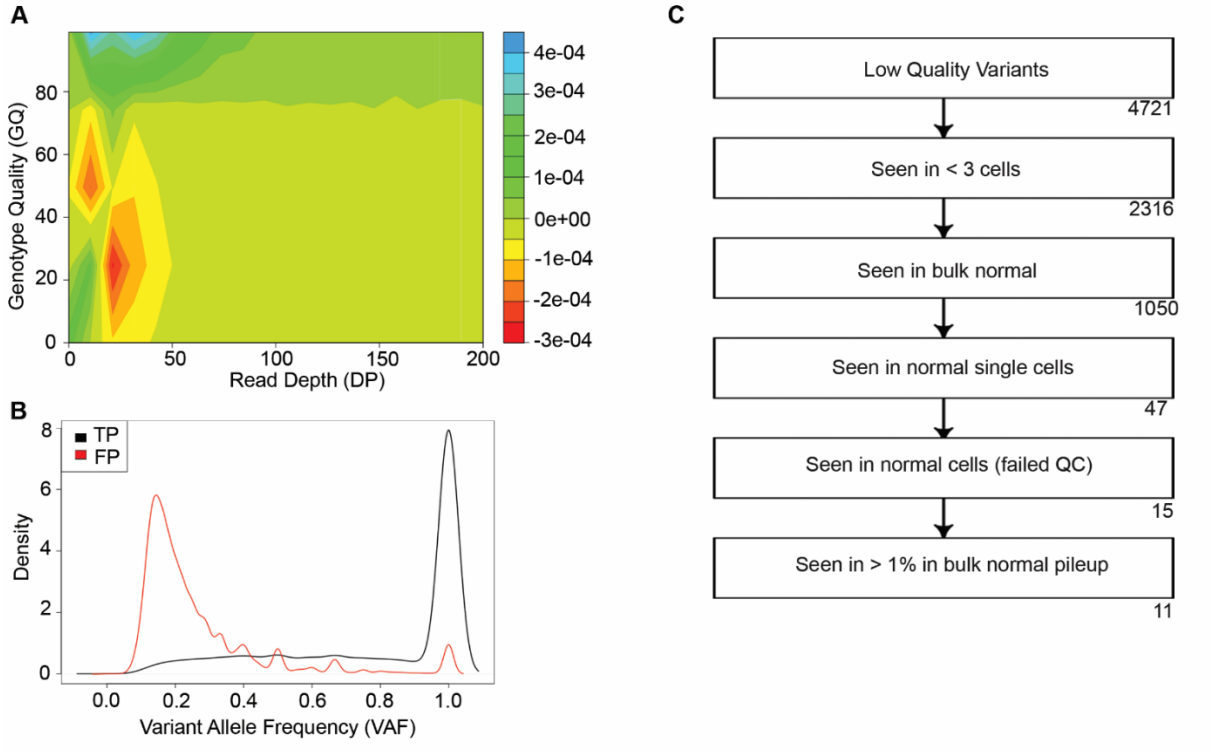
Supplementary Figure 2. Comparison of different metrics for GM12878 single cells amplified by two WGA methods (C1-GE and C1-REPLI) followed by exome sequencing. **(A)** Scatterplot showing *p* as the probability of an allele being detected versus Allelic Drop Out rate. **(B)** Lorenz curves of the ten samples colored based on the WGA method. **(C)** Box plot of Gini coefficients of the Lorenz curves **(D)** Bos plot of Evenness score



Supplementary Figure 3. Quality control statistics of the lung cancer cells which passed QC **(A)** Scatterplot showing *p* as the probability of an allele being detected versus Allelic Drop Out rate. **(B)** Lorenz curves of the cells **(C)** Box plot of Gini coefficients of the Lorenz curves **(D)** Box plot of Evenness score. In each plot the bulk sample performance is included for comparison

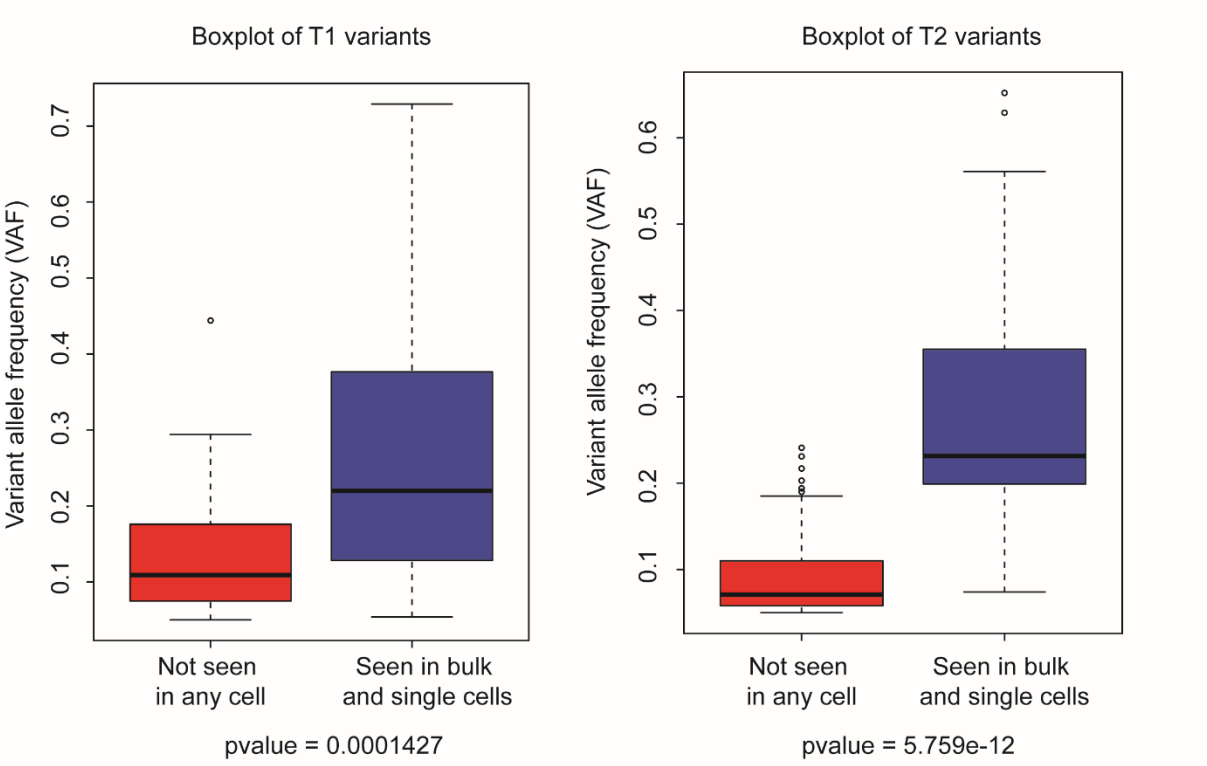


Supplementary Figure 4. Flow chart shows the series of filters applied for the removal of low quality variants. The number of variants that remain after applying each filtering criteria are indicated in the bottom right corner of each box. The downstream filtering of the SNVs is indicated in Fig. 2E.



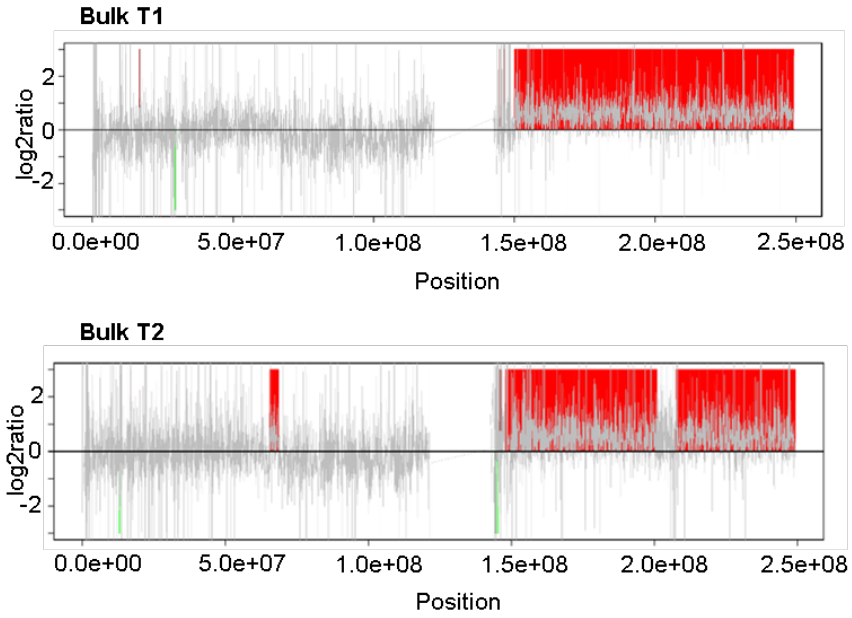
Supplementary Figure 5. Quality control of INDELs detected in single cells. **(A)** Contour plot used to determine the threshold for filtering of low quality INDELs. Red colour indicates region enriched for false positive variants, while blue indicates regions enriched for true positive variants. **(B)** The density plot shows the distribution of variant allele frequency (VAF) between true positives (TP) and false positives (FP). **(C)** The flow chart shows the sequence of serial filters applied to remove germline mutations. Numbers of variants that remained after each step are indicated on the right.

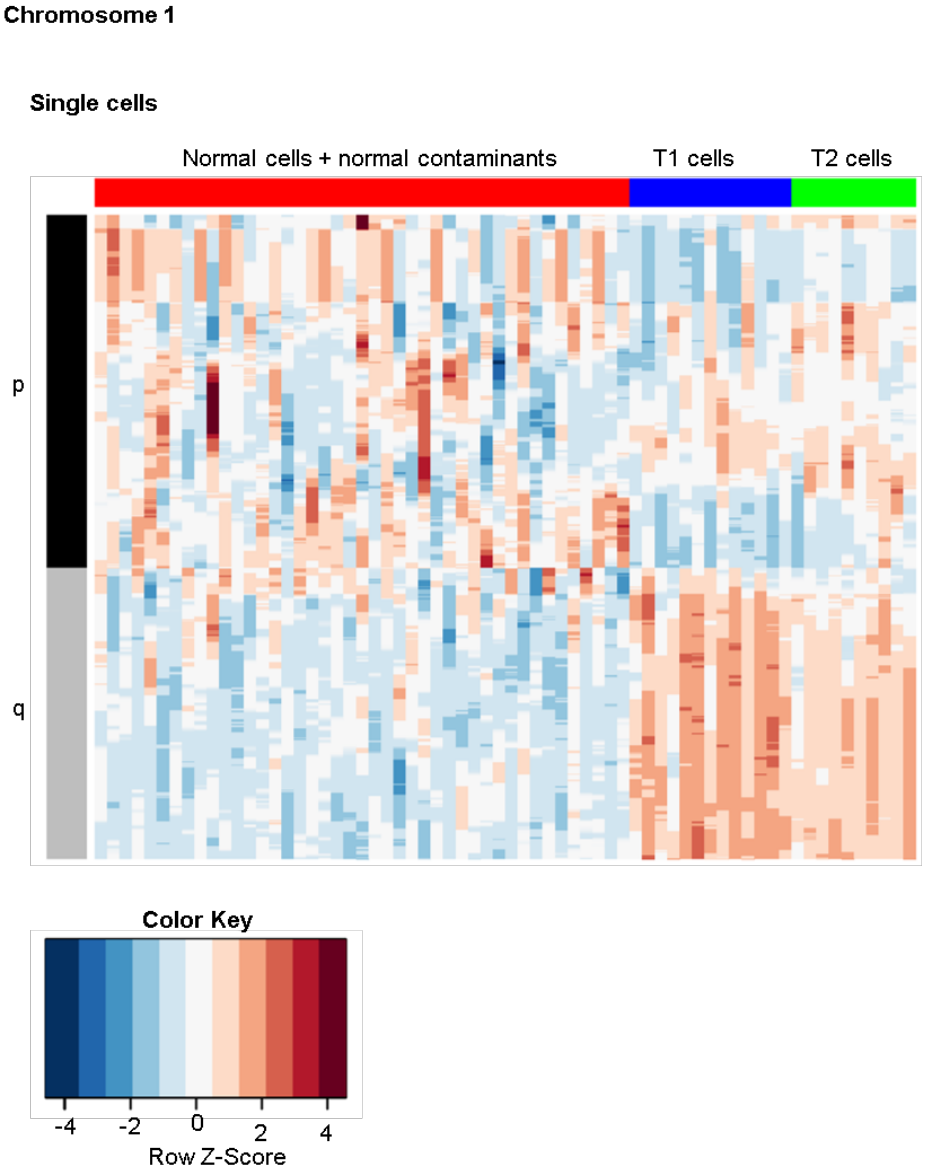
Supplementary Figure 6. Pie chart indicating the reasons why somatic point mutations detected in bulk were not identified in single cells. The inner circle shows the explanation of their presence or absence in single cells at a higher level. This information is further subcategorised in the outer circle and linked to the filters employed. The top plot shows the proportion of variants in T1, while the bottom plot shows the distribution of variants in T2. Majority of the variants in the bulk (T1 and T2) were either not observed or seen in too few cells (red colour). The likelihood of detecting these variants in the single cells can be improved by increasing the number of cells sequenced. Green colour indicates the proportion of variants that were detected in both, bulk and single cells. Blue colour indicates the proportion of variants missed in the single cell due to bioinformatics/experimental limitations. Purple colour indicates the false positives that were detected in the bulk, based on evidence in the bulk normal and normal single cells. The colours for each category are the same for both plots.



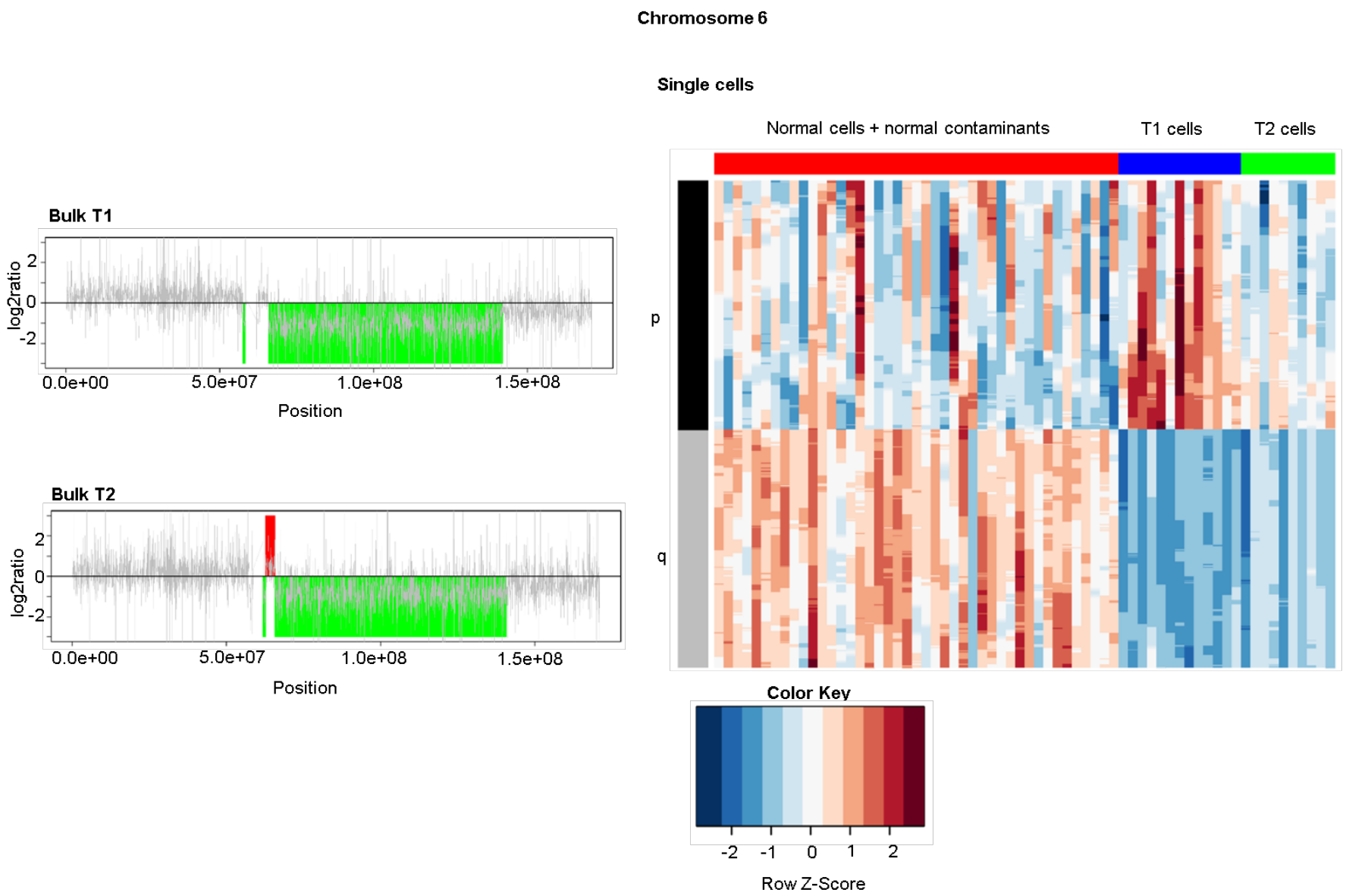
Supplementary Figure 7. Boxplot comparing the VAF of variants detected in bulk. The red colour boxplot shows the distribution of VAF of variants in bulk that were not seen in any single cells, while the blue colour boxplot shows the VAF of variants in bulk that were detected in both, bulk and single cells. In both T1 and T2 tumours, variants that were not seen in any single cells tend to have a lower VAF in the bulk compared to those that were observed.

A

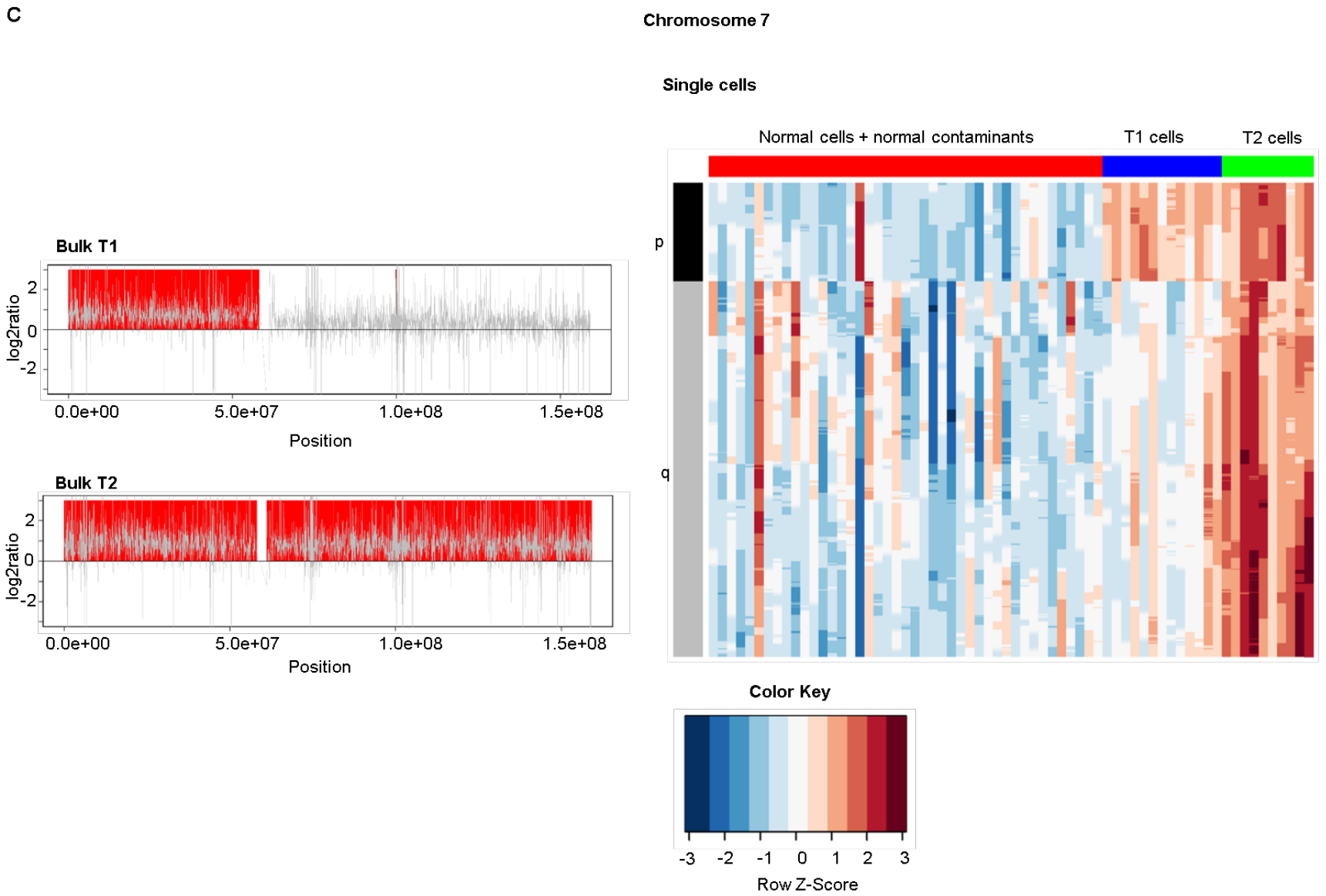

Bulk T1
log2ratio
2
0
-2
0.0e+00
5.0e+07
1.0e+08
1.5e+08
2.0e+08
2.5e+08
Position
Bulk T2
log2ratio
2
0
-2
0.0e+00
5.0e+07
1.0e+08
1.5e+08
2.0e+08
2.5e+08
Position


Chromosome 1
Single cells
Normal cells + normal contaminants
T1 cells
T2 cells
p
q
Color Key
-4
-2
0
2
4
Row Z-Score

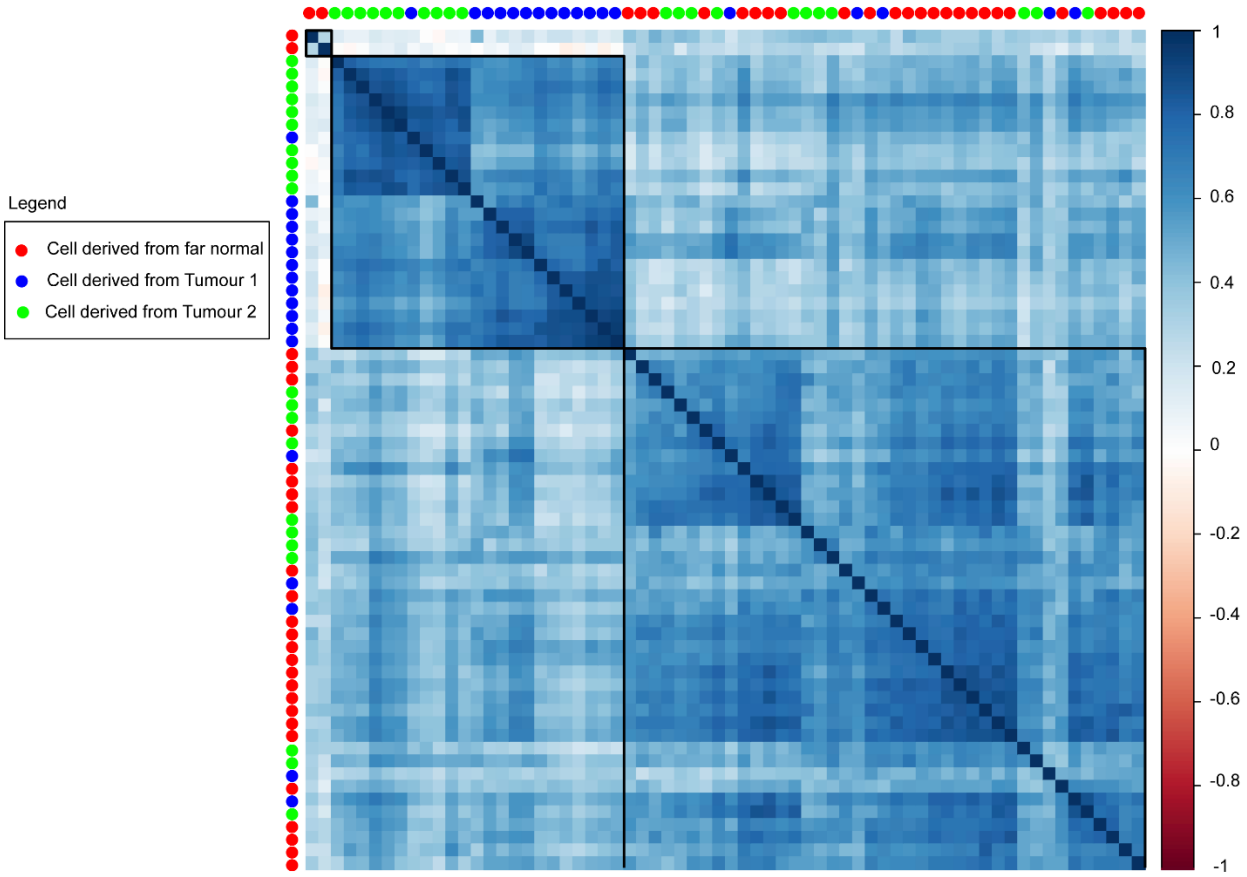
B


Chromosome 6
Single cells
Normal cells + normal contaminants
T1 cells
T2 cells
Bulk T1
log2ratio
2
0
-2
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Position
Bulk T2
log2ratio
2
0
-2
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Position
p
q
Color Key
-2
-1
0
1
2
Row Z-Score

C

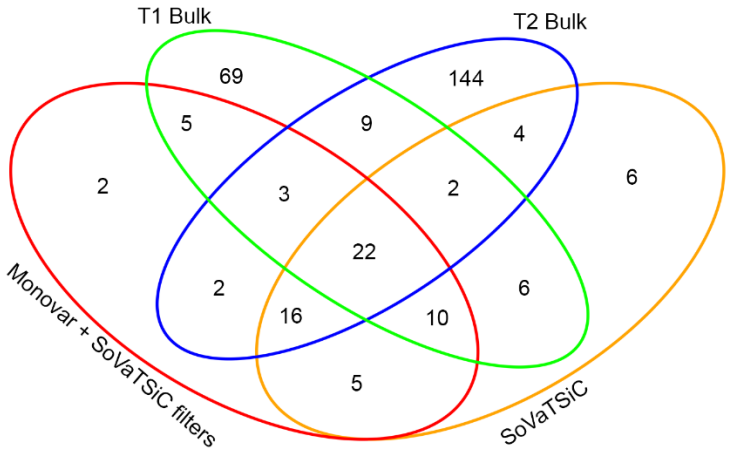

Chromosome 7
Single cells
Normal cells + normal contaminants
T1 cells
T2 cells
p
q
Bulk T1
log2ratio
2
0
-2
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Position
Bulk T2
log2ratio
2
0
-2
0.0e+00
5.0e+07
1.0e+08
1.5e+08
Position
Color Key
-3
-2
-1
0
1
2
3
Row Z-Score

Supplementary Figure 8. Examples of chromosomal copy number changes detected in bulk sectors and supported by single cell data are shown. The copy number changes observed in the bulk are shown on the left side of the figure with chromosome coordinates increasing from left to right (p arm left, q arm right), while the heatmap on the right shows the copy number profiles observed in single cells with chromosome coordinates increasing from top to bottom (p arm top, q arm bottom) and individual cells arranged from left to right corresponding to individual columns. **(A)** The q arm of chromosome 1 was amplified in both bulk sectors. This observation was also seen in cells derived from both tumour sectors but not in the normal cells. **(B)** Deletion of q arm in chromosome 6 was observed in both tumour sectors and supported in the single cells from both sectors as well. **(C)** Sector specific copy number changes were observed in chromosome 7. The p arm was amplified in tumour sector 1, while the entire chromosome 7 was amplified in tumour sector 2. The observation in bulk was supported in the single cells. Cells derived from tumour sector 1 had an amplification of the p arm, while cells derived from tumour sector 2 showed an amplification of the entire chromosome.

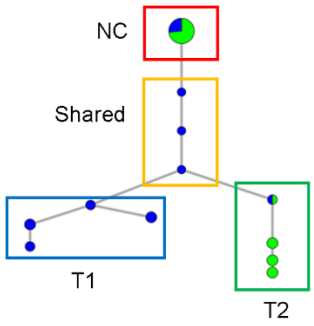


Supplementary Figure 9. Clustering of single cells using hierarchical clustering based on the Pearson correlation coefficient derived from the normalized coverage profiles.

A

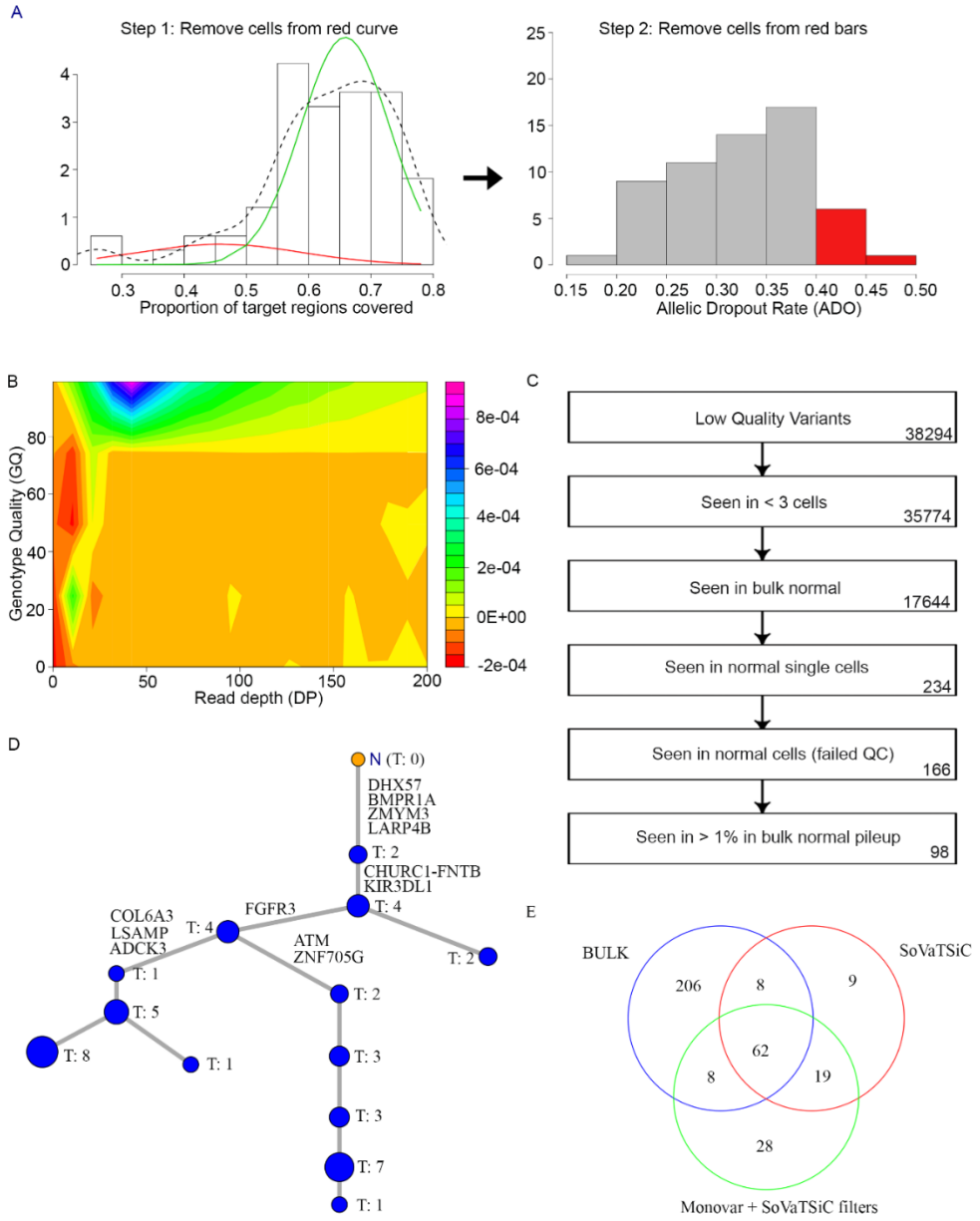


B



Consistency with phylogeny				Our pipeline (SoVaTSiC)			Monovar		
								With our filters	
NC	Shared	T1	T2	All variants	SNVs	INDELs	Monovar after recommendated filters (Unique to Monovar)	Monovar variants overlap with SoVatSiC	Monovar Only sites
Yes	Yes	Yes	Yes	0	0	0	286	0	0
Yes	Yes	Yes	No	0	0	0	22	0	0
Yes	Yes	No	Yes	0	0	0	41	0	0
Yes	No	Yes	Yes	0	0	0	236	0	0
Yes	Yes	No	No	0	0	0	31	0	0
Yes	No	Yes	No	1	0	1	96	0	0
Yes	No	No	Yes	2	0	2	227	0	0
Yes	No	No	No	0	0	0	213	0	1
No	Yes	Yes	Yes	26	22	4	14	16	0
No	Yes	Yes	No	8	7	1	25	2	4
No	Yes	No	Yes	0	0	0	14	0	0
No	Yes	No	No	0	0	0	11	0	0
No	No	Yes	Yes	8	6	2	43	8	3
No	No	Yes	No	12	12	0	58	9	2
No	No	No	Yes	26	24	2	72	18	2
No	No	No	No	0	0	0	70	0	0

Supplementary Figure 10. Comparison of somatic variants detected by Monovar and SoVaTSiC filters. **(A)** Venn diagram shows shared point mutations between bulk tumours and variants detected in single cells by SoVatSiC and Monovar. The variants detected by Monovar were obtained after applying the filters recommended by the Monovar authors as well as our filters. (**B**) Comparison of variants in single cells detected by SoVaTSiC and Monovar with the phylogenetic tree.



Supplementary Figure 11. Somatic variants detected using SoVaTSiC and Monovar on bladder cancer dataset. (**A**) Description of quality control steps for single cells. Gaussian Mixture Model (GMM) was used to cluster the single cells based on exonic coverage. The low coverage clusters were removed from further analysis (cells with coverage ≤ 0.5). In addition, cells were removed based on the allelic dropout rate (ADO) rate. (**B**) Contour plot used to determine the threshold for filtering of low quality INDELs. Red colour indicates region enriched for false positive variants, while blue indicates regions enriched for true positive variants. (**C**) The flow chart shows the sequence of serial filters applied to the SoVaTSiC calls to remove germline mutations. Numbers of variants that remained after each step are indicated on the right. (**D**) Phylogenetic tree based on SoVaTSiC mutation calls depicts the relationship between single cells derived from tumour tissue. The root of the tree (denoted by N) consists of putative normal contaminant cells. In this dataset, none of the tumour cells were determined to be normal contaminants. The tree shows that there are three different tumour clones present within the tumour, with two later clones (boxed by the blue and red rectangle) deriving from an earlier clone. The size of each node is proportional to the number of cells it represents, with the colour representing their source. (**E**) Venn diagram shows the shared somatic point mutations detected in the bulk tumour, tumour single cells detected by SoVaTSiC and Monovar.

Supplementary materials

Tube-bases whole genome amplification

For manual, tube-based whole genome amplification, the following kits were used according to the manufacturer's recommendations: illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences), MALBAC Single Cell WGA kit (Yikon Genomics), Repli-g single cell kit (Qiagen). For the Repli-g single cell kit, 8 hours of amplification time is recommended. In parallel experiments to the 8 hour amplification, 1.5 hour amplification time has been performed.

Detecting somatic CNVs from bulk lung cancer exome sequencing

Sequencing reads were counted and GC normalization was done using Excavator2's EXCAVATORDataPrepare.pl script. A bin size of 50,000 bp was used to partition the off-target regions. Somatic CNVs from bulk sequencing were detected using Excavator2 (D'Aurizio et al. 2016) with default parameters. CNV regions were annotated by PennCNV (Wang et al. 2007) to identify genes that were affected by copy number changes.

Detecting copy number variations from single cell lung cancer exome sequencing

In the single cells, sequencing reads within exonic target regions were counted and GC normalization was performed using Excavator2 EXCAVATORDataPrepare.pl script. To identify regions of copy number changes, we adopted the method by Patel *et al*. to detect copy number variations from single cell RNA-seq (Patel et al. 2014). The exonic target regions were sorted based on their chromosomal location and a moving average of 2001 exonic target regions was used to estimate the copy number changes per chromosome in each cell. The following formula was used to estimate the copy number for each region per chromosome in each cell:

$$Copy\ number\ at\ region\ i\ of\ cell\ k = \frac{\sum_{j=i-1000}^{i+1000} normalized\ read\ count\ of\ j}{2001} \qquad \text{equation S1}$$

Where i is the estimated average copy number change at target region i and j is the exonic region adjacent to i.

For each cell, a z-score is obtained per copy number region using the following formula:

$$Zscore\ at\ region\ i = \frac{copy\ number\ at\ region\ i\ of\ cell\ k - median\ copy\ number\ across\ all\ regions}{sd\ across\ all\ regions}$$ equation S2

The Euclidean distance between each cell was calculated using the dist() function in R and hierarchical clustering using the hclust() function in R was employed to cluster the single cells based on the copy number profiles. Three prominent clusters were observed (**Figure 3D**). To further validate this grouping, we calculated the pearson correlation between cells using the z-score and cluster the cells using hierarchical clustering. Using both methods, cells that were previously clustered at the root of the somatic variant tree were clustered together with normal cells (**Figure 3D and Supplementary Figure 9**).

Lastly, to validate this approach, we compared the copy number profiles observed to those detected in the bulk tumour sectors. The individual cells show general concordance with the sector of origin in both shared as well as sector specific changes (**Supplementary Fig 8A-C**).

Quality control of lung single cells after exome sequencing

Exome sequencing was carried out on 66 single cells from T1, 95 single cells from T2, and 39 single cells from the far normal following the C1-GE protocol. Low quality single cells were removed based on their allelic dropout (ADO) rate, false negative (FN) rate, and percentage of genome covered (≥ 5 reads).

The percentage of genome covered (≥ 5 reads) was computed using GATK DepthOfCoverage using mapped reads which have a minimum mapping quality of 20.

Prior to estimating the ADO and the FN rate for each cell, a set of consensus heterozygous germline variant site in the three bulk samples was used as the true variant set. These heterozygous sites had to fulfil the following criteria: (1) minimum depth of 8 (SNVs) and 5 (INDELs) in all tissues; (2) minimum genotype quality of 30 (SNVs) and 20 (INDELs) for all tissues; (3) each allele is covered by at least 3 reads in all tissues; (4) variant had a minimum variant allele frequency of 0.2 in all tissues.

The ADO rate per cell was calculated using the following formulas using only positions that were covered by at least 5 reads:

$$ADO = \frac{Number\ of\ heterozygous\ sites\ in\ bulk\ detected\ as\ homozygous\ in\ single\ cells}{Number\ of\ heterozygous\ sites\ in\ bulk}$$ equation S3

The FN rate per cells was calculated using the following formula using only positions that were covered by at least 5 reads:

$$FN = 1 - \frac{Number\ of\ heterozygous\ sites\ in\ bulk\ which\ have\ sufficient\ coverage\ in\ single\ cells}{Number\ of\ heterozygous\ sites\ in\ bulk}$$ equation S4

The percentage of genome covered (≥ 5 reads) was used as the first quality control criteria. The cells exhibited a bimodal distribution of coverage (**Figure 2B**). As such, a Gaussian mixed model (GMM) was applied twice to remove single cells based on the percentage of genome covered (≥ 5 reads). The GMM was generated using the mixtools R package (Benaglia et al. 2009). The single cells were separated into three groups, namely cells with coverage less than 10%, cells with coverage between 10% to < 42%, and cells with coverage greater or equal to 42%. Cells belonging to the first two categories were excluded from further analysis. Lastly, the remaining cells were removed if they hadFN rates greater than 0.45 or ADO rates greater than 0.35. This gave a total of 66 cells consisting of 27 far normal cells, 18 cells from T1and 21 cells from T2 were retained for further analysis

Variant detection from single cells exome sequencing

Variants from qualified tumour and normal single cells were detected via GATK haplotypeCaller version 3.5 using the following parameter: Mapping quality (MQ) ≥ 40, Base quality (BQ) ≥ 20. Joint genotyping and variant recalibration were performed on all tumour and normal cells together to produce a single VCF file containing all potential variant sites. Variant sites which passed the variant recalibration were retained. After variant recalibration, tri-allelic sites, singletons, and variants within 10bp of each other were removed to reduce the false positive rate. For SNVs, genotypes with read depth (DP) < 5, genotype quality (GQ) < 30, and variant allele frequency (VAF) < 0.15 were removed. Variant genotypes which failed the GQ filter were re-examined by comparing the difference in Phred likelihood score (PL) between the homozygous reference genotype and the maximum of heterozygous genotype and homozygous alt genotype. If the difference is greater than 30, the genotype will be retained. For INDELS,

genotypes with DP< 5, GQ < 40, and VAF < 0.2 were removed. The thresholds used for filtering of genotypes were determined by using variant calls in the three bulk samples. The variants detected in the bulk samples were used to predefine a truth set and a true negative sites set. The truth set was created using consensus variants in bulk normal and tumour sectors which satisfied the criteria (DP ≥ 8 and GQ ≥ 30 for SNVs, and DP ≥ 5 and GQ ≥ 20 for INDELs). The true negative sites sets were generated using sites whereby no variant was seen in all three bulk samples and were covered with sufficient coverage (DP ≥ 8 for SNVs and DP ≥ 5 for INDELs) in at least one bulk.

Using the two variant sets defined by the bulk samples, the variants identified in the single cells were used to form the true positives and false positives set. The true positives set was defined as variants detected in each single cell that were also seen in the truth set. On the other hand, the false positives set was defined as variants detected in the single cells that were found at the true negative sites.

Lastly, by comparing the density of three features namely, DP, GQ, and VAF between the true positives and the false positives set, a threshold was determined for each feature.

Putative somatic variants were filtered based on the following criteria: (I) variants were seen in less than 3 cells; (II) variants were detected in germline bulk normal tissue; (III) alternative allele was observed in more than one percent of total reads in germline bulk normal tissue pileup data; (IV) variants were detected in normal single cells (we require somatic variant sites to be homozygous reference for all normal cells and have at least 3 normal cells covered); (V) variants were seen in normal single cells which failed QC. The final somatic variants were annotated using ANNOVAR.

Phylogenetic analysis of lung cancer single cells

OncoNEM was used to infer the phylogeny between lung cancer single cells. In order to estimate the false positive rate (FPR) and false negative rate (FNR) based on the data, a maximum likelihood approach was used to identify the best combined parameter across a range of values for both false positives and false negatives. For the false positives estimation, we used a range of values from 0.01 to 0.15, while a range of values from 0.01 to 0.2 was used for the false negative estimation. Based on the maximum likelihood approach, a false positive rate of 0.025 and a false negative rate of 0.14 was estimated to give the highest scoring tree. Using the estimated FNR and FPR, the phylogenetic tree was estimated.

Detecting variants from bladder cancer dataset

Raw sequencing reads of single cells obtained from a muscle-invasive bladder cancer patient (Li et al. 2012) were downloaded from the NCBI short reads archive (SRA051489). The bulk sequencing of the tumour sector and adjacent normal tissue from the same patient were also downloaded. The downloaded raw reads (.sra) file were converted to FASTQ format using SRA Fastq-dump tool. BWA MEM version 0.7.10-r789 (Li 2013) with default parameters was used to align sequencing reads to the Human reference genome Hg19. Picard tool version 1.129 (Picard) was used to sort and mark duplicated reads. GATK version 3.5 (Van der Auwera et al. 2013) with default parameters was used to perform indel realignment and base recalibration to obtain the final bam files.

Germline variants (SNVs and INDELs) in the tumour sector and adjacent normal tissue were detected using GATK haplotypeCaller followed by hard filtering recommended by GATK best practices for both SNVs and INDELs. Germline SNVs were further filtered by removing variants which have DP < 8 or GQ < 30. For germline INDELs, we removed variants which have DP < 5 or GQ < 20. Putative somatic SNVs were detected by comparing bulk tumour samples with the adjacent normal tissue via MuTect using the default parameters (Cibulskis et al. 2013) and were annotated using ANNOVAR (Wang, Li, and Hakonarson 2010).

Variants in each single cells were detected using GATK haplotypeCaller with the following parameter (Mapping quality (MQ) $\geqslant$ 40, Base quality (BQ) $\geqslant$ 20). Joint genotyping was performed on all the tumour and normal cells together to produce a single VCF file containing all potential variant sites. Lastly, variant recalibration was performed to remove low quality variant sites. GATK variant recalibrator was used to filter the output at 99.9% sensitivity level for both SNVs and INDELs. Recalibration training databases used include dbSNP build 138, Omni 2.5M, 1000 genome phase 1 SNPs, Hapmap version 3.3, and Mills and 1000 genome gold standard INDELs. For SNVs, annotations used for recalibration training include variant quality score by read depth (QD), strand bias (FS), mapping quality rank sum score (MQRankSum), read position rank sum score (ReadPosRankSum), and mapping quality (MQ). For INDELS, the annotations used for recalibration training include variant quality score by read depth (QD), strand bias (FS), mapping quality rank sum score (MQRankSum), and read position rank sum score (ReadPosRankSum). Lastly, only variant sites which were indicated as PASS by the VariantRecalibrator were retained for further analysis.

Cell QC was performed according to the pipeline described for the lung cancer dataset. For the bladder cancer dataset, Gaussian mixture model was performed once using the percentage of genome covered with at least 5 reads. As a result, seven cells with percentage of genome covered less or equal to fifty percent were removed. In the second step of the cell QC, seven cells were removed due to high allelic dropout rate (ADO ≥ 0.50). This gives a total of 52 single cells, comprising of 9 normal cells and 43 tumour cells. Compared to the original analysis by Li *et al*. (Li et al. 2012) whereby 11 cells were removed, 14 cells were removed in this analysis.The differences in the number of cells removed could be due to the different parameters that were utilized for the QC steps.

Using the 52 good quality cells, joint genotyping and variant recalibration were performed to identify potential variant sites, and sites that passed the variant recalibration step were retained for further analysis. Using the sites, threshold to remove low quality genotypes were determined. For SNVs, we retained genotypes that have a minimum read depth (DP) of 5, minimum genotype quality (GQ) of 30, and minimum variant allele frequency (VAF) of 0.15. For INDELs, genotypes that have a minimum DP of 5, minimum GQ of 40, and minimum VAF of 0.2 were retained. For both SNVs and INDELs, variant genotypes which failed the GQ filter were re-examined by comparing the difference in phred likelihood score (PL) between the homozygous reference genotype and the maximum of the heterozygous genotype and homozygous alternate genotype. If the difference is greater than 30 for SNVs and 40 for INDELs, the genotype will be retained.

Lastly, to determine somatic variations from the bladder cancer dataset, the filters that were used for the lung cancer dataset were employed. After removing variants that were within 10 bp of each other, tri-allelic sites, and singletons sites, putative somatic variants were filtered based on the following criteria: (I) variants were seen in less than 3 cells after removing low quality genotypes; (II) variants were detected in germline bulk normal tissue; (III) variants were detected in normal single cells (We require variant site to be homozygous reference for all normal cells and at least 3 normal single cells were covered.); (IV) variants were seen in normal single cells which failed QC; (V) the alternate allele was observed in more than one percent of the total reads in germline bulk normal pileup data. With this, a total of 98 somatic SNVs and 53 somatic INDELs were detected. Phylogenetic analysis was performed using OncoNEM (Ross and Markowetz 2016) to determine the tumour evolution trajectory.

Lastly, to compare the performance of our pipeline with that of Monovar, the aligned bam files were used for variant calling using Monovar. Monovar was run with default parameters using sequencing reads with mapping quality (MQ) ≥ 40 and base alignment quality (BAQ) ≥ 20. Variants within 10bp of each other and singletons were removed to reduce the variant false positives rate. Based on the recommendations given by the authors, variants with read depth less than 10 were removed. For variants with 10 ≥ DP < 20, a minimum of 3 alternative reads were required. For variants with 20 ≥ DP < 100, a minimum variant allele frequency of 0.15 is required. Lastly, for variants with DP ≥ 100, a minimum variant allele frequency of 0.10 is needed. In addition, putative somatic variants were filtered based on germline bulk tissues. Lastly, to ensure that the results were comparable with those detected by GATK, we applied the following filters: (I) variants were seen in less than 3 cells; (III) alternate allele was observed in more than one percent of total reads in germline bulk normal tissue pileup data; (IV) variants were detected in normal single cells (require sites to be homozygous reference for all normal cells and have at least 3 normal cells covered); (V) variants were seen in normal single cells which failed QC.

Based on the comparison, Monovar detected 117 SNVs of which 70 (60%) overlap with the bulk tumour. On the other hand, SoVaTSiC detected 98 SNVs of which 70 (71.4%) overlap with the bulk (**Supplementary Figure 11E**). Of the 117 SNVs detected by Monovar, 28 were unique to Monovar. A close inspection of the 28 variants showed that 12 of them failed variant recalibration, 11 were not called by GATK, 1 was called as an INDEL, 3 were seen in less than 3 cells after filtering, and 1 was seen within 10bp of another variant.

Phylogenetic analysis of bladder cancer single cells

OncoNEM was used to infer the phylogeny between bladder cancer single cells. In order to estimate the false positive rate (FPR) and false negative rate (FNR) based on the data, a maximum likelihood approach was used to identify the best combined parameter across a range of values for both false positives and false negatives. For the false positives estimation, we used a range of values from 0.01 to 0.15, while a range of values from 0.01 to 0.2 was used for the false negative estimation. Based on the maximum likelihood approach, a false positive rate of 0.1 and a false negative rate of 0.09 was estimated to give the highest scoring tree and these parameters were used to estimate the phylogenetic tree.

Statistical analyses

All statistical analyses were done using Microsoft Office Excel or R. Two sample t-test was used to compare the performance of different kits. Two sample t-test was also applied to compare the number of mutations observed in cells derived from tumour sector 1 and tumour sector 2.

All p-values reported are based on the two-tailed tests.

References

Benaglia, Tatiana, Didier Chauveau, David Hunter, and Derek Young. 2009. 'mixtools: An R package for analyzing finite mixture models', *Journal of Statistical Software*, 32: 1-29.

Cibulskis, Kristian, Michael S Lawrence, Scott L Carter, Andrey Sivachenko, David Jaffe, Carrie Sougnez, Stacey Gabriel, Matthew Meyerson, Eric S Lander, and Gad Getz. 2013. 'Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples', *Nat Biotechnol*, 31: 213-19.

D'Aurizio, Romina, Tommaso Pippucci, Lorenzo Tattini, Betti Giusti, Marco Pellegrini, and Alberto Magi. 2016. 'Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2', *Nucleic Acids Res*, 44: e154-e54.

Li, Heng. 2013. 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', *arXiv preprint arXiv:1303.3997*.

Li, Y., X. Xu, L. Song, Y. Hou, Z. Li, S. Tsang, F. Li, K. M. Im, K. Wu, H. Wu, X. Ye, G. Li, L. Wang, B. Zhang, J. Liang, W. Xie, R. Wu, H. Jiang, X. Liu, C. Yu, H. Zheng, M. Jian, L. Nie, L. Wan, M. Shi, X. Sun, A. Tang, G. Guo, Y. Gui, Z. Cai, J. Li, W. Wang, Z. Lu, X. Zhang, L. Bolund, K. Kristiansen, J. Wang, H. Yang, M. Dean, and J. Wang. 2012. 'Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer', *GigaScience*, 1: 12.

Patel, Anoop P, Itay Tirosh, John J Trombetta, Alex K Shalek, Shawn M Gillespie, Hiroaki Wakimoto, Daniel P Cahill, Brian V Nahed, William T Curry, and Robert L Martuza. 2014. 'Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma', *Science*, 344: 1396-401.

Picard. 'Picard Tools', Accessed 23 February http://broadinstitute.github.io/picard.

Ross, Edith M, and Florian Markowetz. 2016. 'OncoNEM: inferring tumor evolution from single-cell sequencing data', *Genome Biol*, 17: 69.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. 'From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline', *Curr Protoc Bioinformatics*, 11: 11.10.1-11.10.33.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S. F. A. Grant, H. Hakonarson, and M. Bucan. 2007. 'PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data', *Genome Res*, 17: 1665-74.

Wang, K., M. Li, and H. Hakonarson. 2010. 'ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data', *Nucleic Acids Res*, 38: e164.