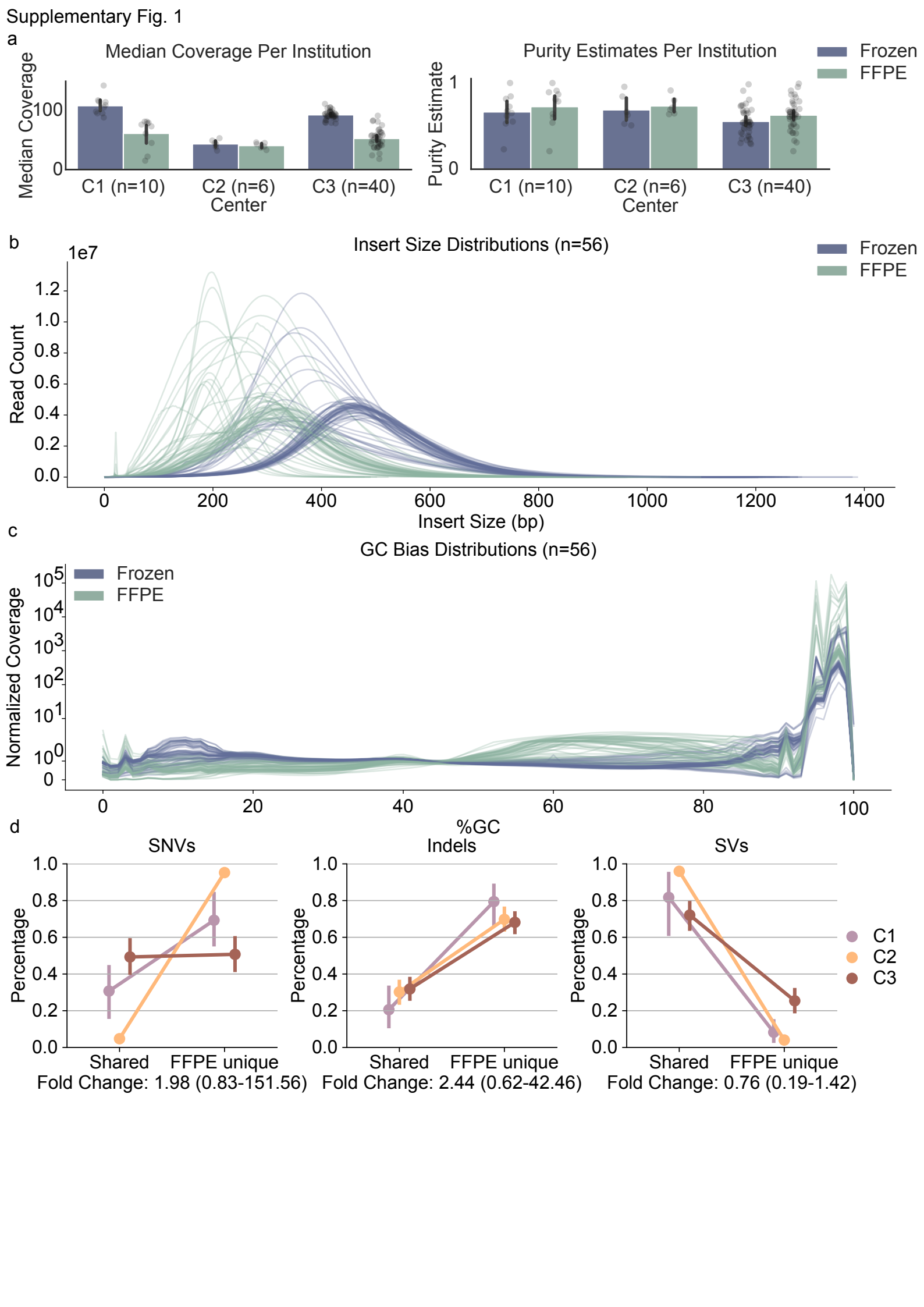
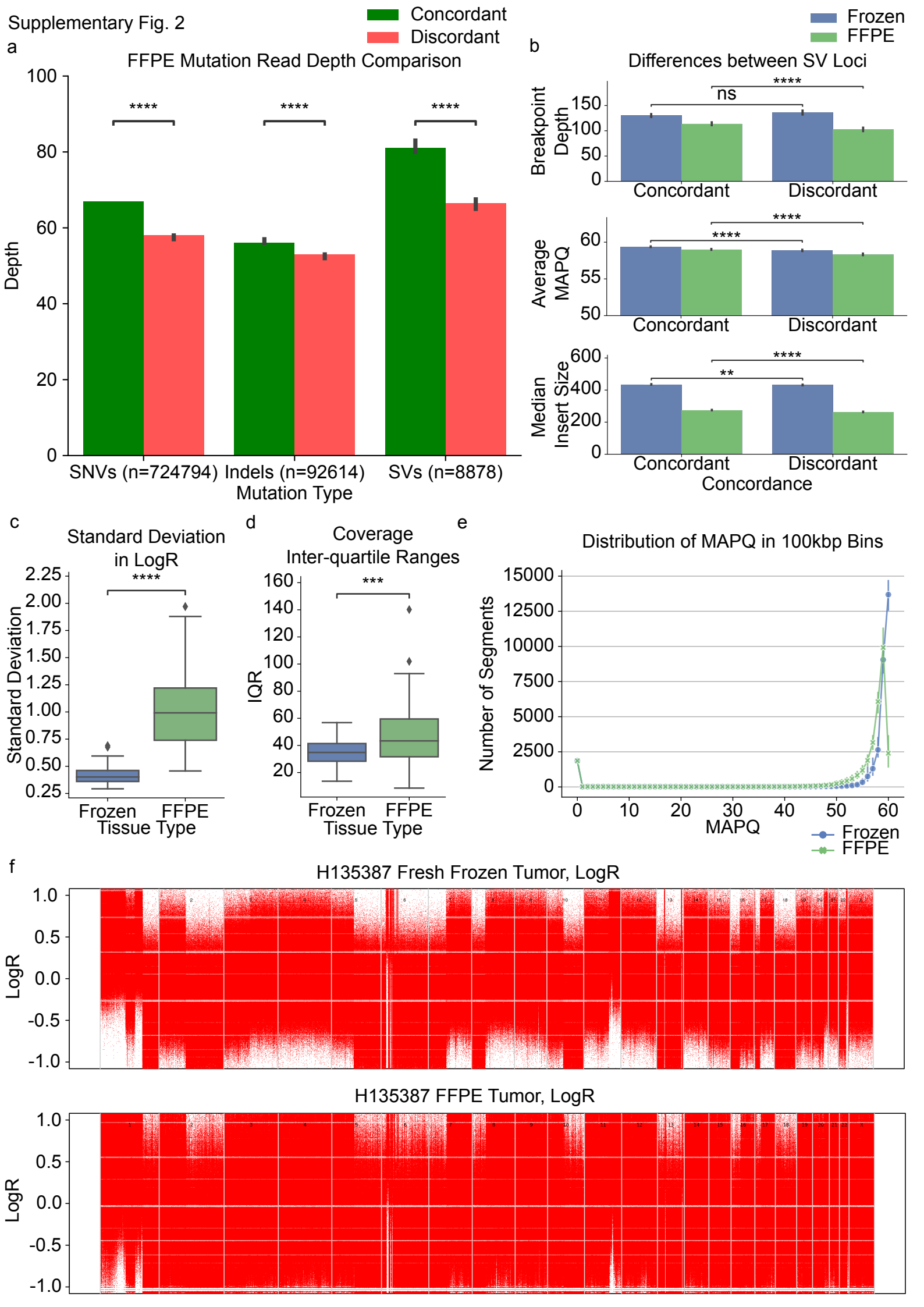
Supplementary Fig. 1



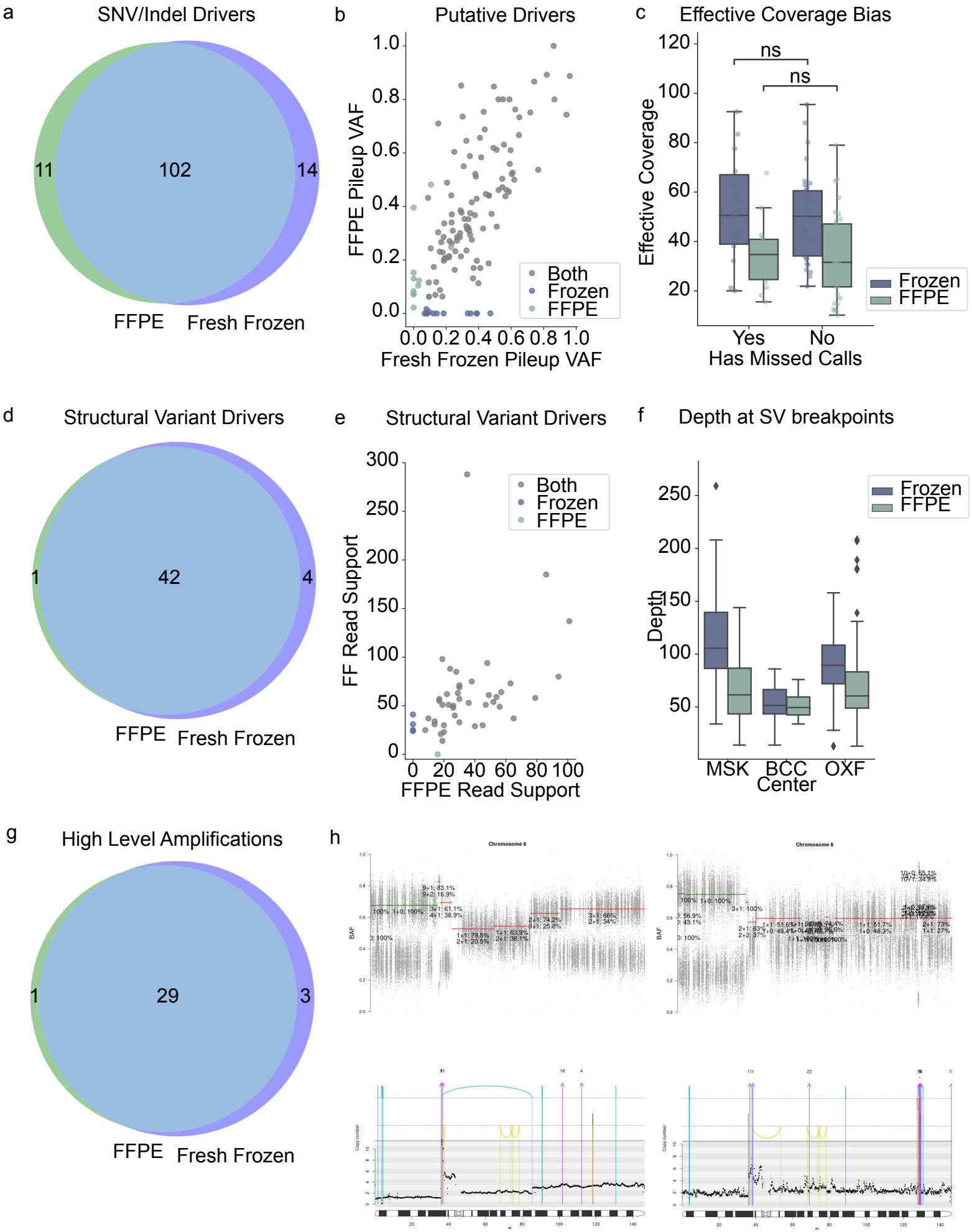
Supplementary Figure 1: Comparative Sequencing Results. a. Left barplot showing median coverage per matched pair from each center (C1: Memorial Sloan Kettering Cancer Center, n=10, C2: BC Cancer Agency, n=6, C3: Oxford University Hospitals, n=40) grouped by source material (FF or FFPE) (Mann-Whitney U test, p=1.889e13). Right barplot showing tumor purity estimation per sample (Mann-Whitney U test, p=4.533e-02). **b.** Kernel density plot showing insert size distributions for all matched FF/FFPE samples. **c.** Kernel density plot showing GC bias distributions. **d.** Pointplots demonstrating mutations shared between matched FF and FFPE versus FFPE unique calls for SNVs, indels, and SVs. FF, Fresh Frozen; FFPE, Formalin-fixed, paraffin-embedded; SNV, single nucleotide variant; indel, small insertion or deletion; SV, structural variant.

Supplementary Fig. 2



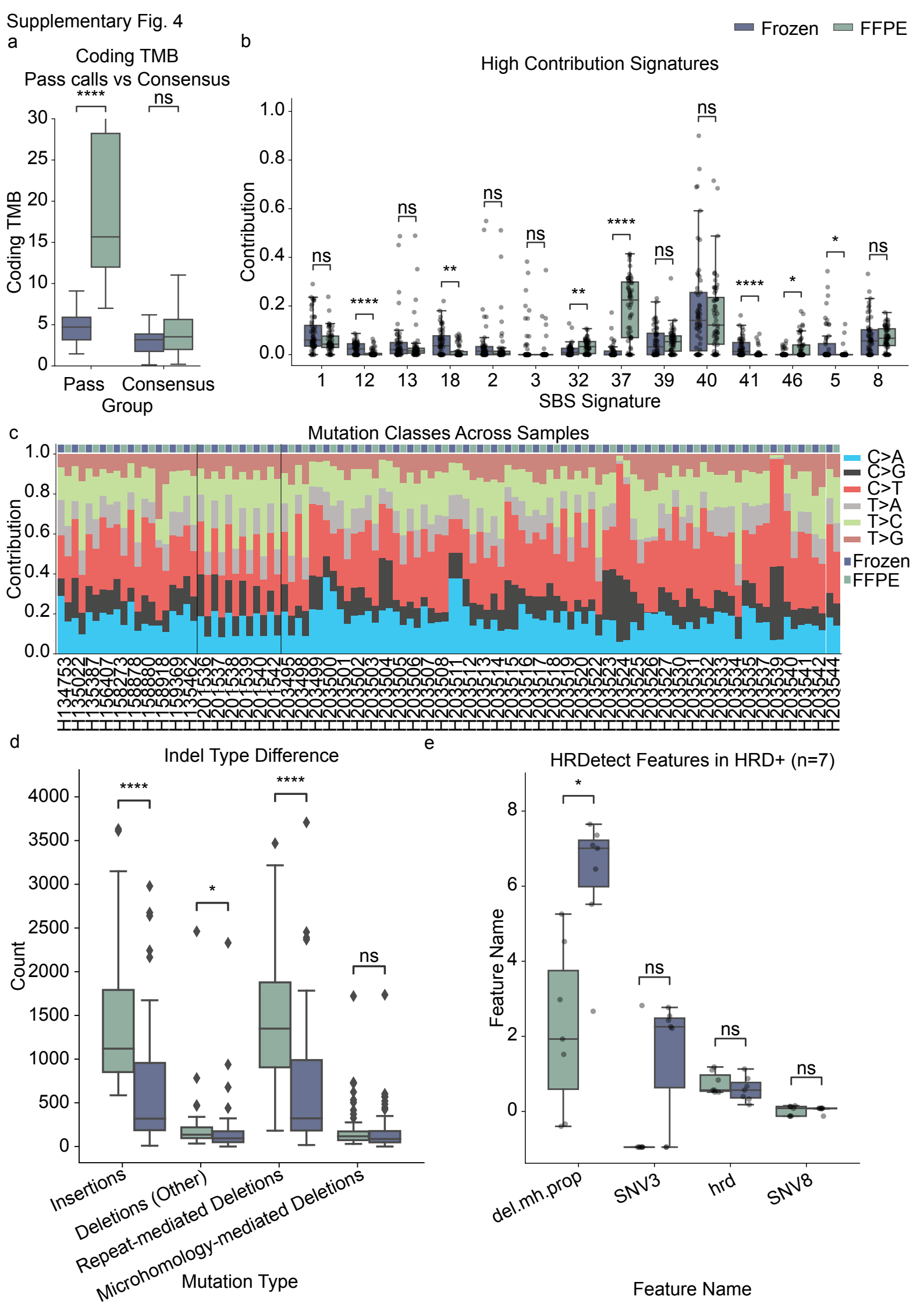
Supplementary Figure 2: Coverage Related Inconsistencies. a. Barplot showing distributions of loci depth in FFPE for each mutation class concordant between FF and FFPE versus discordant (FF only) excluding hypermutator samples (two-sided Mann–Whitney U test, ns = 5e-02 < p <= 1, * = 1e-02 < p <= 5e-02, ** = 1e-03 < p <= 1e-02, *** = 1e-04 < p <= 1e-03, **** = p < 1e-04). **b.** Barplots showing differences between concordant/discordant SV loci in breakpoint depth, average mapping quality and median insert size between FF and FFPE. **c.** Boxplot showing standard deviations in Log2 ratio (LogR) between FF and FFPE samples (FF median: 0.99, range: 0.46-1.97, FFPE median: 0.40, range: 0.29-0.69, Mann-Whitney U test, p=2.041e-18). **d.** Boxplot showing interquartile range distribution for coverage across all FF and FFPE samples. (Mann-Whitney U test, p=9.888e-04) **e.** Line plot showing median number of segments per sample at each mapping quality across 100 kbp segments for both FF and FFPE. **f.** Example LogR plot for FF and FFPE sample from patient H135387. FF, fresh frozen; FFPE, formalin-fixed, paraffin-embedded; SV, structural variant.

Supplementary Fig. 3



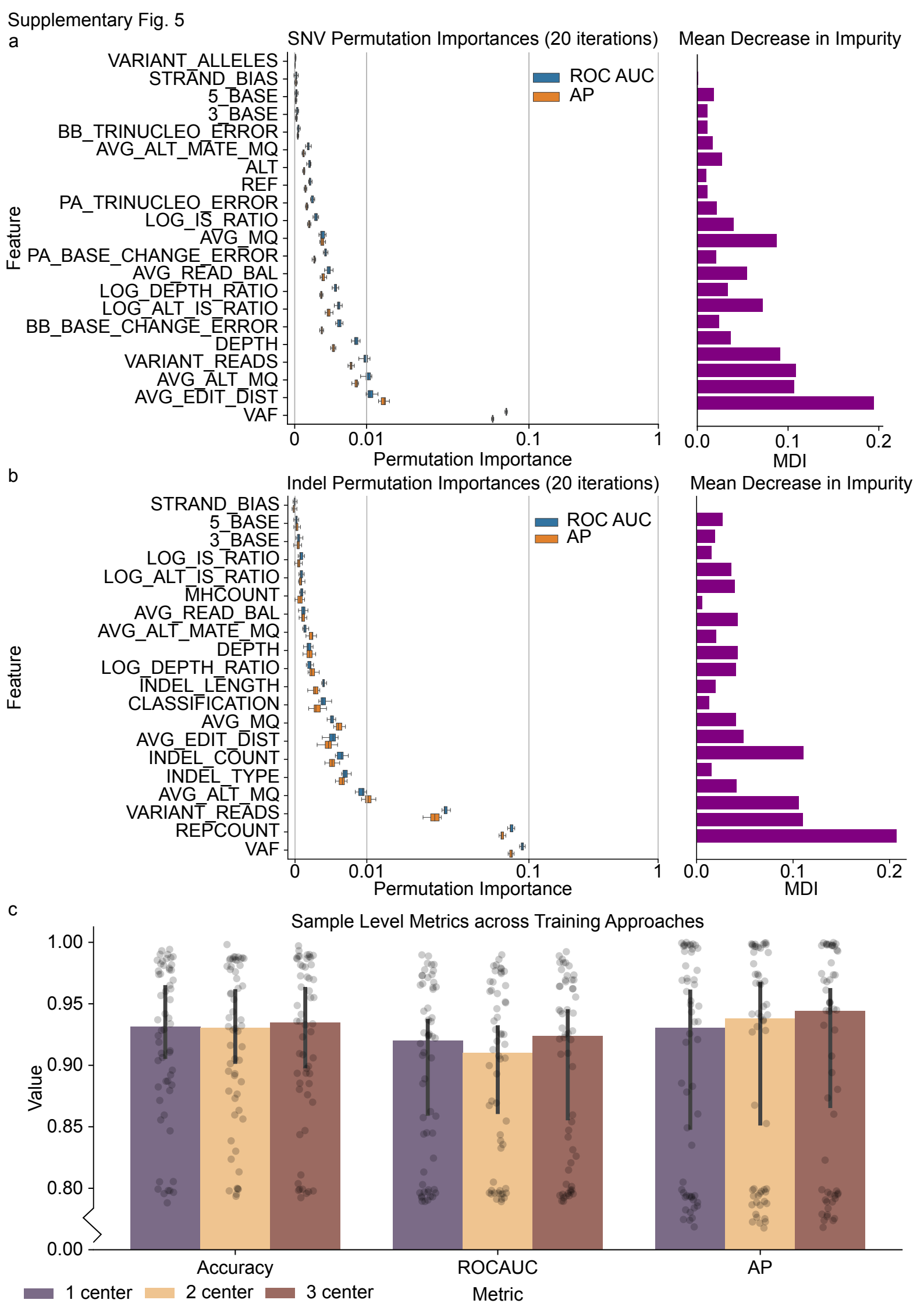
Supplementary Figure 3: Driver Alterations. a. Venn diagram showing overlap in called SNV/indel drivers. **b.** Scatterplot showing pileup VAF in FF and FFPE for all SNV/indel drivers colored by whether they were called de novo in FF, FFPE or both. **c.** Boxplot displaying WGS effective coverage (purity x coverage) for samples with discordant SNV/indel driver mutations (unique to FFPE or unique to FF, n=14) and samples with all driver mutations shared (n=33). NS, not significant. **d.** Venn diagram showing overlap in called SV drivers. **e.** Scatterplot showing read support in FF and FFPE for all SV drivers colored by whether they were called de novo in FF, FFPE or both. **f.** Boxplot showing differences in loci depths between FF and FFPE for driver SVs per each center. **g.** Venn diagram showing overlap in called high level amplification drivers. **h.** Example of a high level MYC amplification identified in the FFPE sample for a squamous cell carcinoma patient (H203504) that was not present in the matched FF likely due to ITH. SNV, single nucleotide variant; indel, insertion/deletion; VAF, variant allele frequency; FF, fresh frozen; FFPE, formalin-fixed, paraffin-embedded. WGS, whole genome sequencing; SV, structural variant.

Supplementary Fig. 4



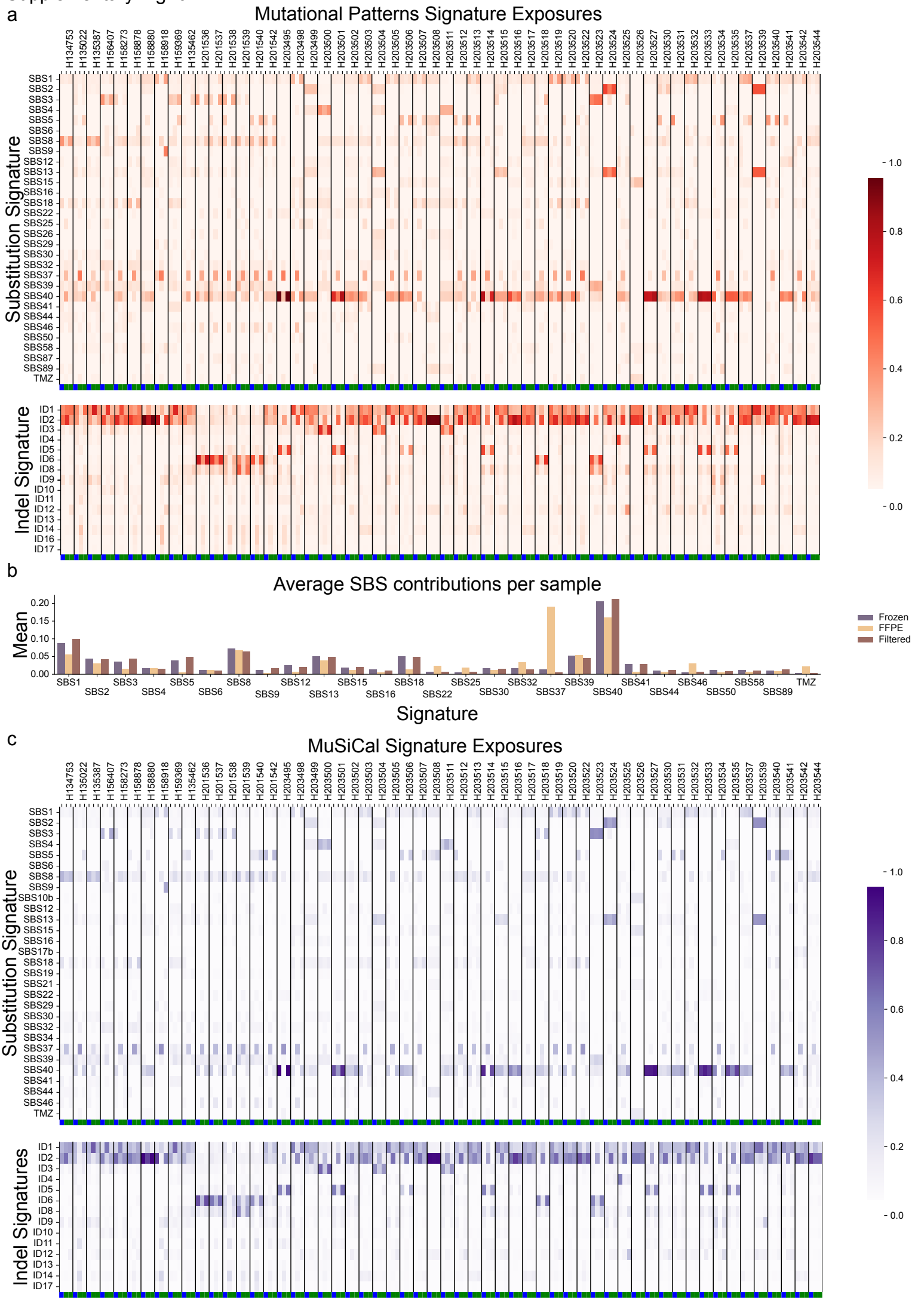
Supplementary Figure 4: Signatures of FFPE damage. a. Boxplot showing difference in distributions of coding TMB in all pass calls from any caller versus high confidence consensus calls in FF vs. FFPE. (two-sided Mann-Whitney U test, all pass calls: p=4.619e-13, consensus calls: p=3.968e-01) **b.** Boxplot showing difference between matched FF/FFPE in contributions of most prominent SBS signatures within the cohort with statistical annotations between FF/FFPE for each signature (two-sided Mann–Whitney U test, ns = 5e-02 < p <= 1, * = 1e-02 < p <= 5e-02, ** = 1e-03 < p <= 1e-02, *** = 1e-04 < p <= 1e-03, **** = p < 1e-04). **c.** Stacked barplot demonstrating breakdown of substitution changes per sample. Each sample has a bar for FF sample and a bar for matched FFPE denoted by the blocks at the top of the plot. **d.** Boxplot showing distribution of indel types between matched FF/FFPE (two-sided Mann–Whitney U test, Insertions: p=1.978e-07, Deletions (Other): p=1.293e-02, Repeat-mediated Deletions: p=3.246e-07, Microhomology-mediated Deletions: p=4.007e-01). **e.** Matched FF/FFPE distribution comparison for impacted HRDetect features for 7 cases identified as HRD positive from FF scores. TMB, tumor mutational burden, FF, fresh frozen; FFPE, formalin-fixed, paraffin-embedded; SBS, single base substitution; indel, insertion/deletion; HRD, homologous recombination deficiency.

Supplementary Fig. 5



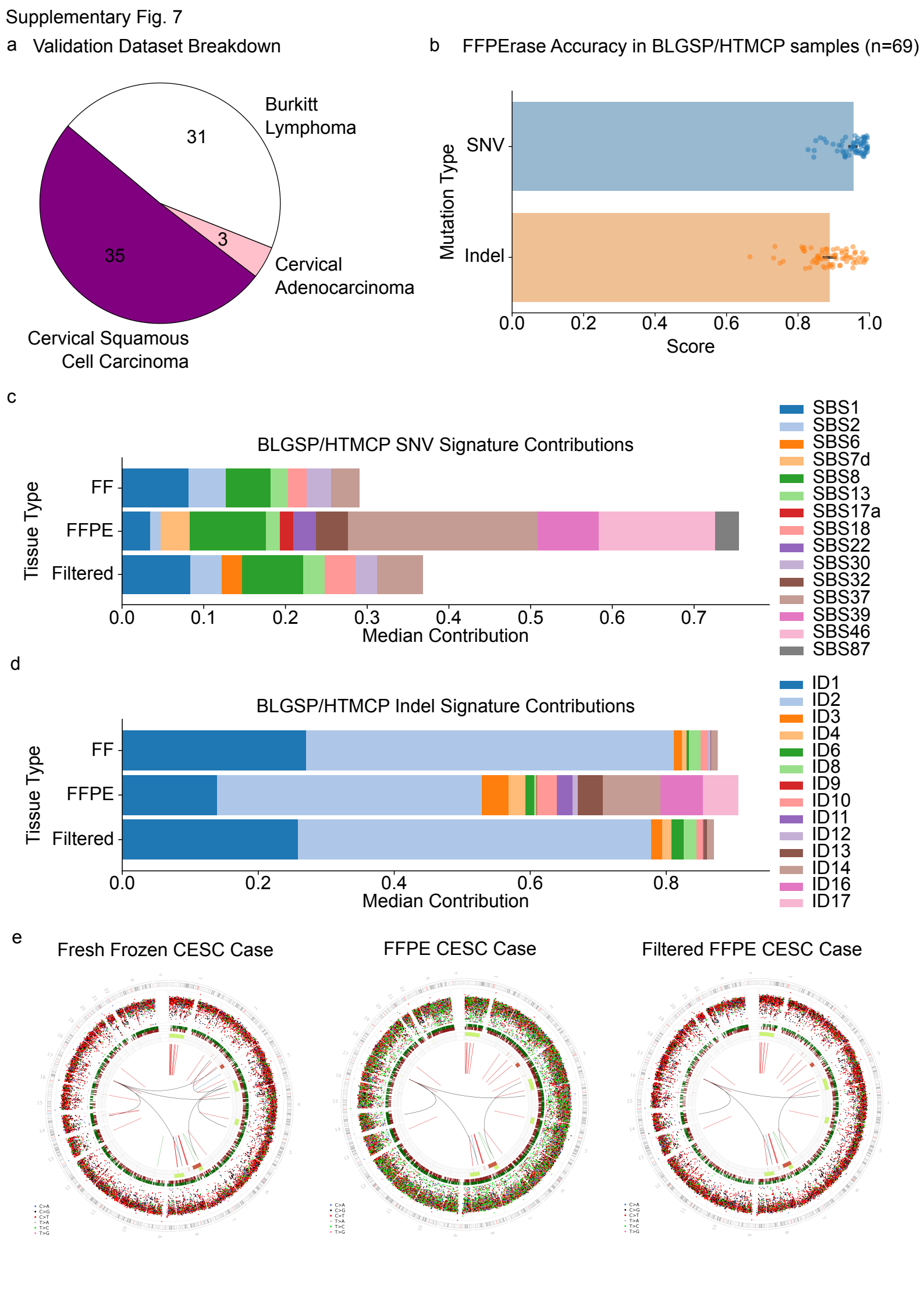
Supplementary Figure 5: Classifier Benchmarking. a. On the left, boxplots showing permutation importance calculated for 20 random iterations for each SNV feature. On the right, barplots showing mean decrease in impurity calculated for the same features. **b.** Same for indel features. **c.** Barplots showing performance metrics for each training/testing strategy per sample across the cohort from 0-1.

Supplementary Fig. 6



Supplementary Figure 6: Post Filtration Signatures. a. Signature exposures generated with MutationalPatterns. Top heatmap COSMIC SBS contributions with three bars per sample denoting first the FF, followed by the FFPE and then the filtered FFPE across all signatures present**.** Bottom heatmap same as described but for COSMIC ID contributions. **b.** Barplots showing difference in SBS contributions for top signatures throughout the cohort in the FF, unfiltered FFPE, and post filtered FFPE. **c.** Same as a. but with signature exposures generated with MuSiCal.

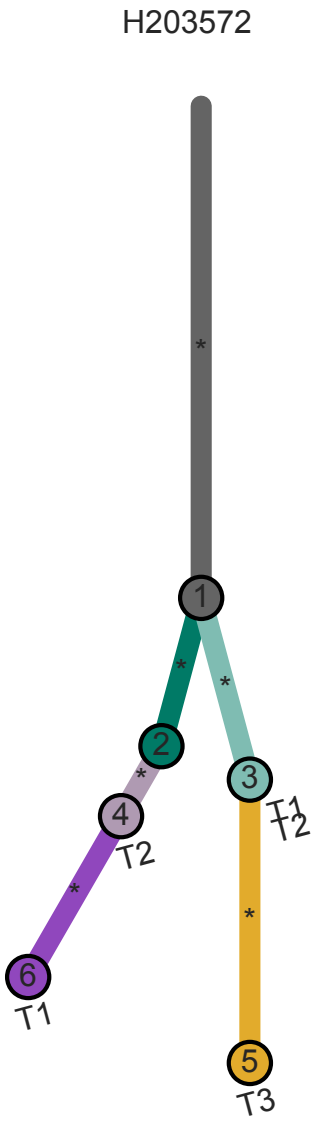
Supplementary Fig. 7



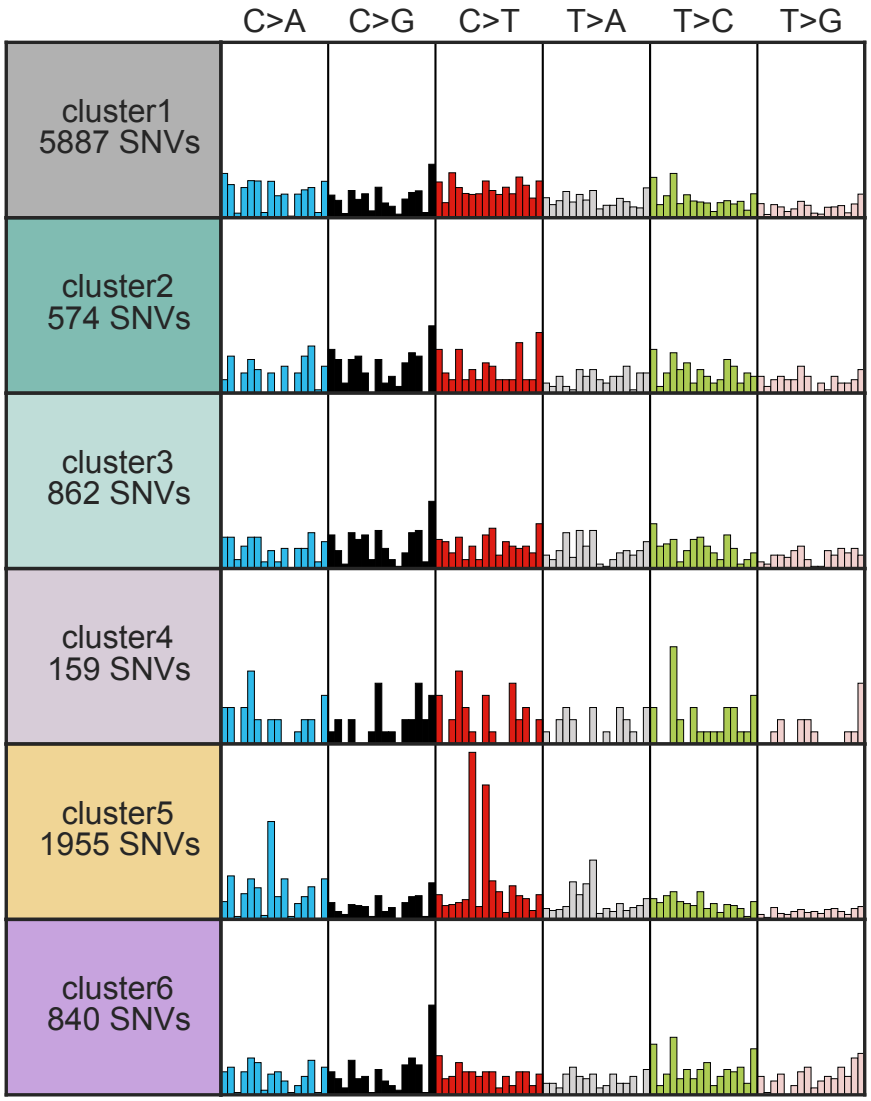
Supplementary Figure 7: FFPolish Validation Dataset. a. Pie chart showing disease breakdown for validation set. **b.** Barplot showing accuracy in artifact detection from FFPErase on matched trios. **c.** Stacked bar plot showing median contribution percentage of highest contribution signatures in SNVs for FF, FFPE, and filtered FFPE in validation cohort. **d.** Same as c. for indel signatures. **e.** Example circos plots as described in Figure 2 for fresh frozen, FFPE, and post-filtered by FFPErase FFPE samples from a Cervical Squamous Cell Carcinoma case which presented with increased artifact burden. FFPE, formalin-fixed, paraffin-embedded. SNV, single nucleotide variant; BLGSP, Burkitt Lymphoma Genome Sequencing Project; HTMCP, Human Tumour Molecular Characterization Project.

Supplementary Fig. 8

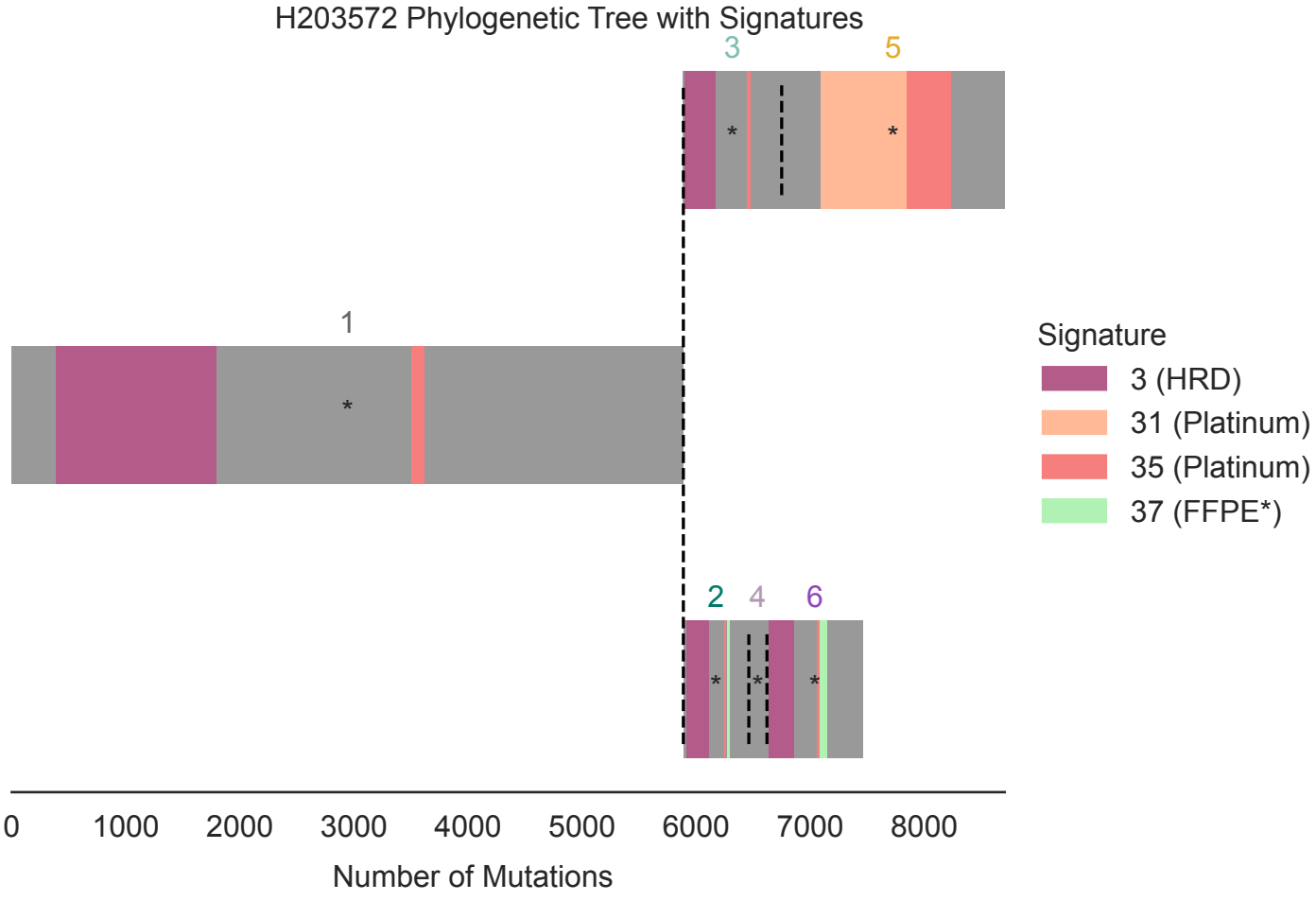
a



b



c



Supplementary Figure 8: Post Filtration Comparisons. a. Individual-level clonality analysis for H203572 post filtration of FFPE post-therapy sample. Left shows phylogenetic tree representation of branches designated as SNV clusters with branch length associated with total number of SNVs. Dark grey branch designates the trunk represented by the mutations present at a CCF of 100% in all samples. Terminal branches are annotated with the sample(s) where the corresponding clone is found. **b.** Clone-level mutational signature analysis for post-filtered clonality analysis as described in Fig. 6. **c.** Phylogenetic tree for post-filtered clonality analysis as described in Fig. 6. FFPE, formalin-fixed, paraffin-embedded. SNV, single nucleotide variant; CCF, cancer cell fraction.

Supplementary Notes

Identification of putative oncogenic events

Substitutions and indels were annotated according to their presence in OncoKB and COSMIC databases(Chakravarty et al. 2016; Forbes et al. 2017). Candidate oncogenic variants were identified as those annotated to be oncogenic/putative oncogenic in OncoKB and/or present in at least 6 tumors in COSMIC version 92.

Notes on Driver Mutations

High-confidence SVs were designated as those that were passed by at least 2 callers and were either translocations or greater than 600 base pairs in length.

Of the 117 SNV/indel drivers called in FF (median variant allele frequency (VAF): 34%, range: 7-96%), 103 (88%) were detected in FFPE (median VAF: 38%, range: 6-100%) outperforming the 83% genome-wide concordance for small mutations. Coverage across all FF calls averaged at 85x in FF and 65x in FFPE. There were 25 discordant variants: 14 FF-specific (median VAF: 13%, range: 7%-47%) and 11 FFPE-specific (median VAF: 13%, range: 2%-50%). 11 FF-specific and 7 FFPE-specific calls had no read support in the corresponding sample (Supplementary Fig. 3a-b, Supplementary Table 4) suggesting intratumor heterogeneity(Shukla et al. 2022) (Supplementary Fig. 3c) and necessitating validation experiments, not possible in this study.

Of 46 driver SVs in FF, 42 (91%) were detected in FFPE (Supplementary Fig. 3d-f, Supplementary Table 5). 1/4 FF-specific calls (NF1 deletion) and 1 FFPE-specific BRCA2 duplication had no read support in the matched counterpart. Importantly, the BRCA2 event was found in an HRD+ breast cancer patient and was likely missed in FF owing to lower tumor purity (30%/77% in FF/FFPE).

Of the 32 oncogene amplifications identified in FF, 29 (91%) were captured in FFPE (Supplementary Fig. 3g, Supplementary Table 5). Four discordant calls were found, notably an FFPE-specific MYC amplification in a squamous cell carcinoma (H203504) unobservable

in FF (Supplementary Fig. 3h). Conversely, 3 FF-specific calls were filtered in FFPE due to noise, but confirmed present through manual review.

Notes on Artifact Classifier Benchmark

Several strategies were explored to determine the best model architecture for detection of artifact mutations derived from FFPE. One being the random forest architecture that was ultimately selected for the final model, another being a convolutional neural network.

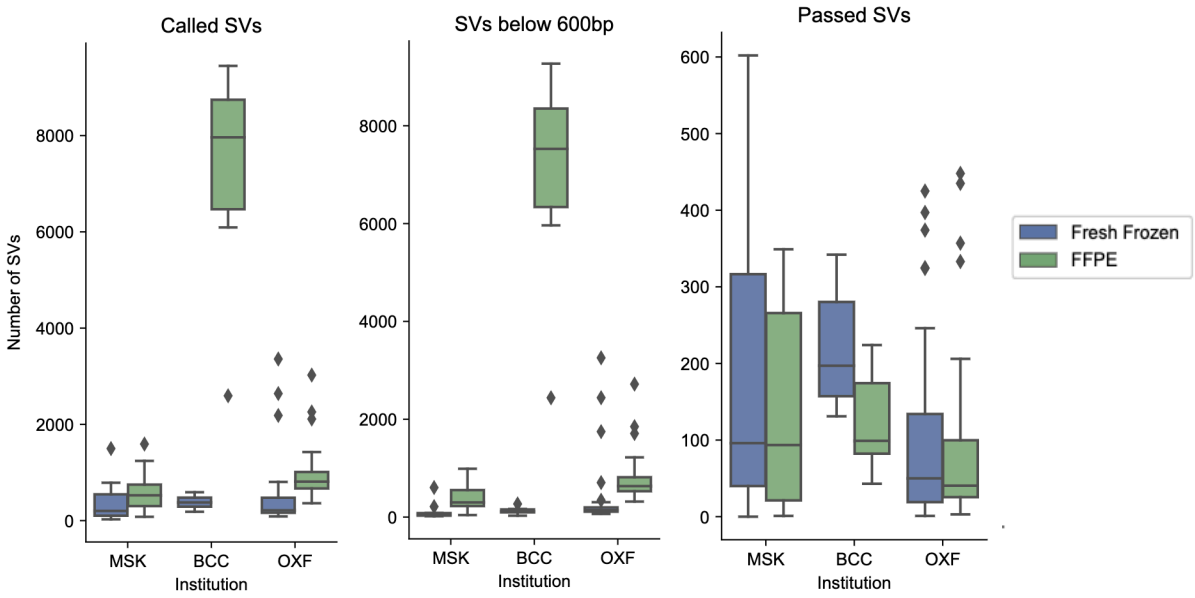
The random forest architecture was implemented using using scikit-learn (v0.24.1, https://scikit-learn.org/0.24/) and imbalanced-learn (v.0.8.0, https://imbalanced-learn.org/stable/). It is built with a default of 100 estimators and leverages imbalanced-learn's BalancedRandomForestClassifier module which internally undersamples the majority class in each bootstrap sample. This was necessary as in many FFPE samples, assessment of artifact mutations on the basis of overlap with matched FF demonstrates drastically larger totals of artifact mutations over shared ones.

The convolutional neural network approach utilized Google's DeepVariant(Poplin et al. 2018) (v1.2.0, https://github.com/google/deepvariant) and is detailed in the ICML 2023 Computational Biology workshop paper 116. (Domenico & Asimomitis et al. 2023)

Although model architectures achieved comparable results, the balanced random forest implementation was ultimately chosen due to less computationally expensive data preprocessing and model complexity.

Notes on low confidence structural variant enrichment

We observed an increased number of structural variants specific to a single caller with an enrichment for SV events <600 bp in length (median 5x increase (-0.8-266.5)), highlighting an increase in fragmented calls and the need for a consensus calling approach.



Notes on comparison to MSK-IMPACT

For the 23 cases with FFPE-derived WGS data and matched MSK-IMPACT targeted data available, 62 mutations of clinical relevance were identified between the assays (Supplementary Table 13). Of these, 58 were called de novo by both assays while 1 was specific to MSK-IMPACT and 3 were specific to WGS. The single call by MSK-IMPACT was present in WGS (VAF: 0.24) but filtered during post-processing on the basis of having >=3 mutant alleles in at least 1% of samples within a panel of 100 unmatched blood normal (https://github.com/cancerit/cgpCaVEManPostProcessing). Of the 3 WGS specific calls, two were designated as artifact calls with scores of 0.509 and 0.963. The threshold for which we determined artifact mutations was 0.50 or higher, showing that the confidence for the first call being an artifact was on the lower side of this range and potentially suggesting an indeterminate call.

Notes on comparison to other classifiers

Two other established tools for FFPE artifact clean up in WGS data were tested against our method, FFPolish and FFPESig.(Dodani et al. 2022; Guo et al. 2022) Additionally an adapted

version of FFPEimpact, a scoring tool which derives a damage score for a sample on the basis of FFPE signatures, was added to the comparison.(Basyuni et al. 2024)

FFPolish, a SNV artifact classifier, was tested out of the box using the model and instructions provided here: https://github.com/morinlab/FFPolish. We compared the results of our 3-center approach to the results of this model's classification of all FFPE samples in our cohort using matched FF as ground truth.

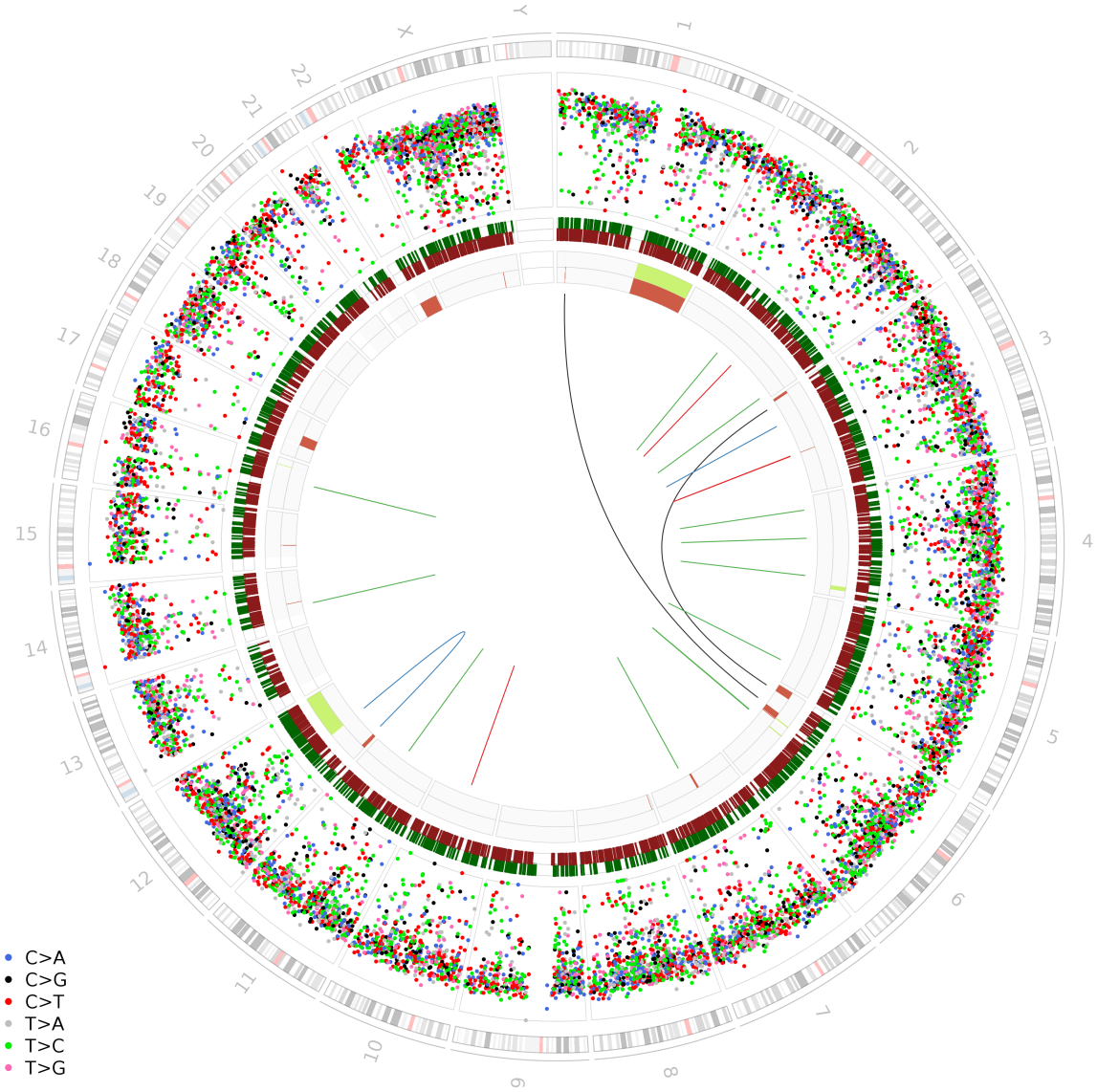
FFPESig is a clean up tool for mutation contexts at the trinucleotide level and therefore direct performance on the basis of specific mutation calls was not possible. This tool uses a previously extracted FFPE signature to correct the trinucleotide context of a sample. For this tool we calculated trinucleotide contexts for all of our samples prior to any correction strategy and then compared the post-corrected context from FFPESig by calculating cosine similarity for context as well as for contributions of mutational signatures calculated for COSMIC version 3 SBS signatures.(Forbes et al. 2017; Alexandrov et al. 2020) We generated the same resulting cosine similarities for samples post filtration with FFPErase and FFPolish and then compared all three methods against one another to see any improvements from the original cosine similarity of the matched FF/FFPE.

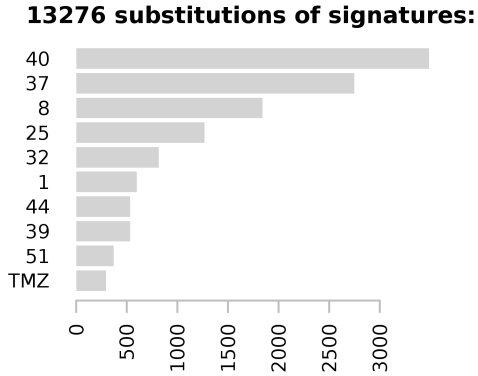
FFPEimpact (https://github.com/Nik-Zainal-Group/FFPE_impact), works similarly to FFPESig using a previously extracted FFPE signature to assign exposures. The tool requires mutations and a categorical organ site as input and outputs an FFPEimpact score for the sample, a cleaned indel context, a corrected micro-homology indel ratio, and signature exposures on the basis of fitting to organ specific signatures plus a custom FFPE signature for SNVs, and COSMIC ID signatures plus a custom FFPE signature for indels. In order to compare this tool directly with the others for signature fitting, we reduced SNVs to produce a cleaned SNV context using the same logic that is used for indels. This outputs a cleaned 96 mutational context which can be directly compared to the FF mutational context using cosine similarity.

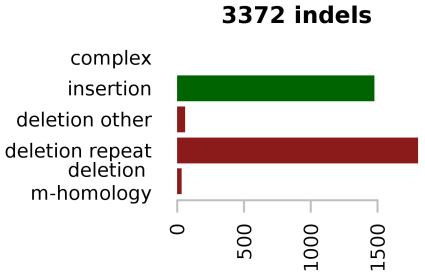
Circos plots for cohort

Included below are circos plots of genome-wide calls for all 56 of the matched trios used for training in the manuscript. The outermost ring shows the intermutation distance for all SNVs color-coded by the pyrimidine partner of the mutated base. The middle ring shows small insertions (green) and deletions (red). The innermost ring shows copy number changes, and the arcs show SVs. The barplot in the upper right shows the absolute number of mutations attributed to the five mutational signatures with the highest exposure in the tumor. The barplot in the middle right shows the distribution of insertions and deletions by categorical type. The barplot in the bottom right shows the distribution of structural variant types.

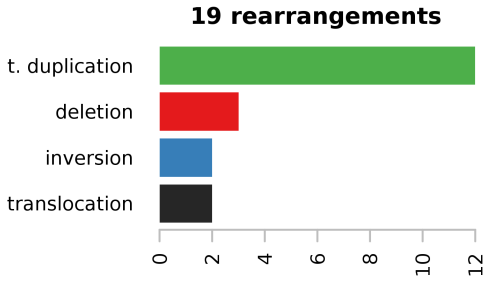
H134753 - FFPE



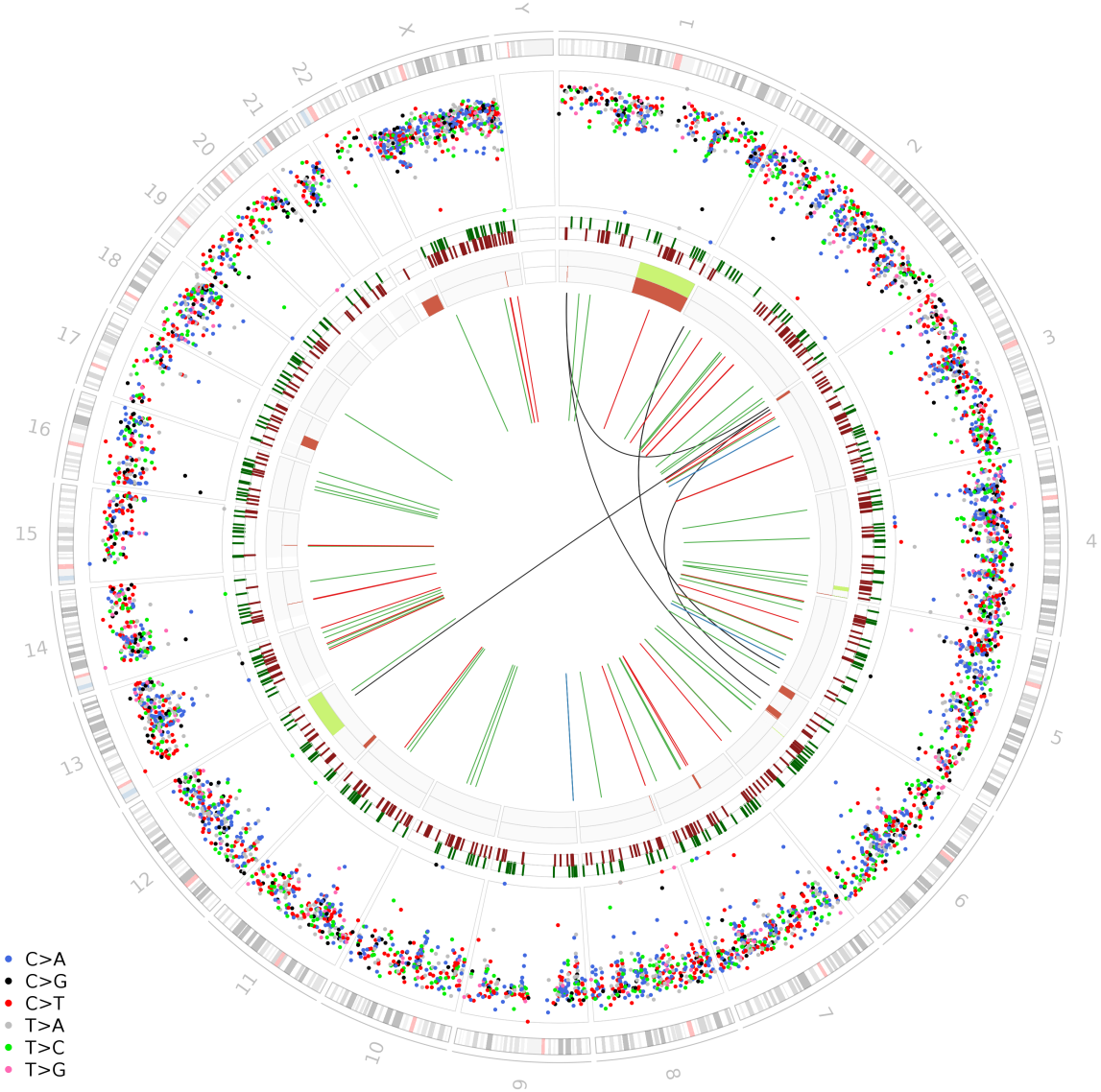




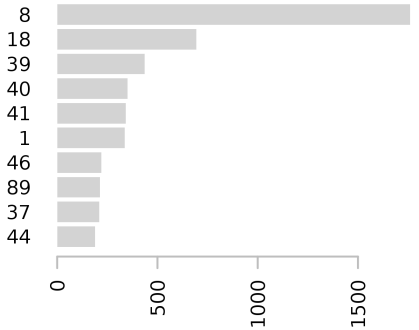




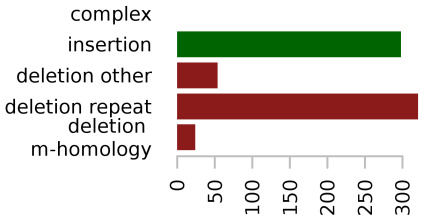
H134753 - Fresh Frozen



5359 substitutions of signatures:



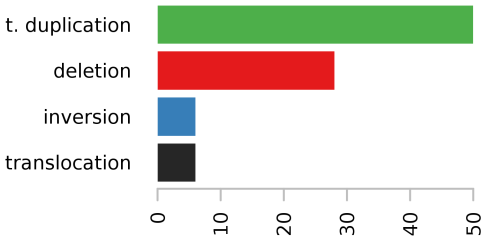
697 indels



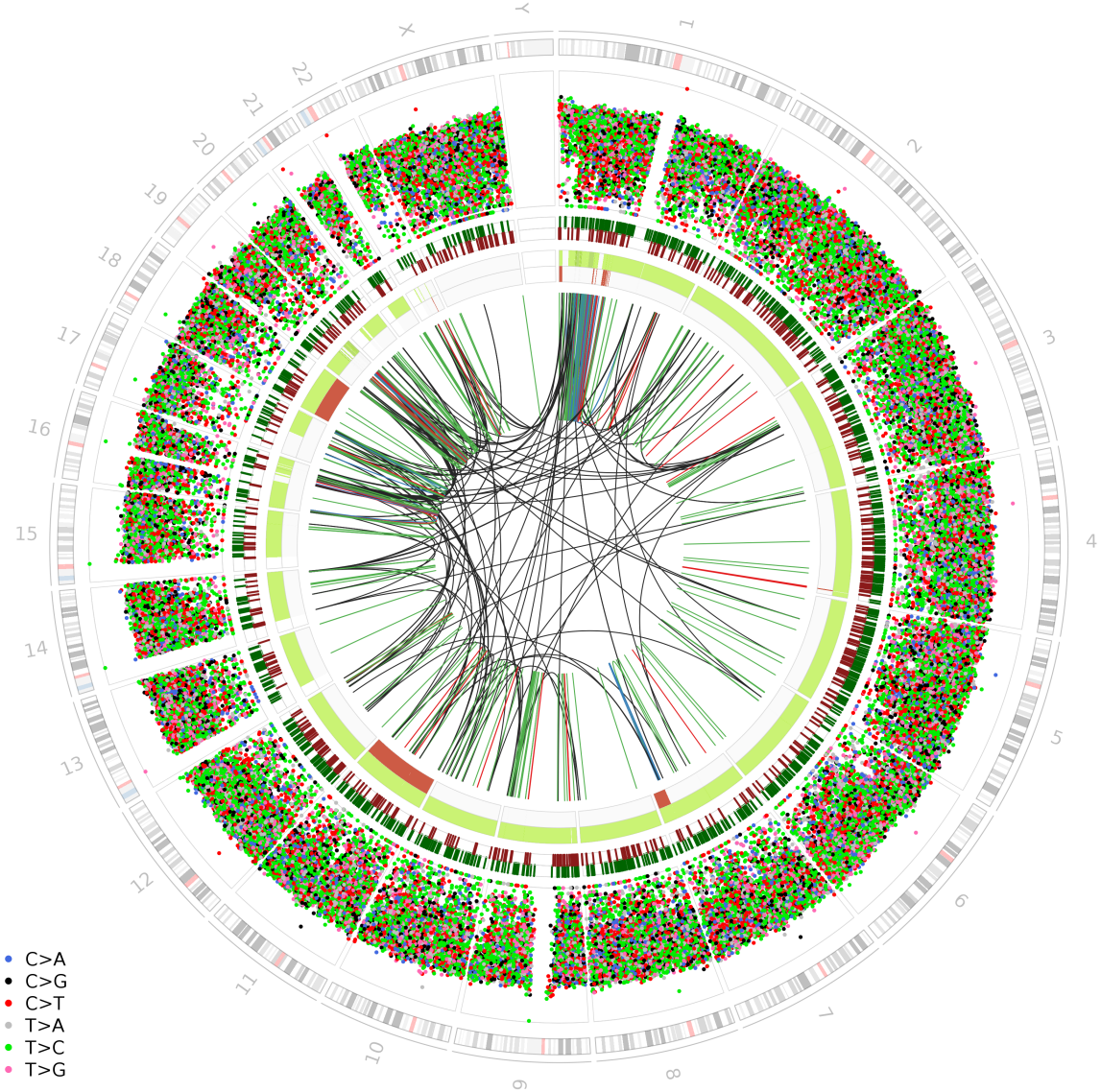
karyotype (purity 0.53, ploidy 2.1)



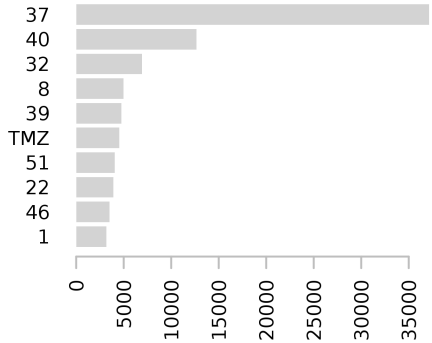
90 rearrangements



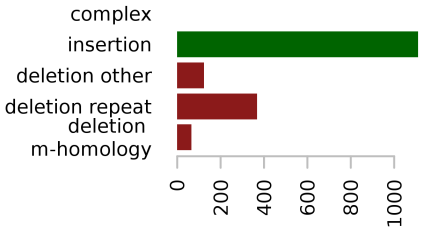
H135022 - FFPE



95003 substitutions of signatures:



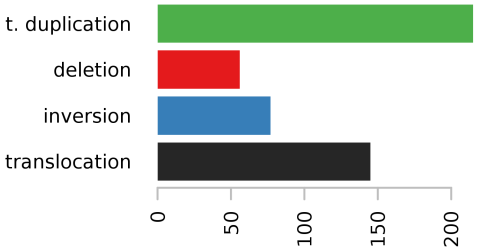
1669 indels



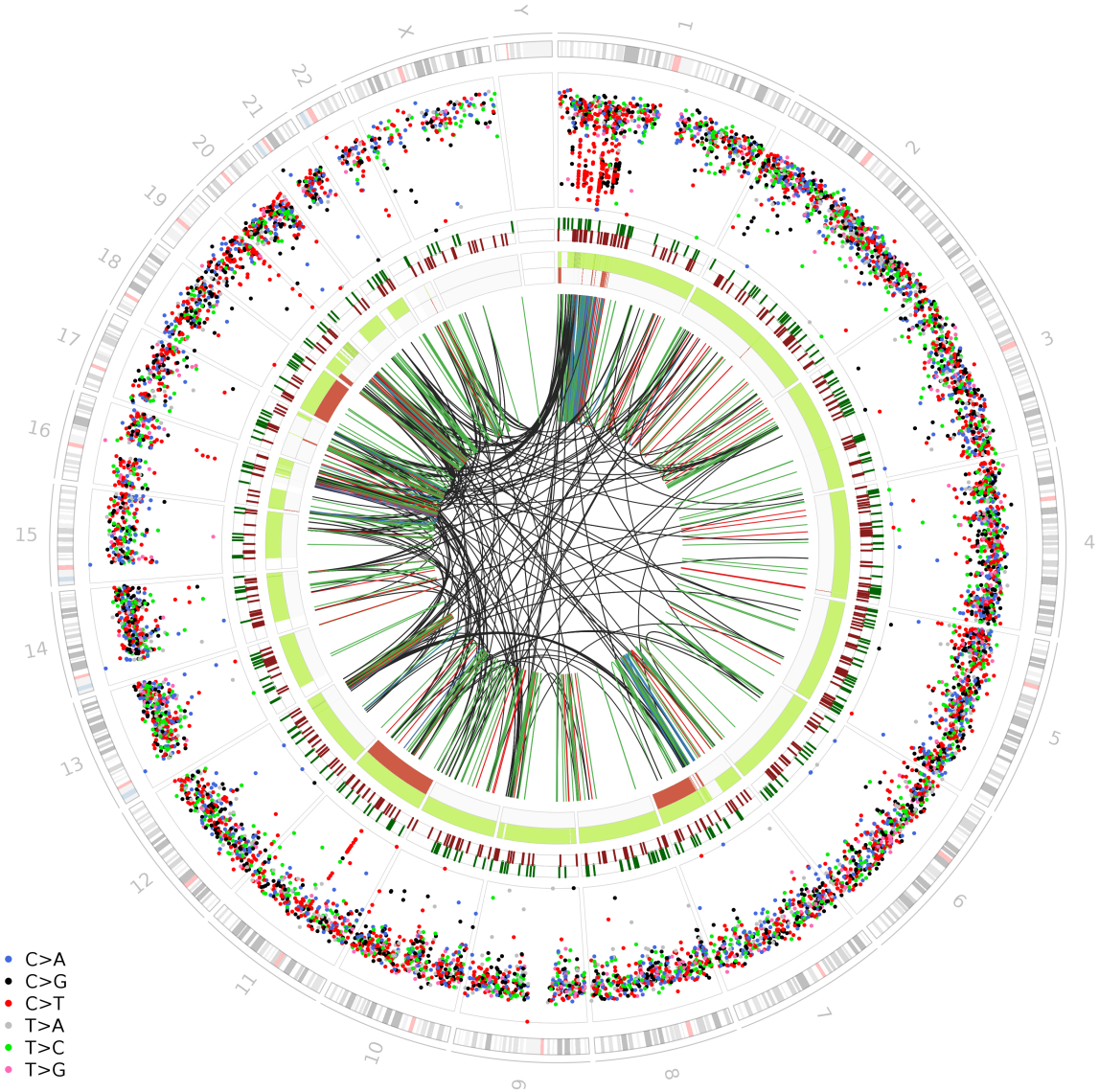
karyotype (purity 0.6, ploidy 3.4)



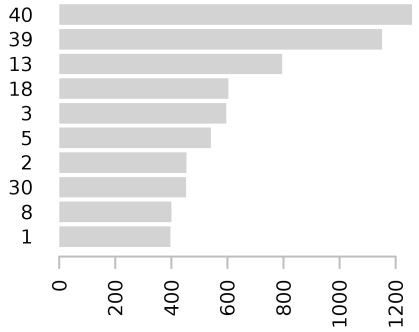
493 rearrangements



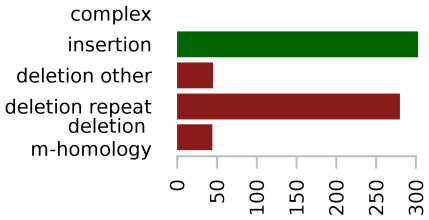
H135022 - Fresh Frozen



7916 substitutions of signatures:



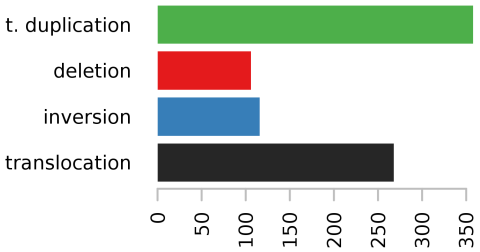
672 indels



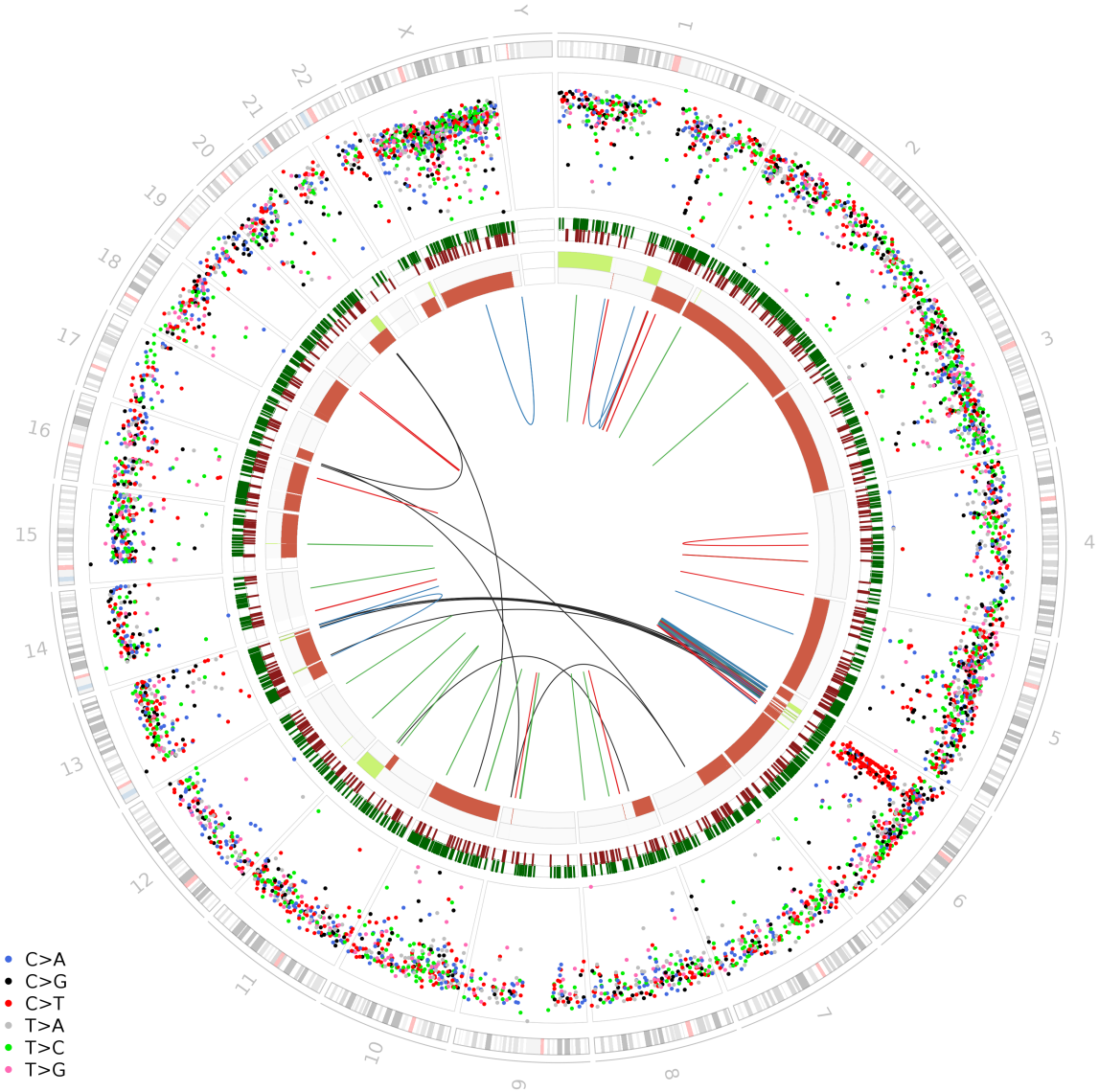
karyotype (purity 0.84, ploidy 3.2)

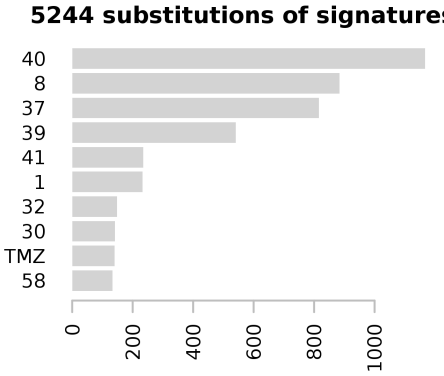


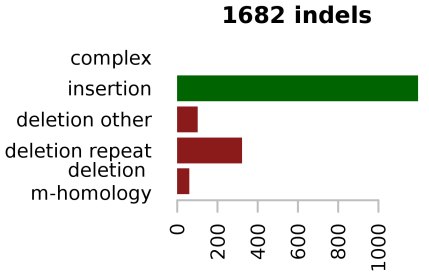
848 rearrangements



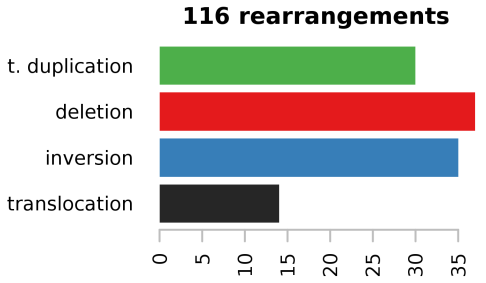
H135387 - FFPE



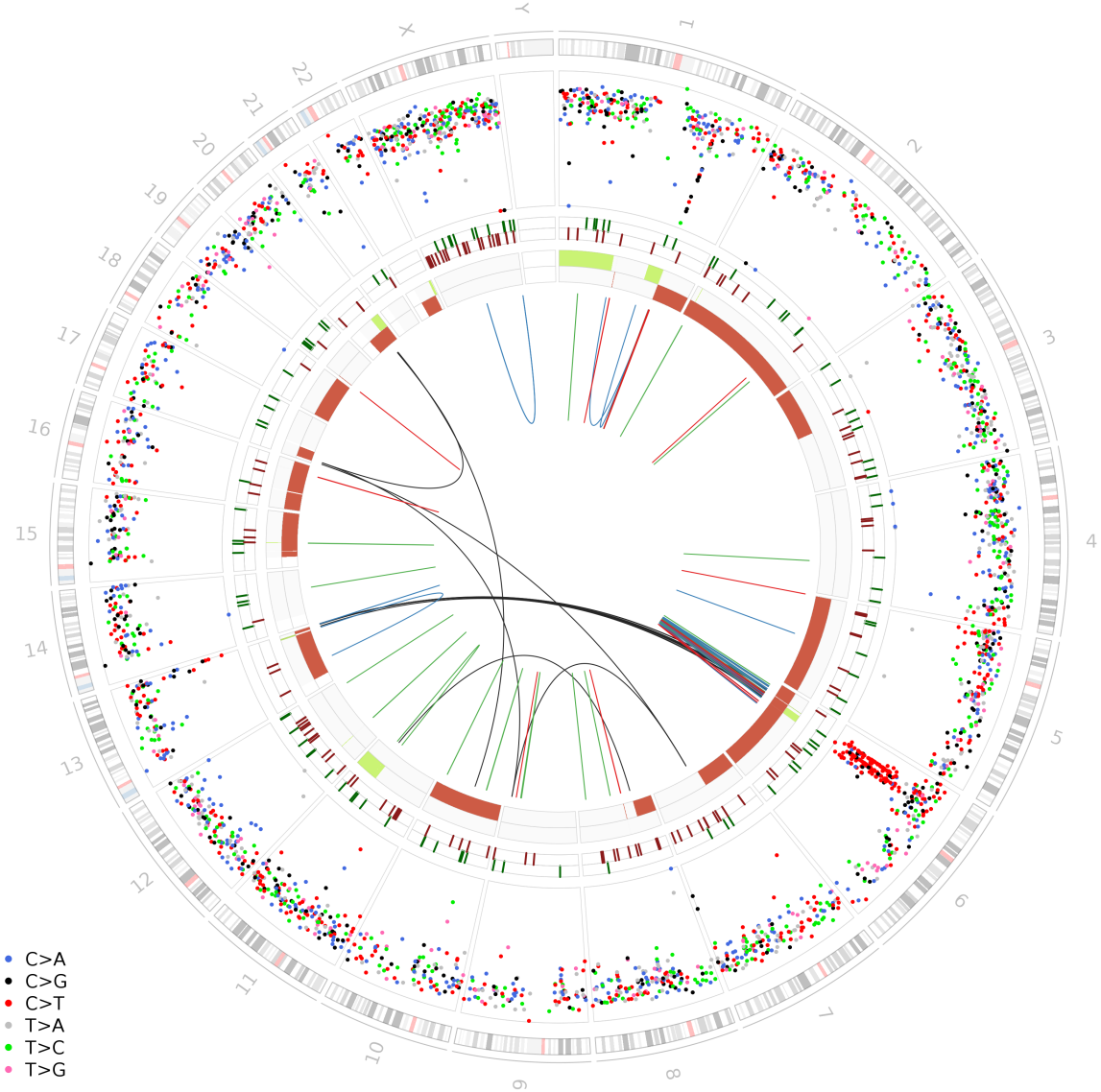




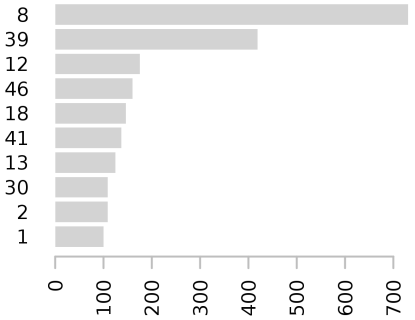




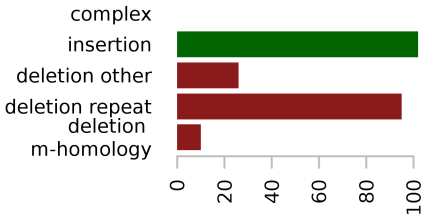
H135387 - Fresh Frozen



2866 substitutions of signatures:



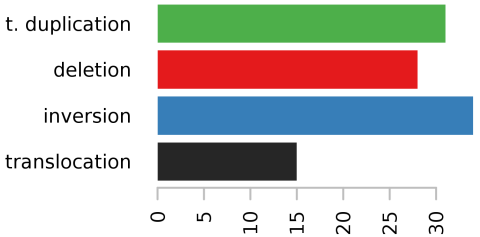
233 indels



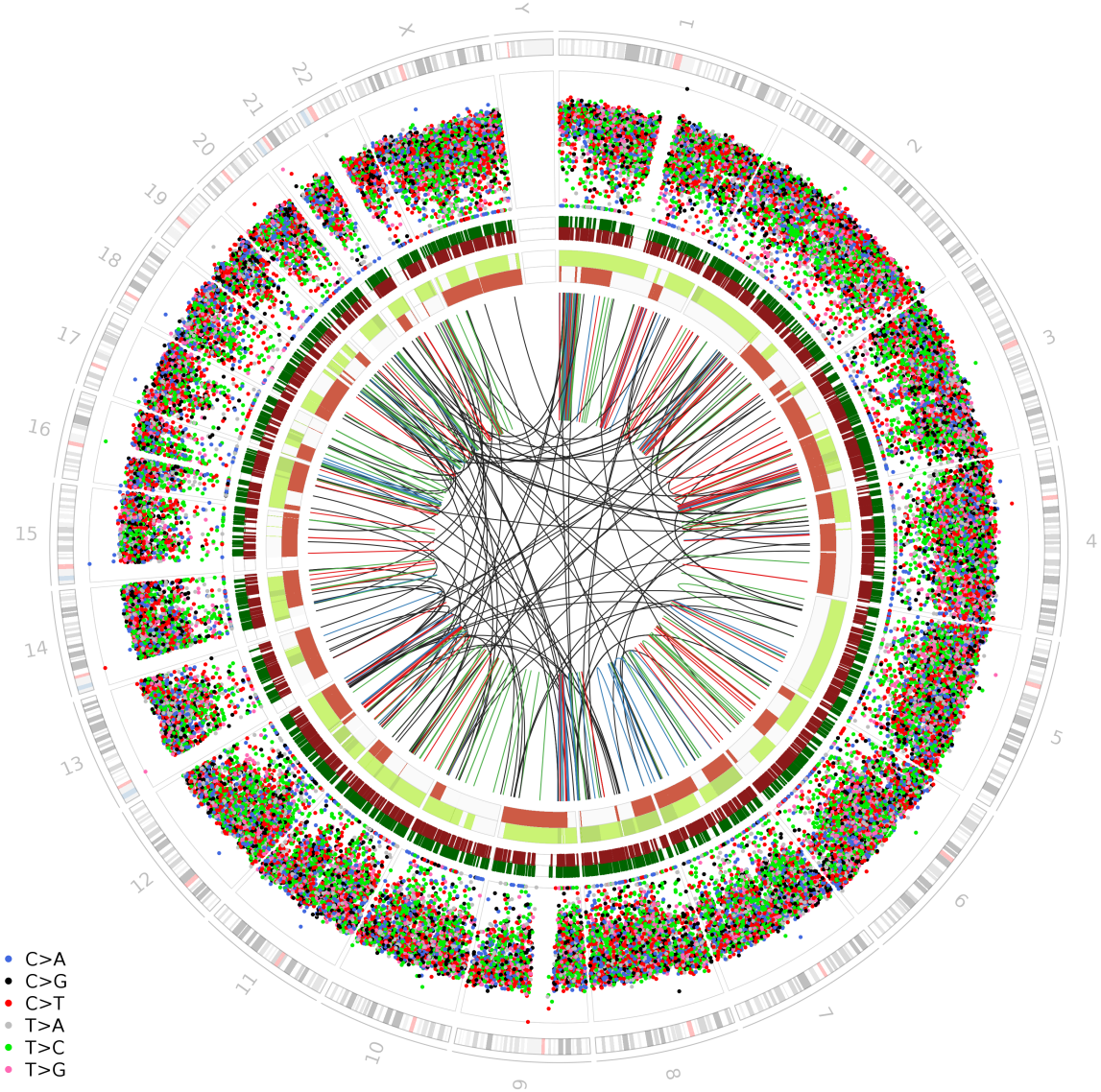
karyotype (purity 0.55, ploidy 1.8)

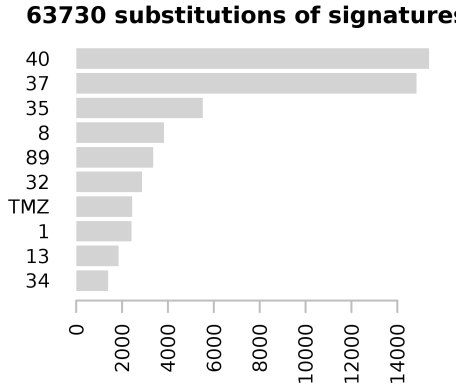


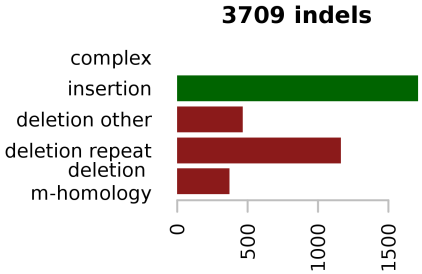
108 rearrangements



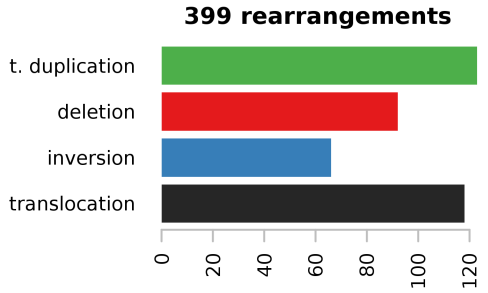
H135462 - FFPE



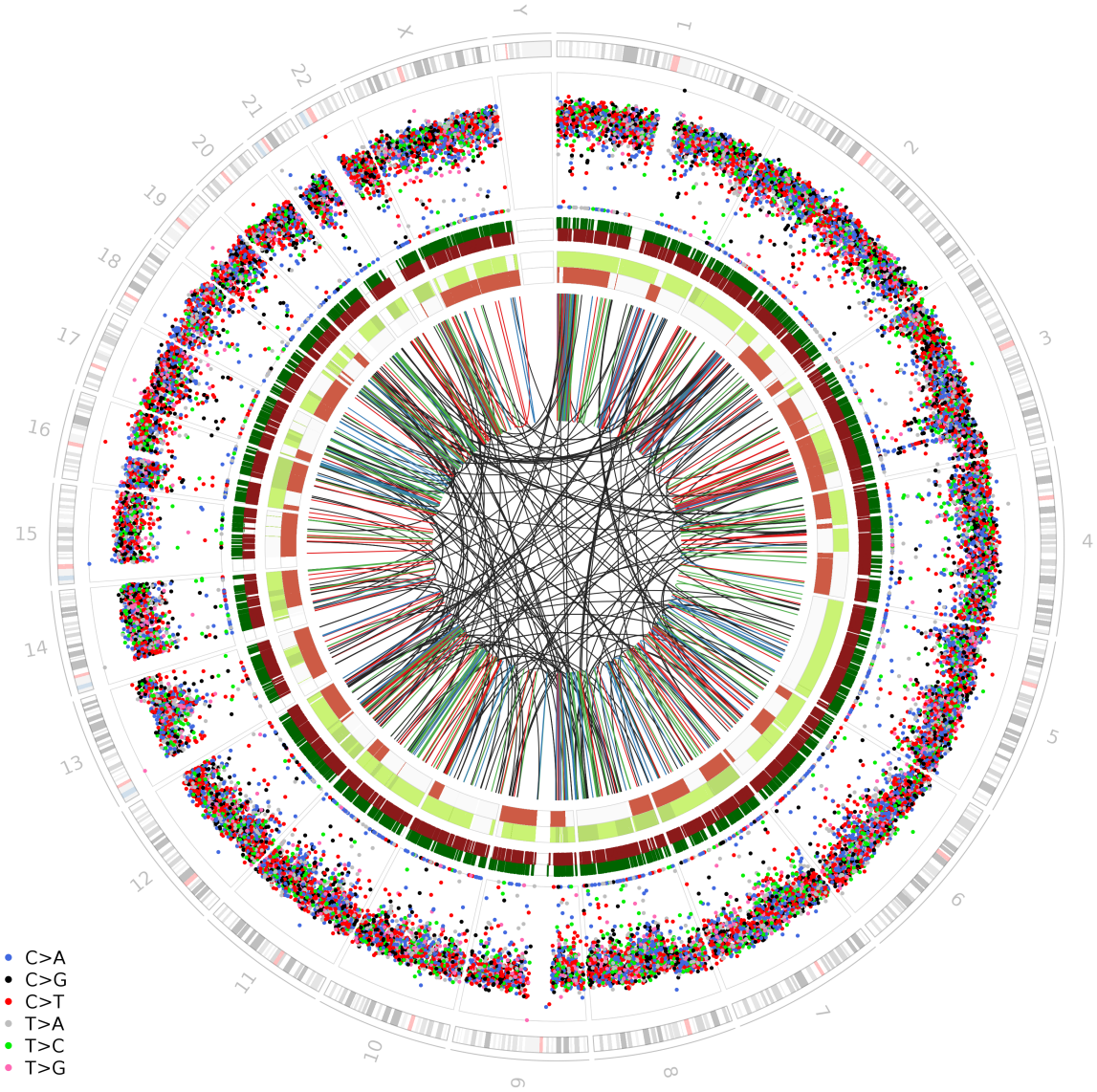




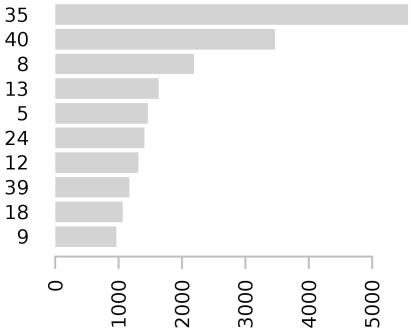




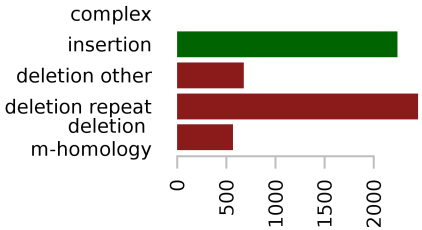
H135462 - Fresh Frozen



30349 substitutions of signatures:



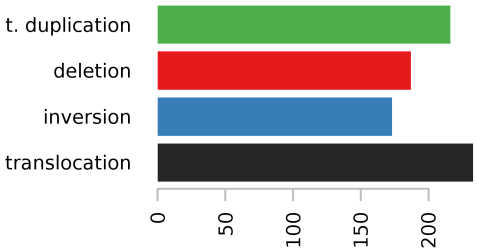
5941 indels



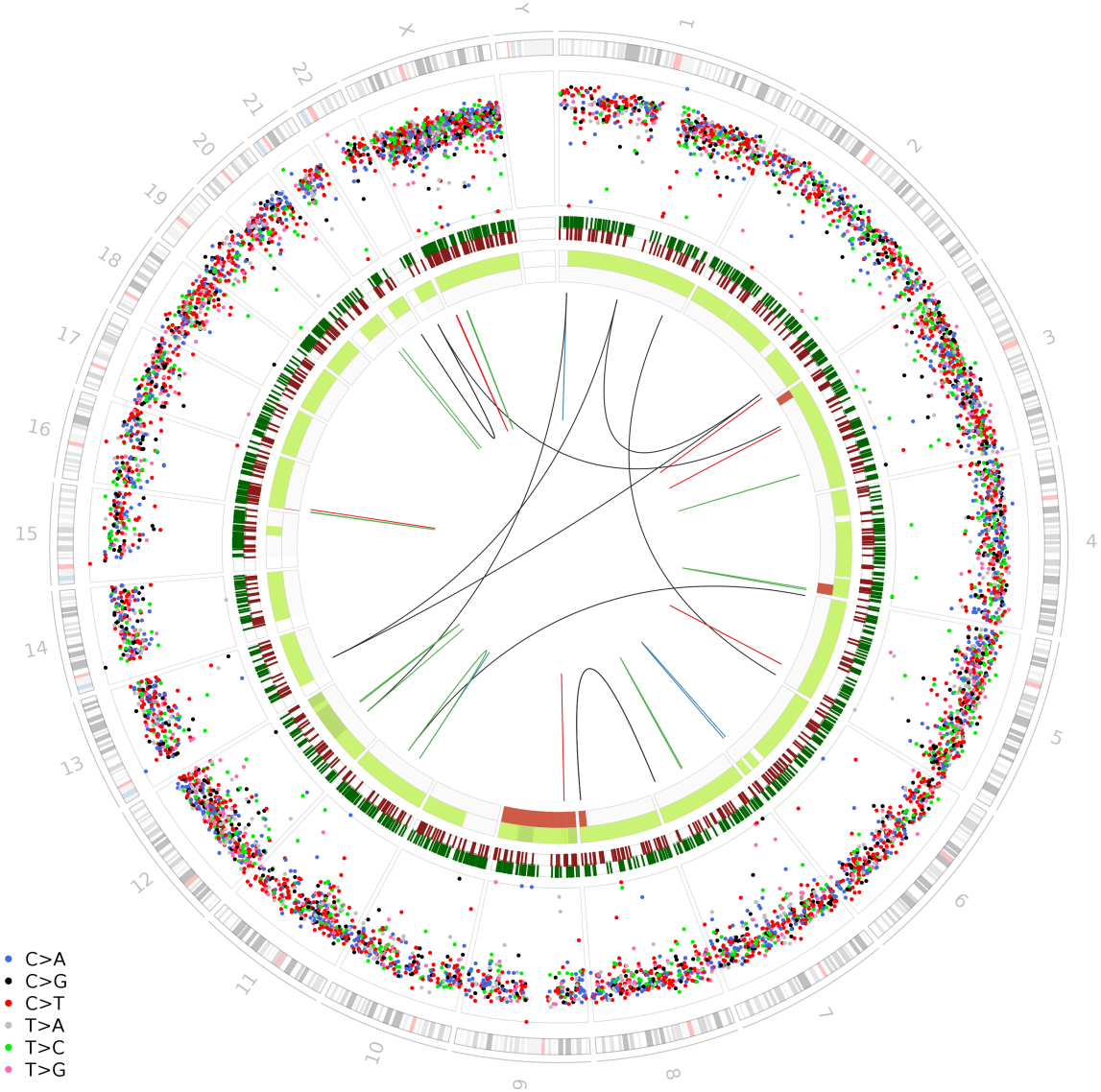
karyotype (purity 0.23, ploidy 3.2)

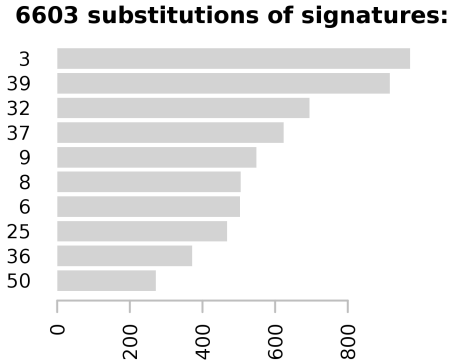


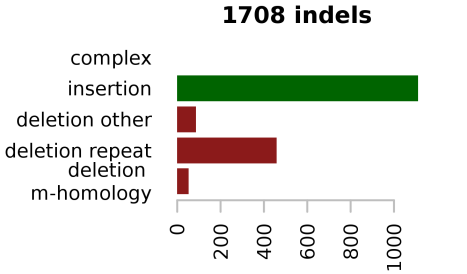
809 rearrangements



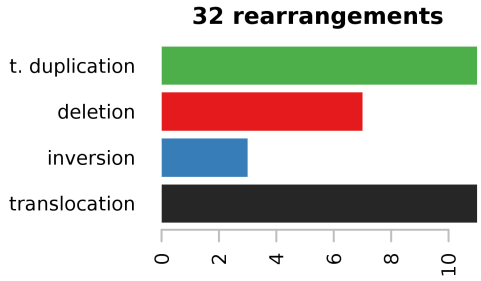
H156407 - FFPE



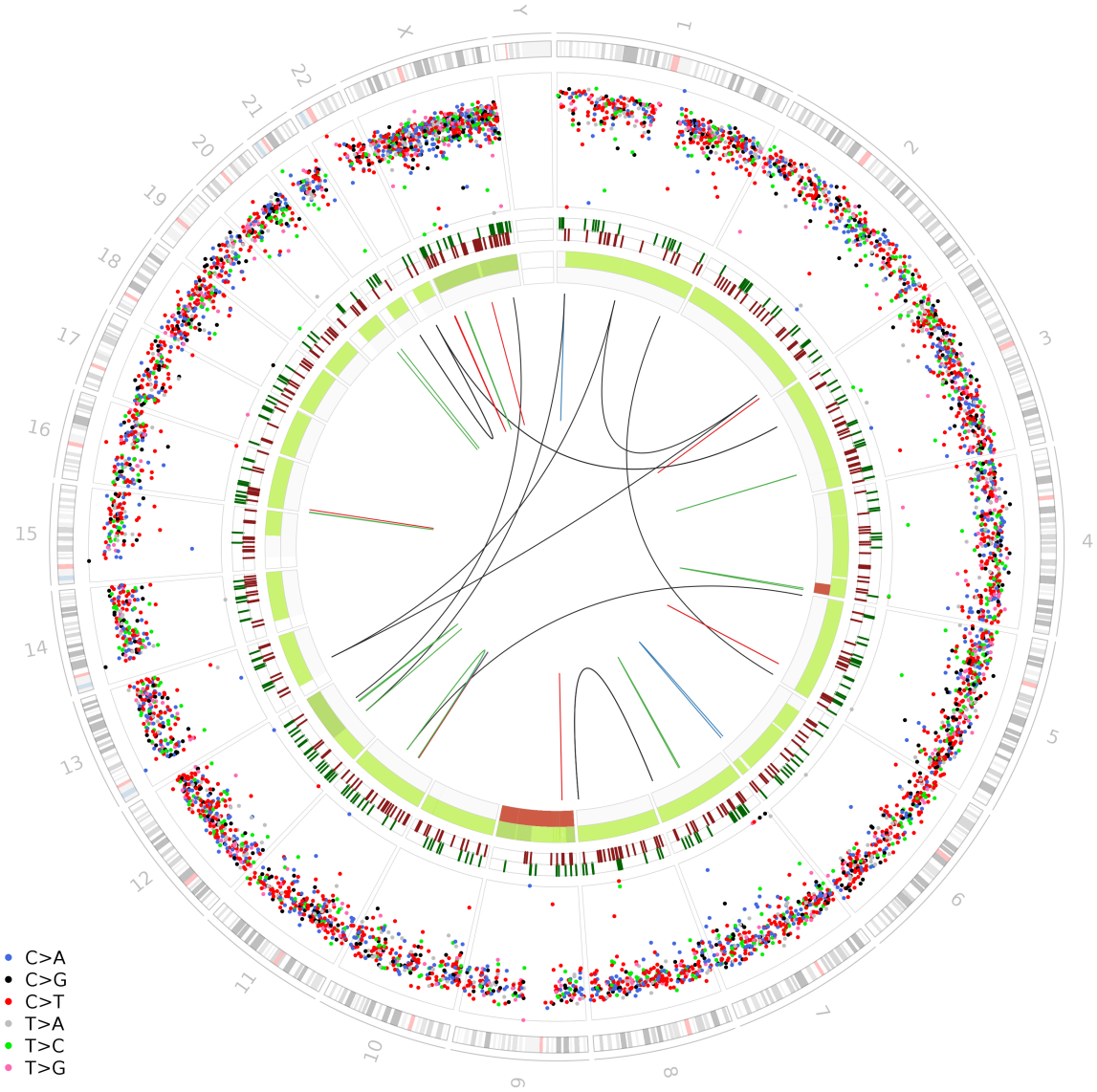




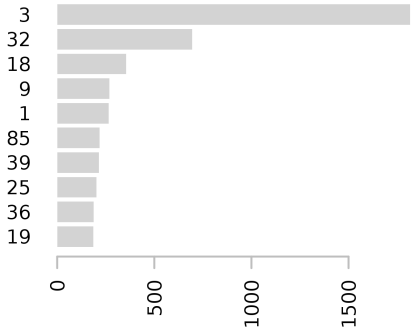




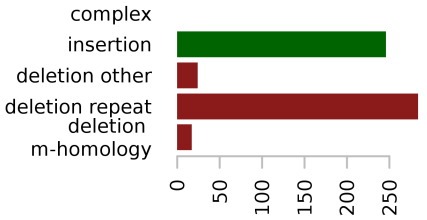
H156407 - Fresh Frozen



5433 substitutions of signatures:



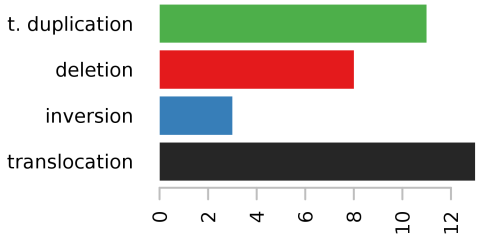
571 indels



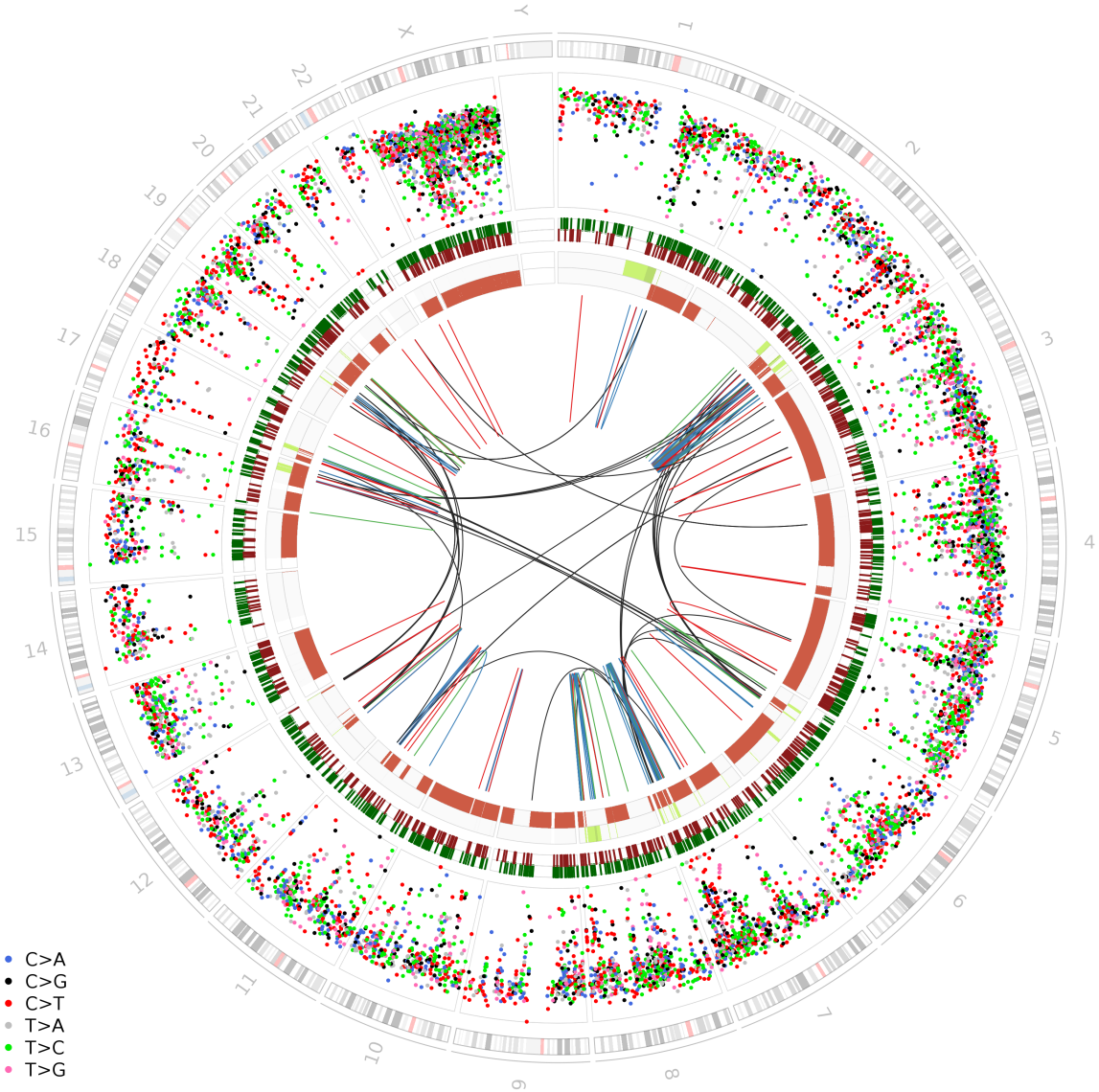
karyotype (purity 0.62, ploidy 3.9)



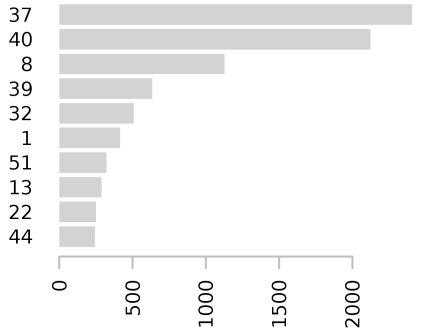
35 rearrangements



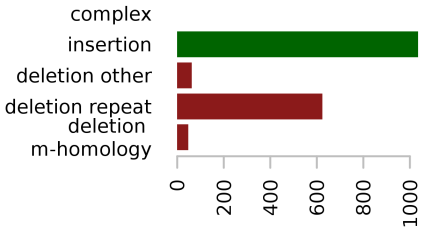
H158273 - FFPE



9689 substitutions of signatures:



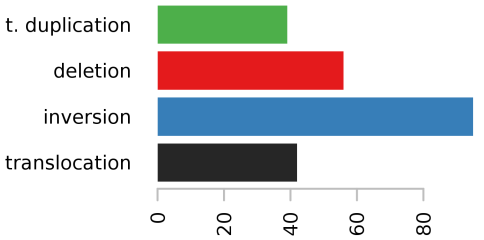
1770 indels



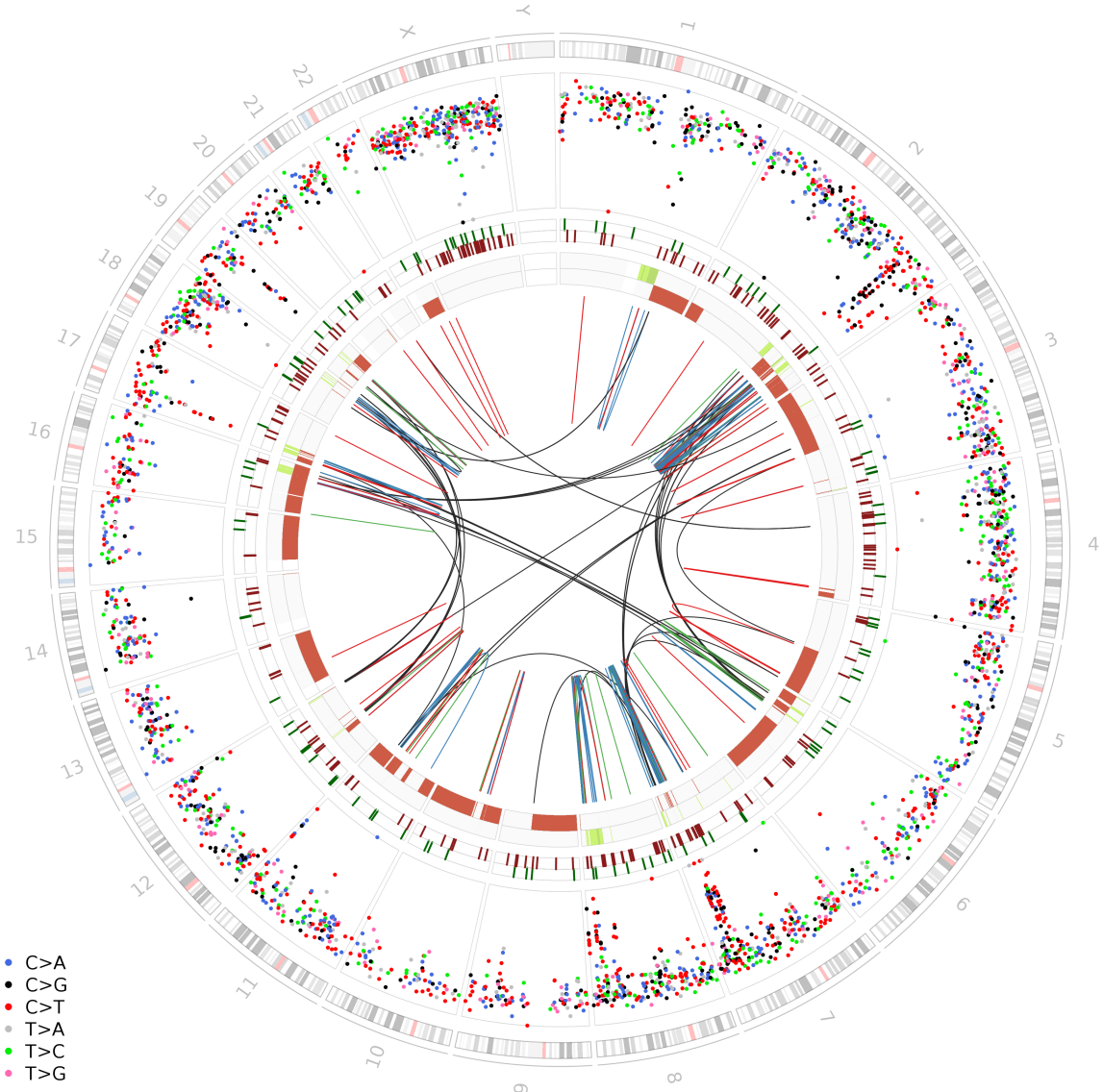
karyotype (purity 0.91, ploidy 1.8)



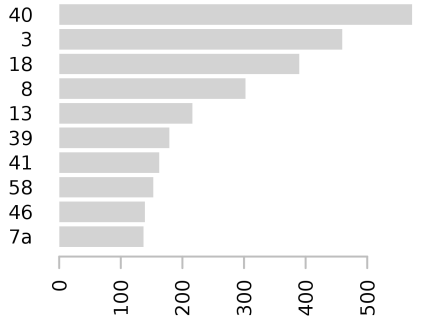
232 rearrangements



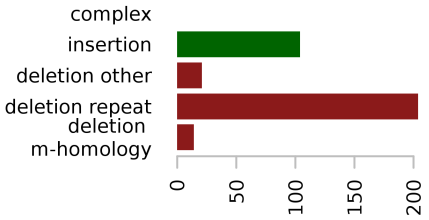
H158273 - Fresh Frozen



3082 substitutions of signatures:



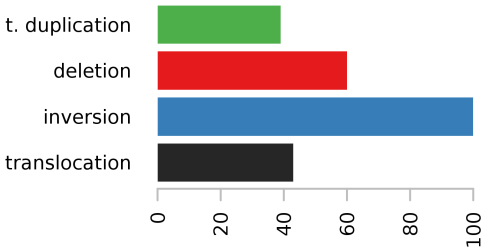
343 indels



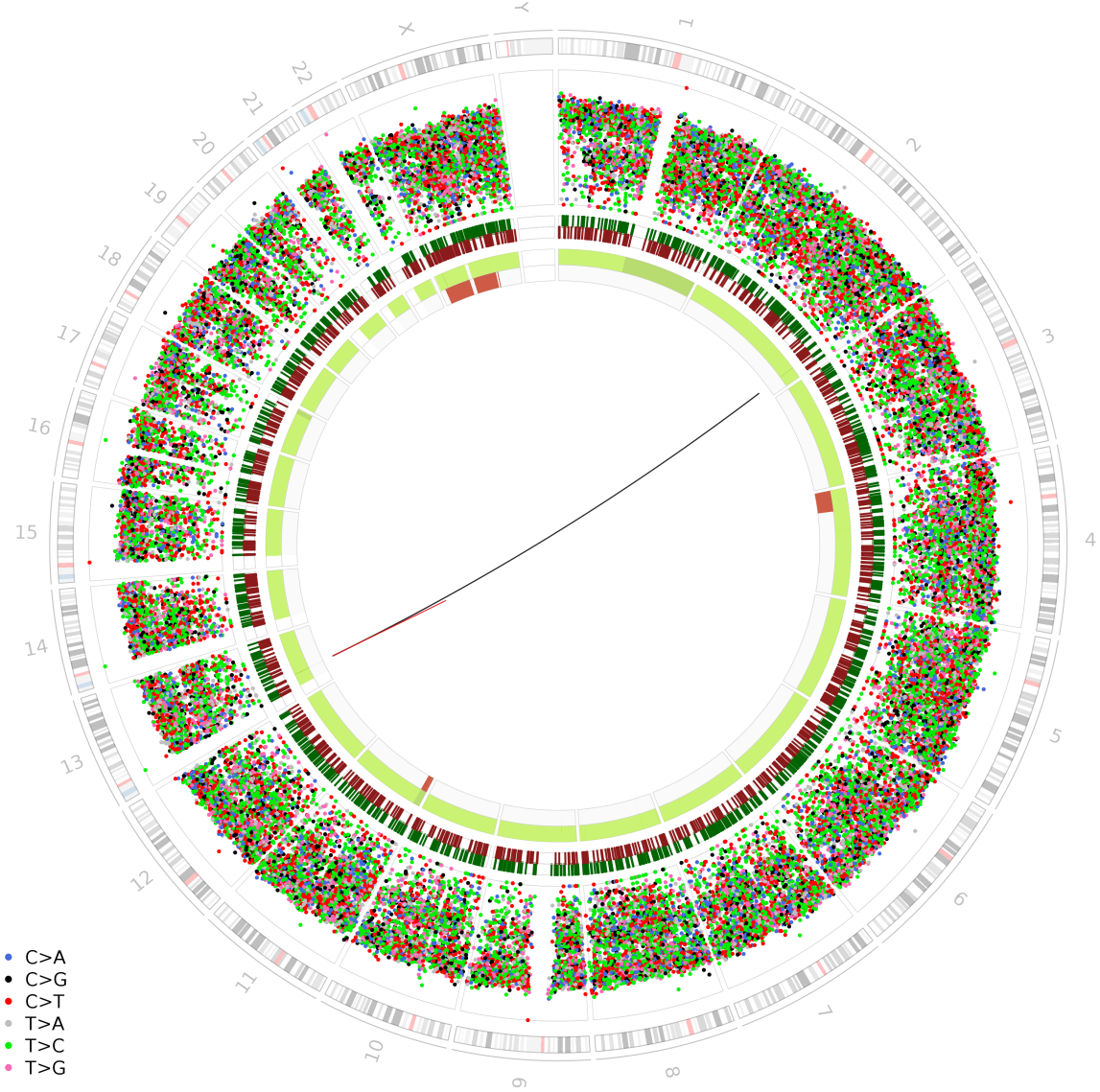
karyotype (purity 0.82, ploidy 1.7)



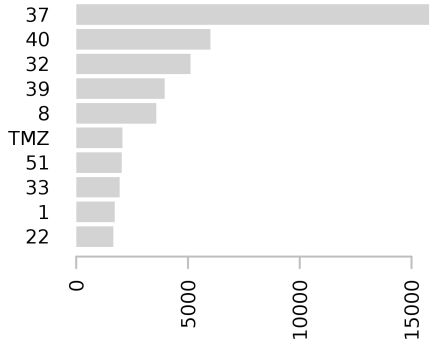
242 rearrangements



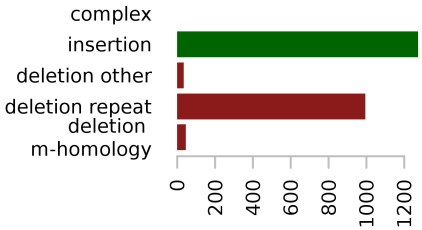
H158878 - FFPE



52228 substitutions of signatures:



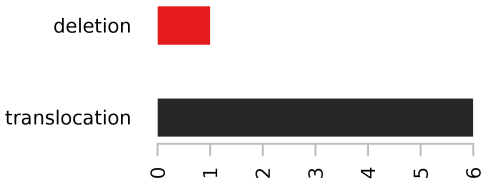
2347 indels



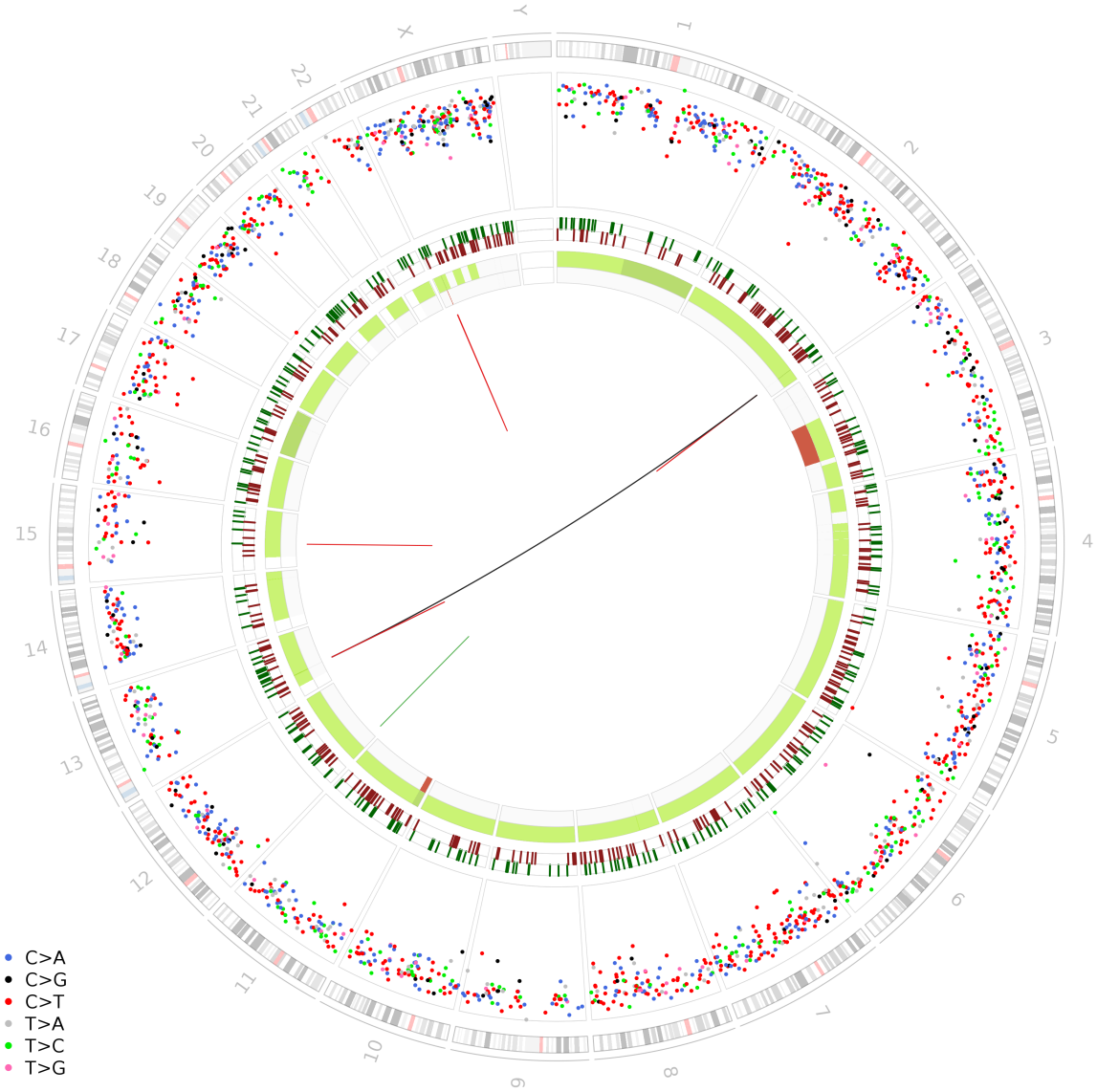
karyotype (purity 0.54, ploidy 4.1)



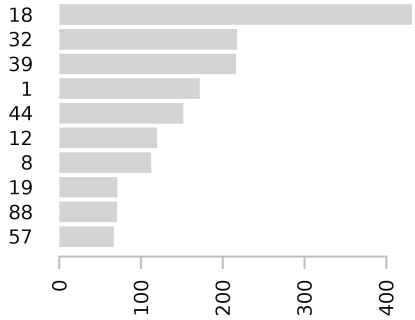
7 rearrangements



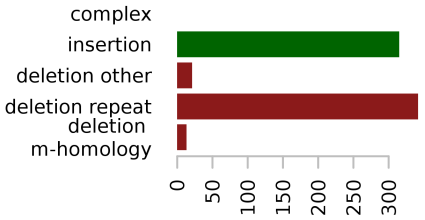
H158878 - Fresh Frozen



1953 substitutions of signatures:



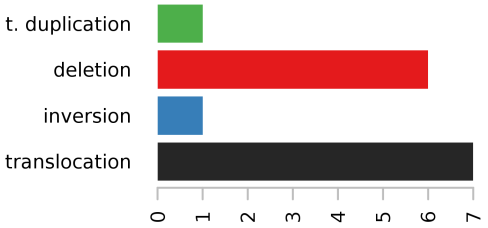
691 indels



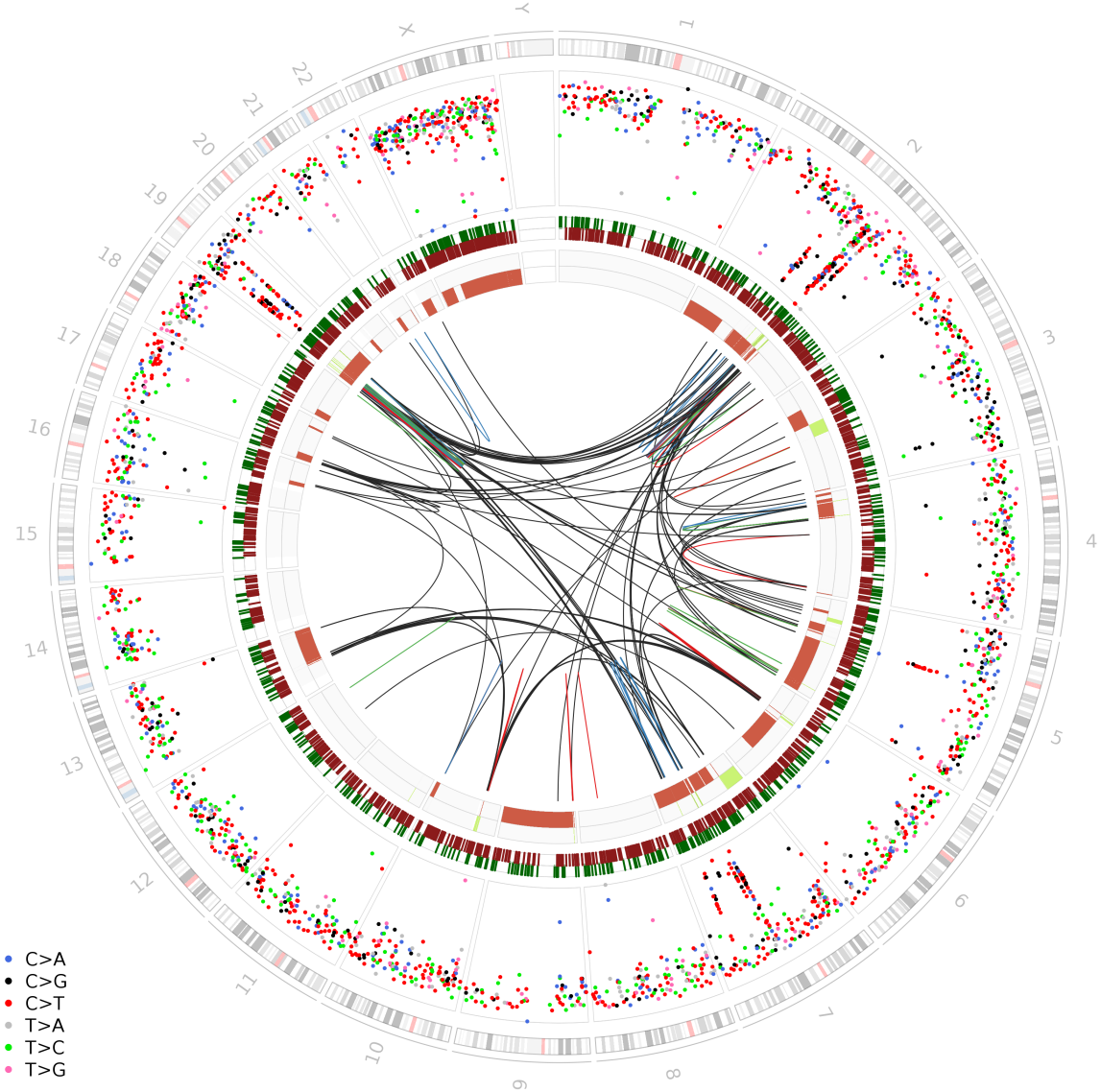
karyotype (purity 0.65, ploidy 4.2)



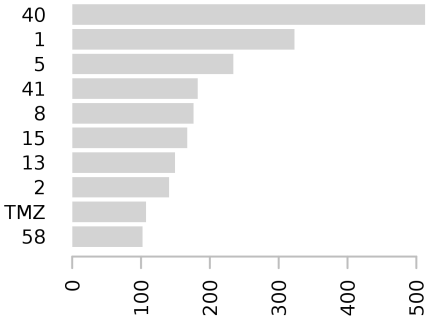
15 rearrangements



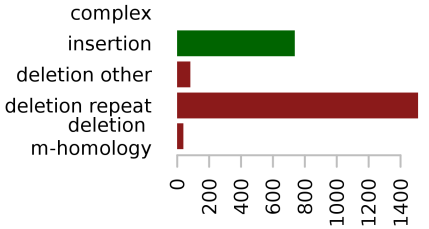
H158880 - FFPE



2602 substitutions of signatures:



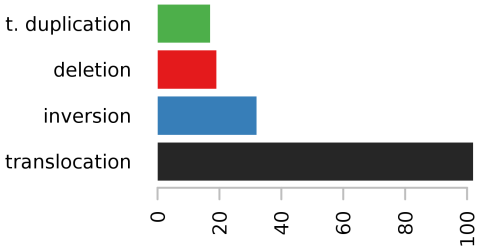
2368 indels



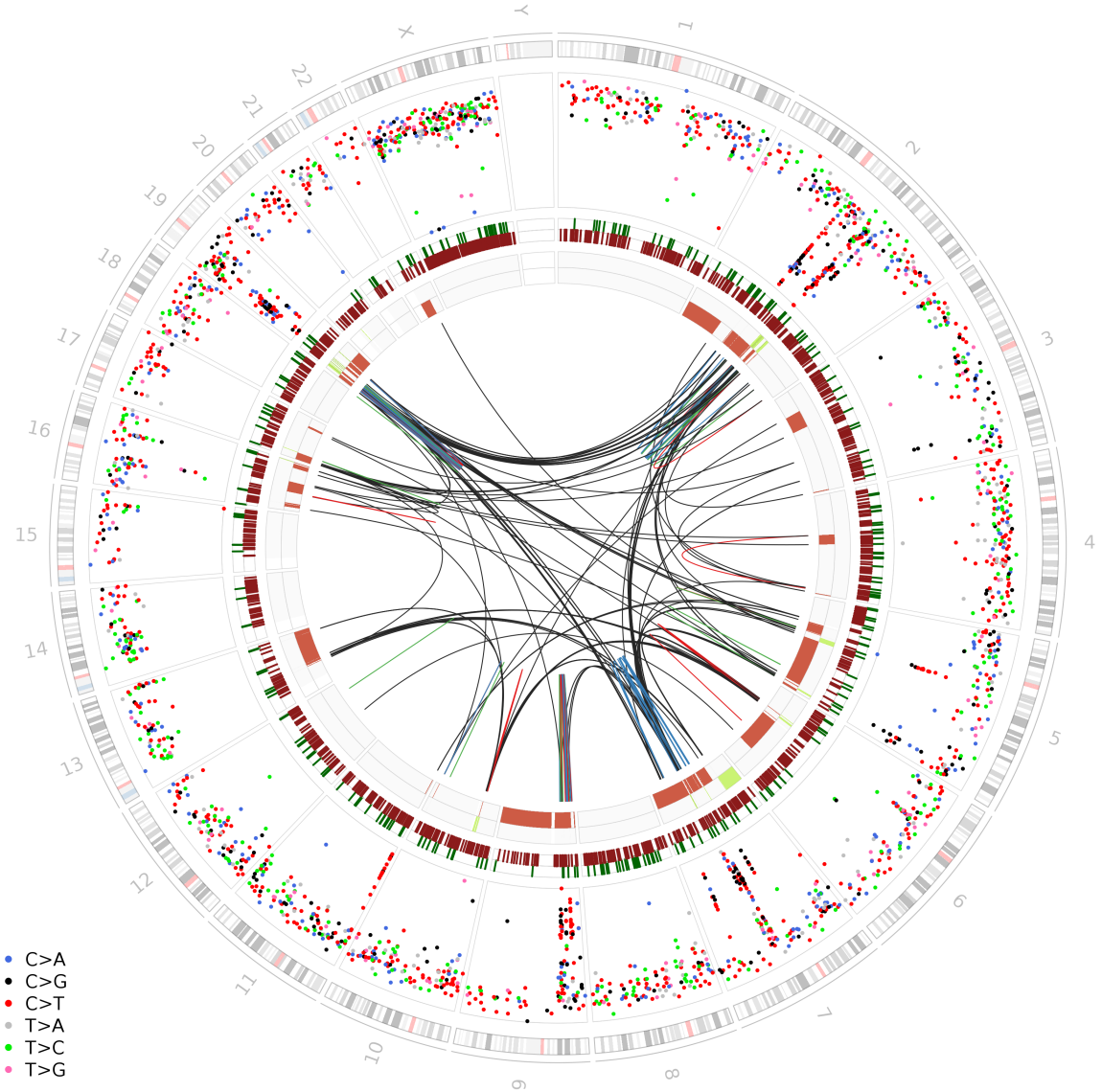
karyotype (purity 0.87, ploidy 2)



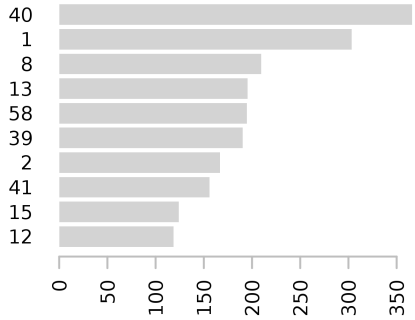
170 rearrangements



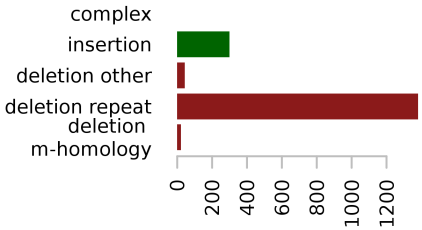
H158880 - Fresh Frozen



2388 substitutions of signatures:



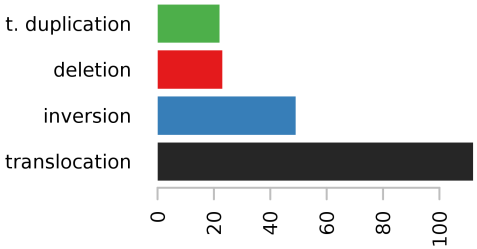
1746 indels

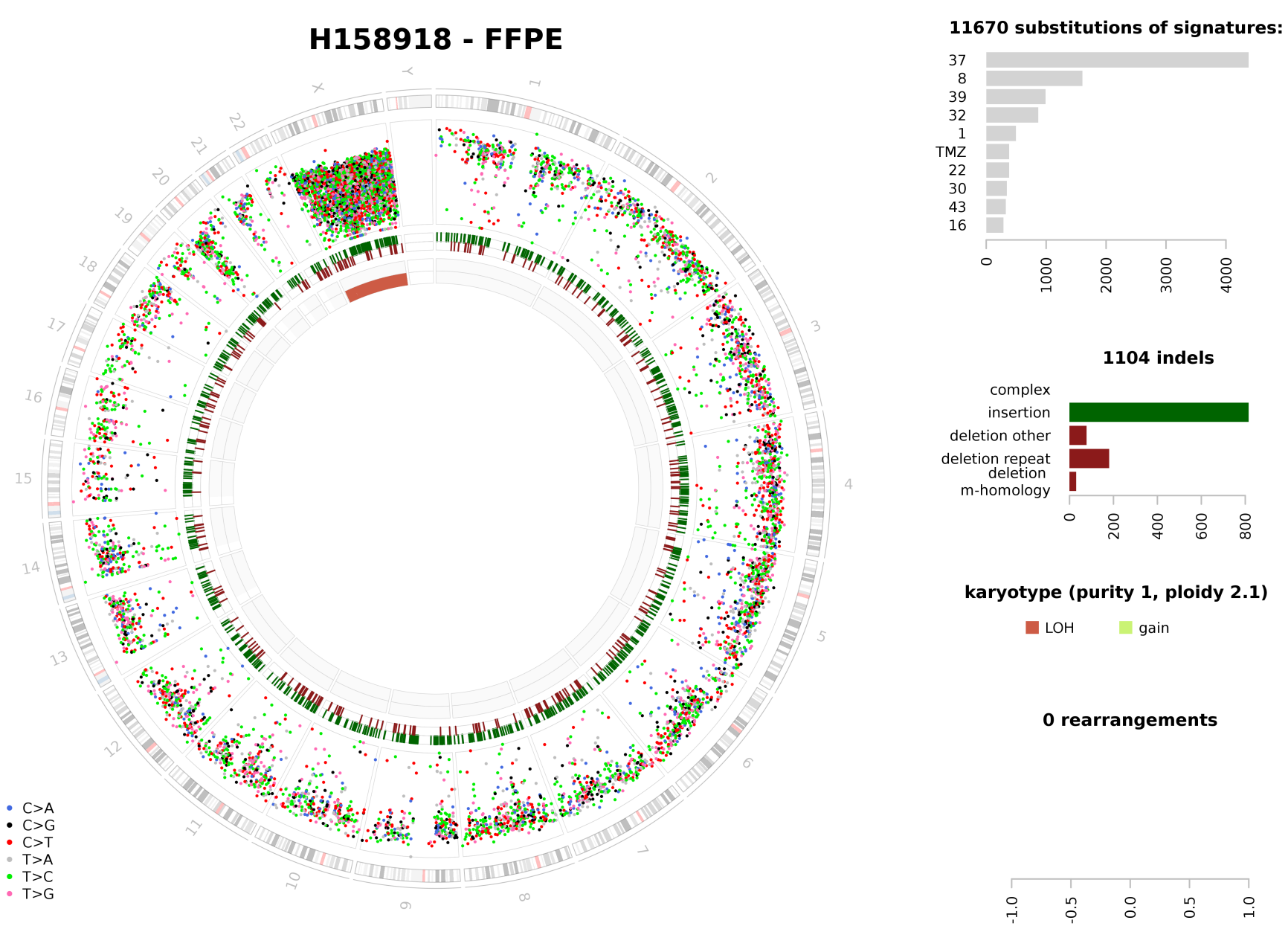


karyotype (purity 0.73, ploidy 2)

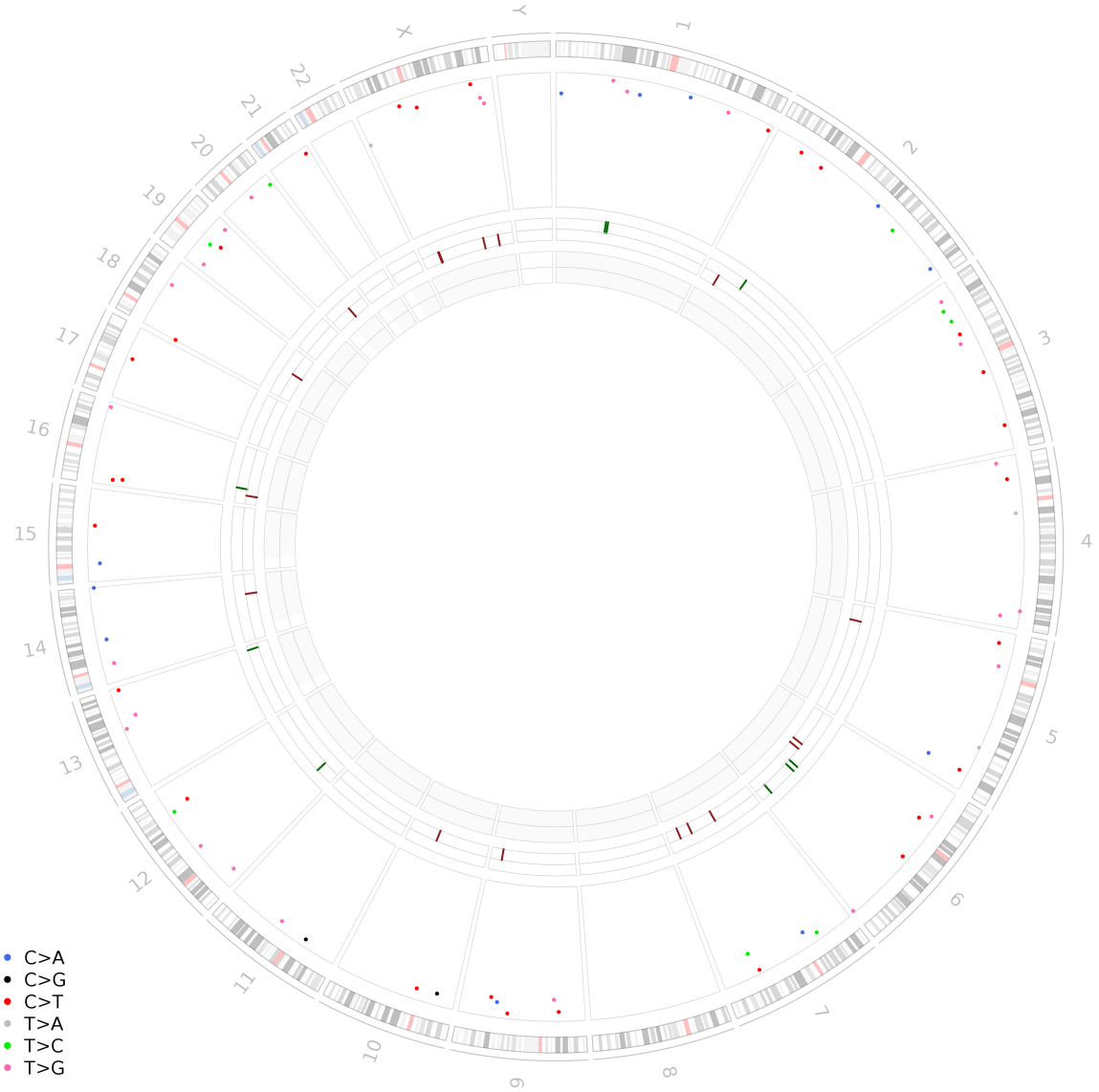


206 rearrangements

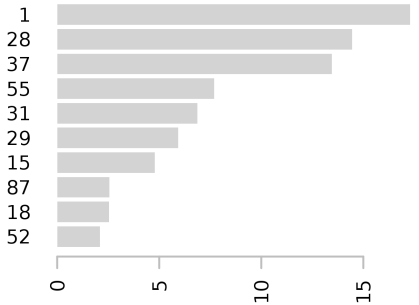



H158918 - FFPE
C>A
C>G
C>T
T>A
T>C
T>G
11670 substitutions of signatures:
37
8
39
32
1
TMZ
22
30
43
16
0
1000
2000
3000
4000
1104 indels
complex
insertion
deletion other
deletion repeat
deletion
m-homology
0
200
400
600
800
karyotype (purity 1, ploidy 2.1)
LOH
gain
0 rearrangements
-1.0
-0.5
0.0
0.5
1.0

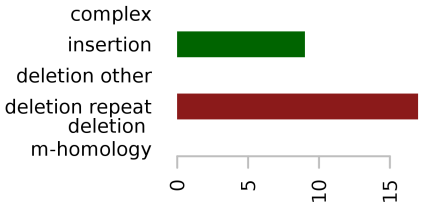
H158918 - Fresh Frozen



77 substitutions of signatures:



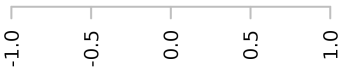
26 indels



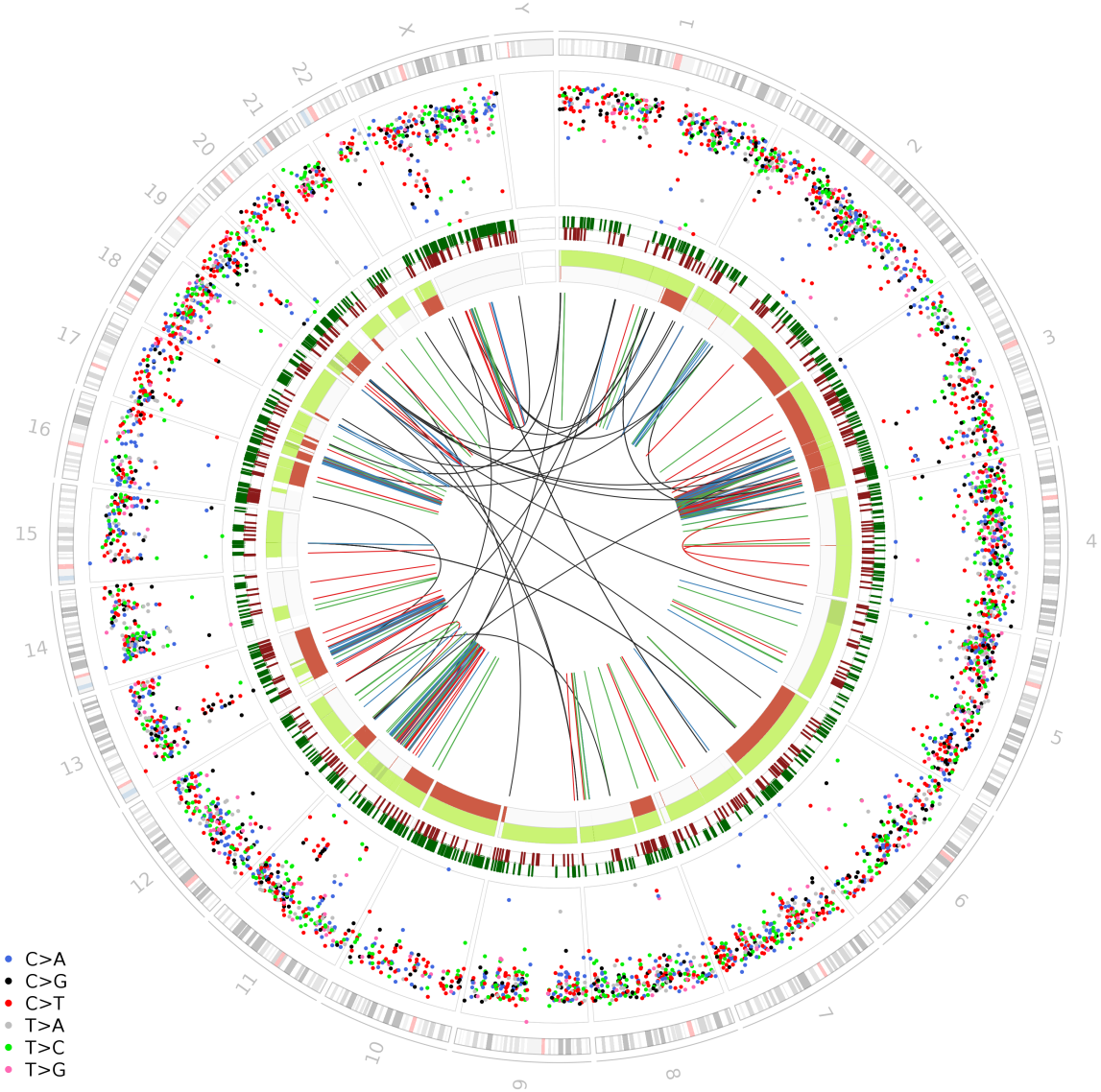
karyotype (purity 1, ploidy 2)

LOH gain

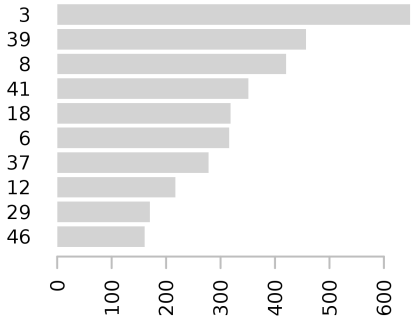
0 rearrangements



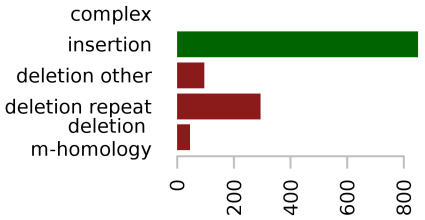
H159369 - FFPE



4201 substitutions of signatures:



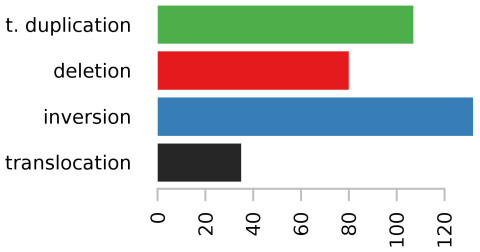
1286 indels



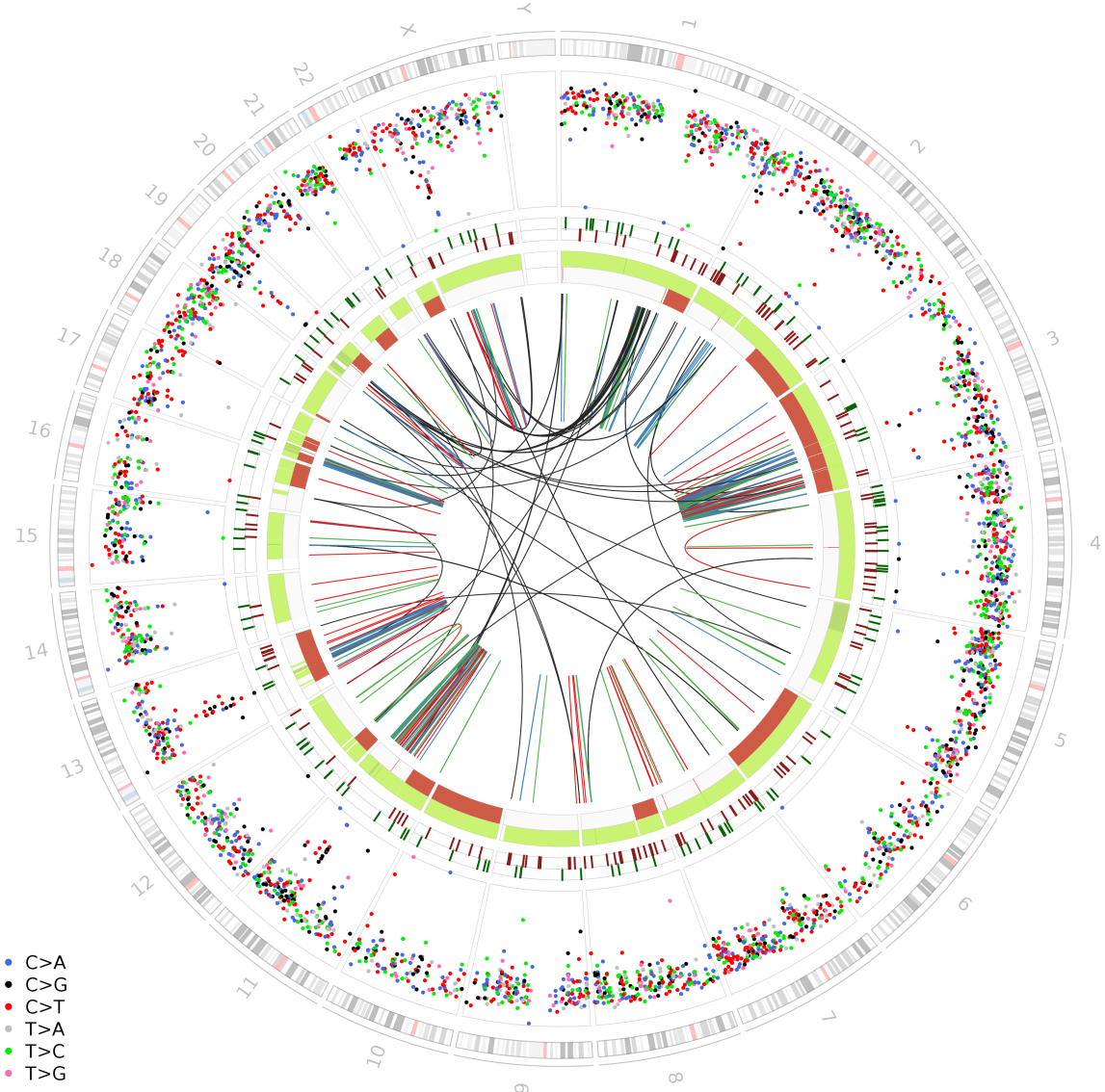
karyotype (purity 0.89, ploidy 3.5)



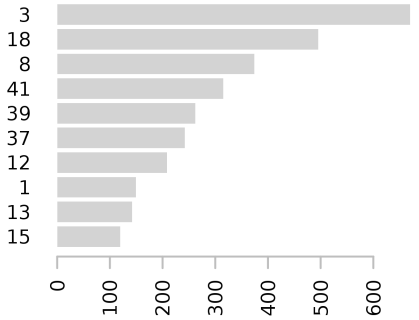
354 rearrangements



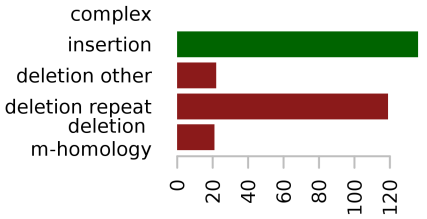
H159369 - Fresh Frozen



3746 substitutions of signatures:



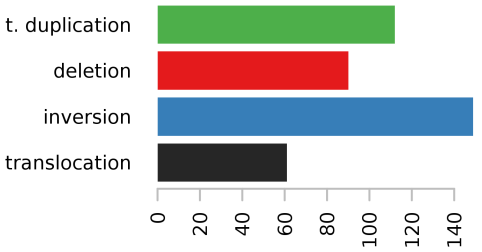
298 indels

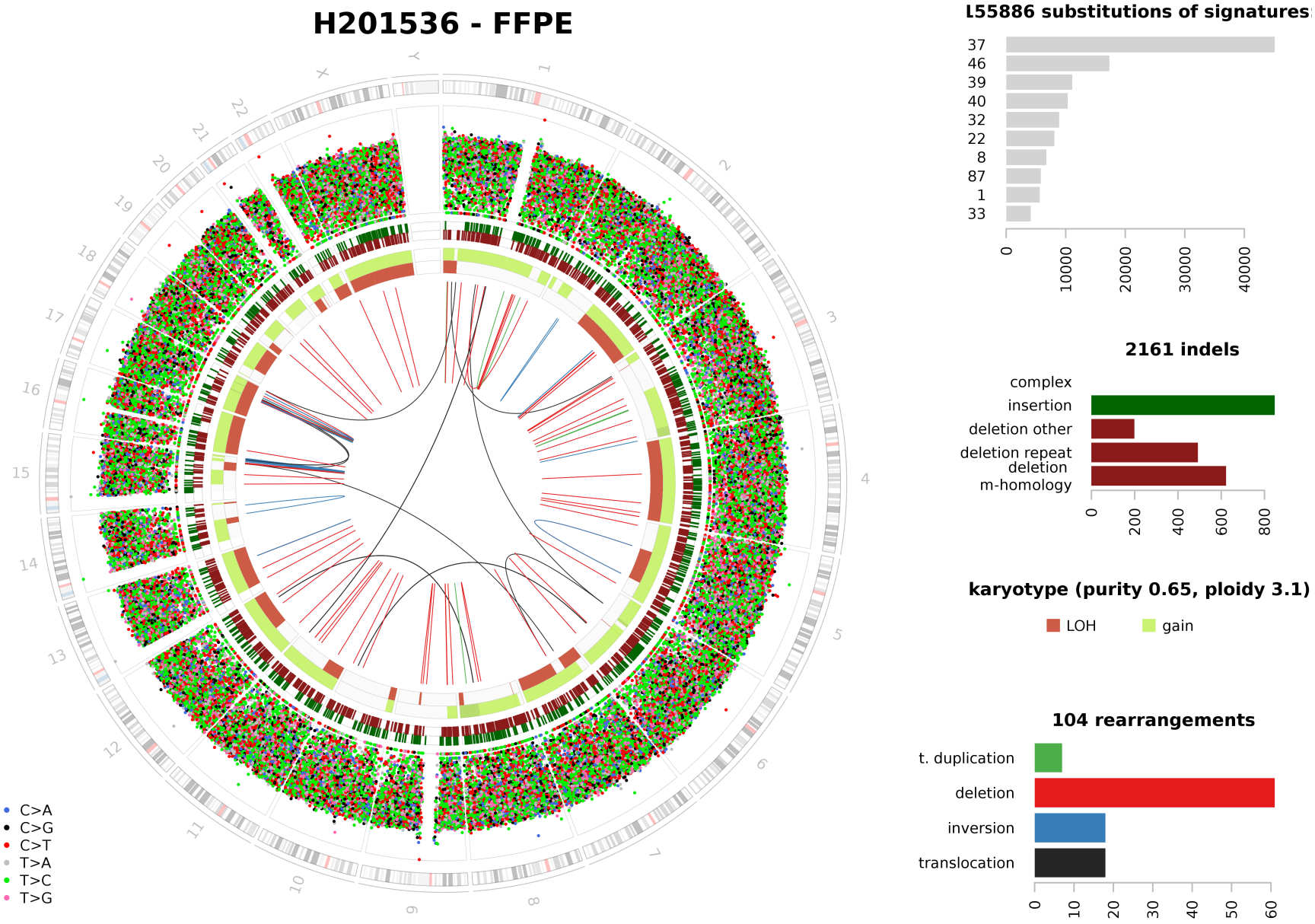


karyotype (purity 0.67, ploidy 3.4)

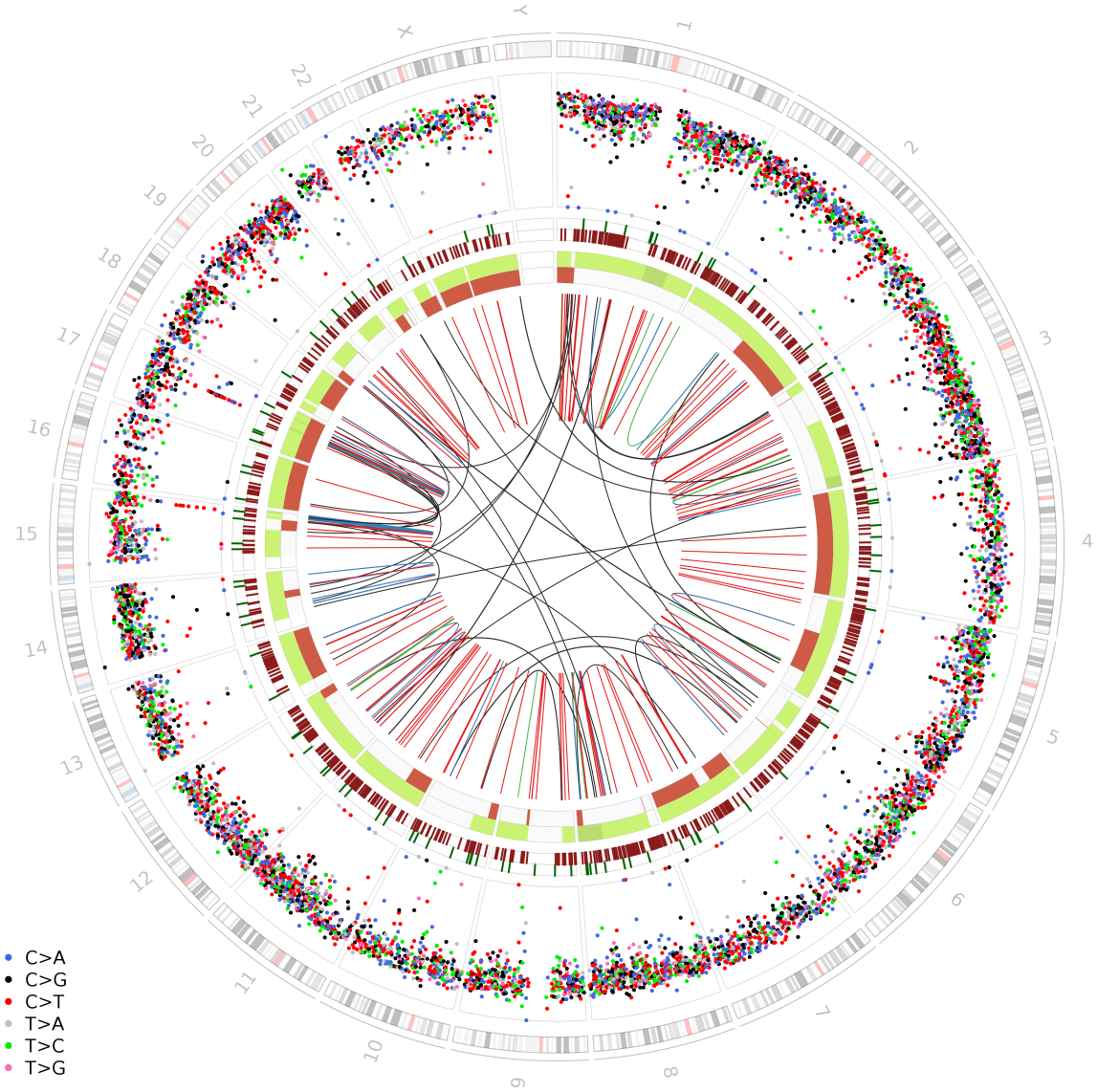


412 rearrangements

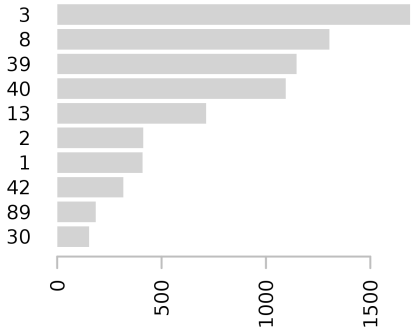



H201536 - FFPE
C>A
C>G
C>T
T>A
T>C
T>G
L55886 substitutions of signatures
37
46
39
40
32
22
8
87
1
33
0
10000
20000
30000
40000
2161 indels
complex
insertion
deletion other
deletion repeat
deletion
m-homology
0
200
400
600
800
karyotype (purity 0.65, ploidy 3.1)
LOH
gain
104 rearrangements
t. duplication
deletion
inversion
translocation
0
10
20
30
40
50
60

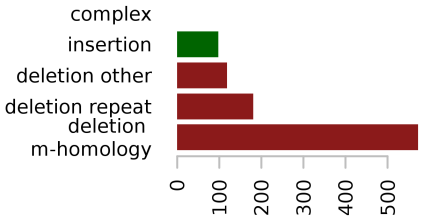
H201536 - Fresh Frozen



8367 substitutions of signatures:



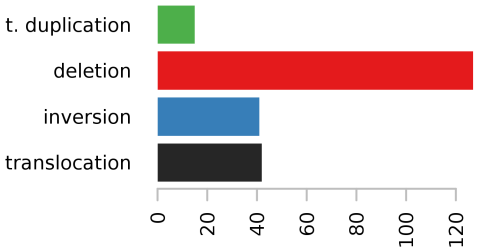
970 indels

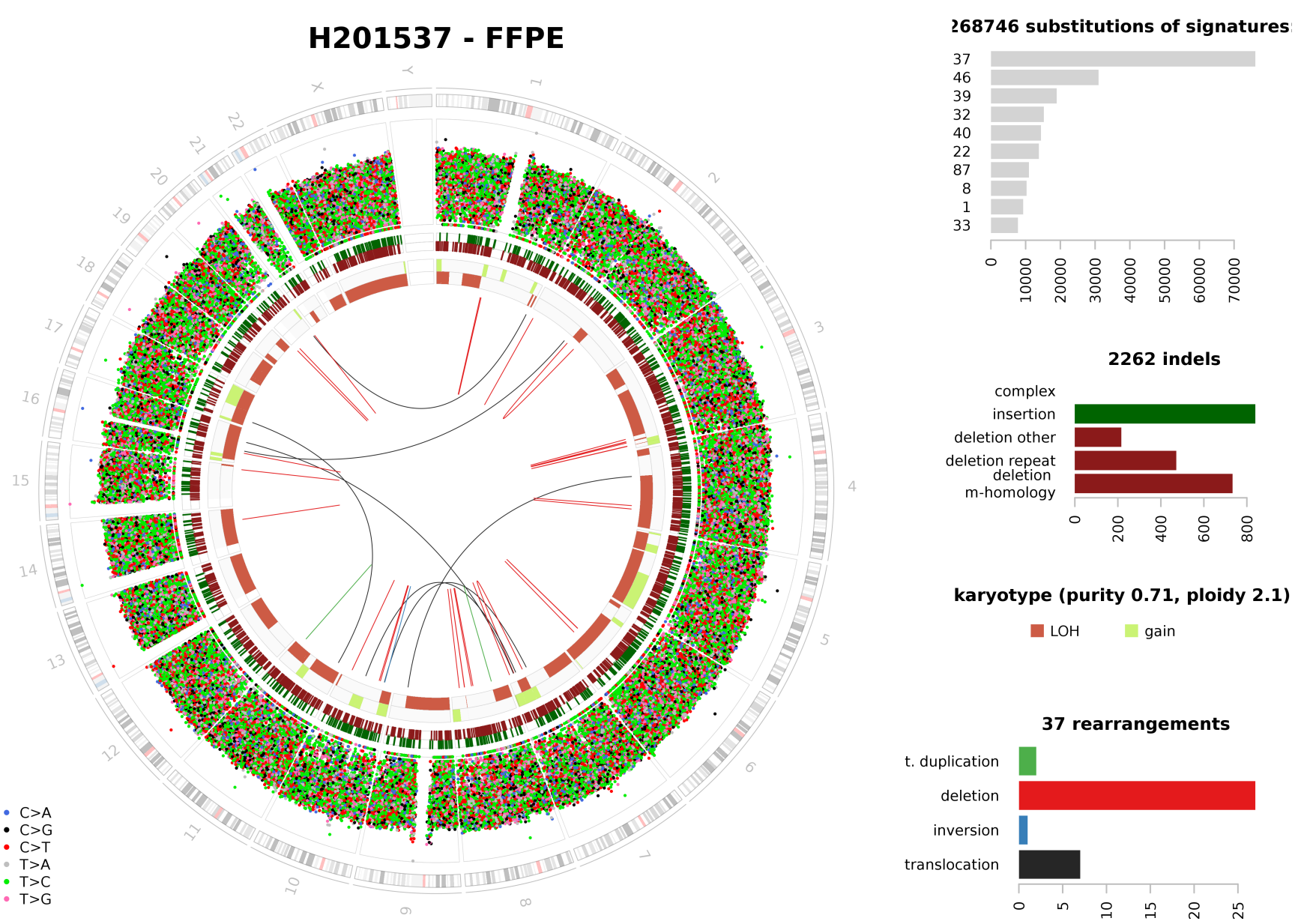


karyotype (purity 0.66, ploidy 3.2)

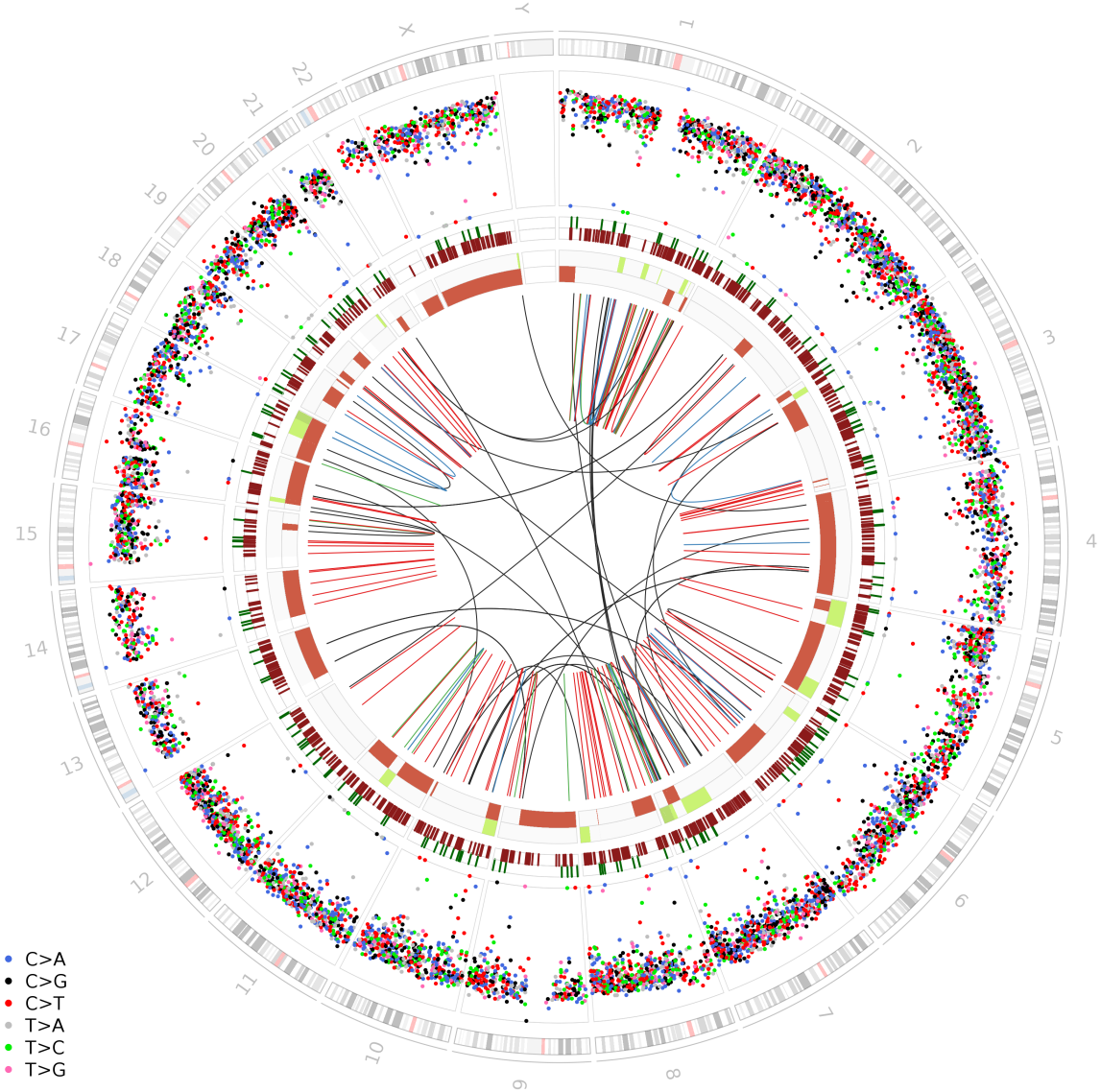


225 rearrangements

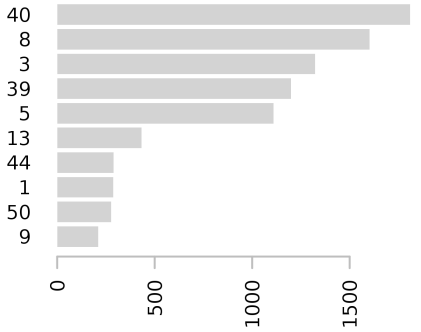



H201537 - FFPE
C>A
C>G
C>T
T>A
T>C
T>G
268746 substitutions of signatures
37
46
39
32
40
22
87
8
1
33
0
10000
20000
30000
40000
50000
60000
70000
2262 indels
complex
insertion
deletion other
deletion repeat
deletion
m-homology
0
200
400
600
800
karyotype (purity 0.71, ploidy 2.1)
LOH
gain
37 rearrangements
t. duplication
deletion
inversion
translocation
0
5
10
15
20
25

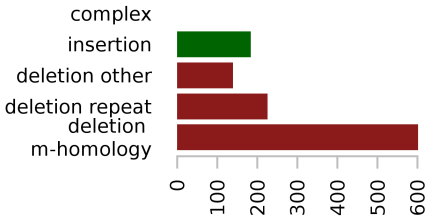
H201537 - Fresh Frozen



9009 substitutions of signatures:



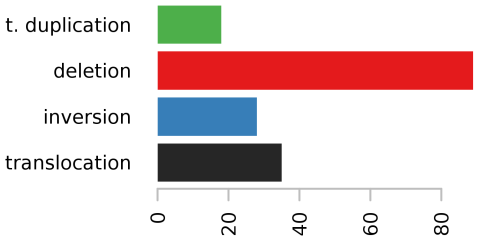
1151 indels

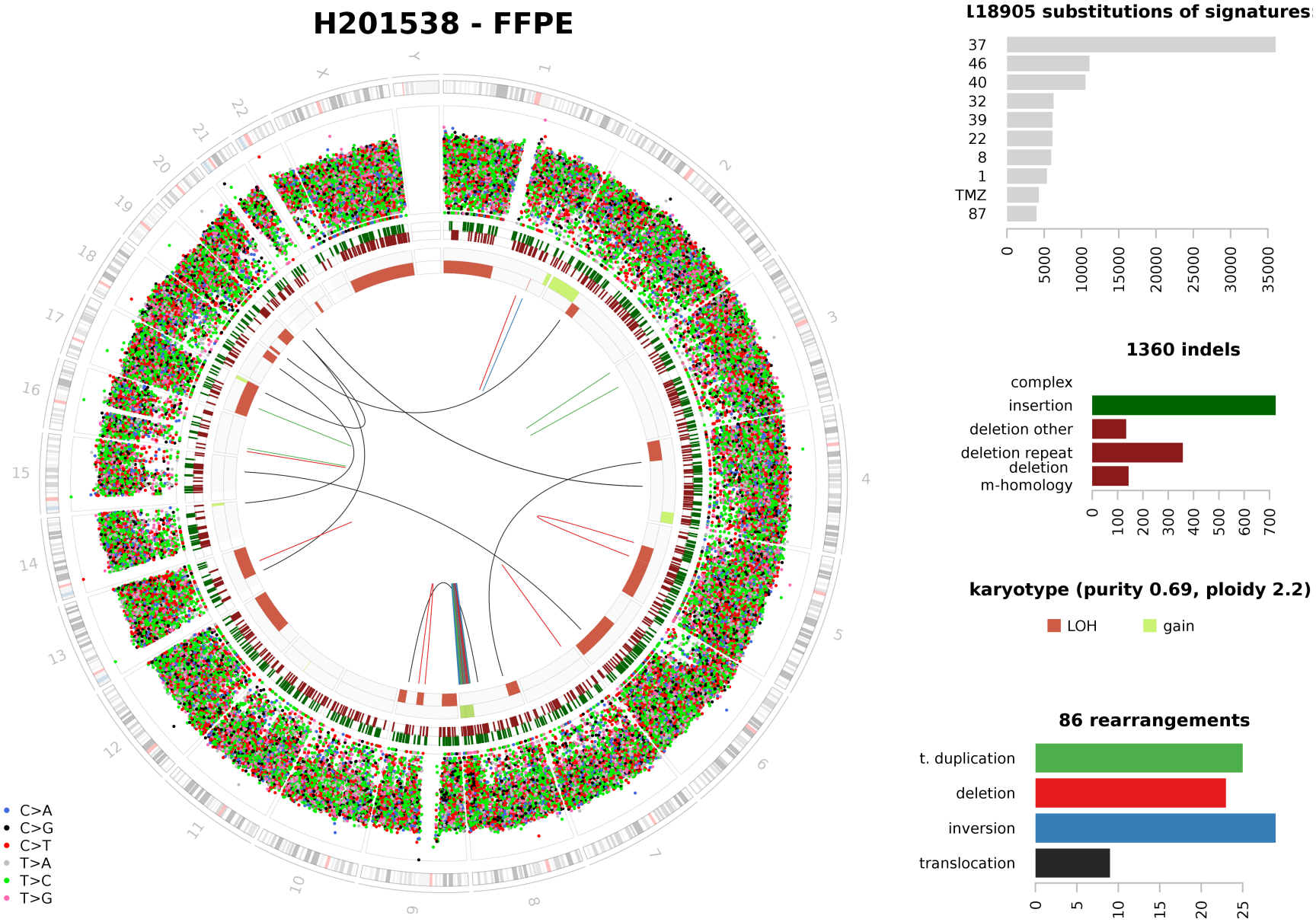


karyotype (purity 0.53, ploidy 2.1)

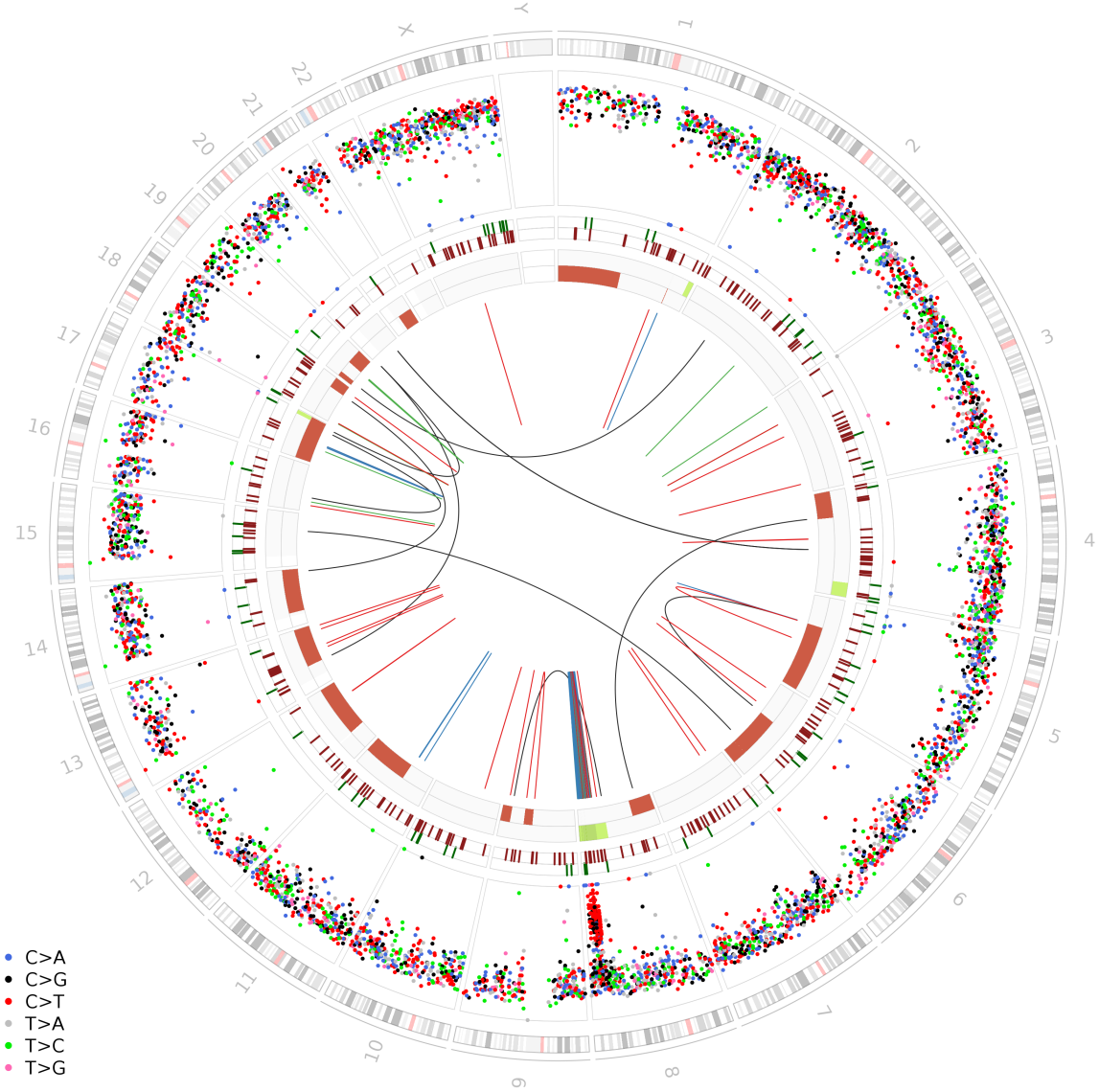


170 rearrangements

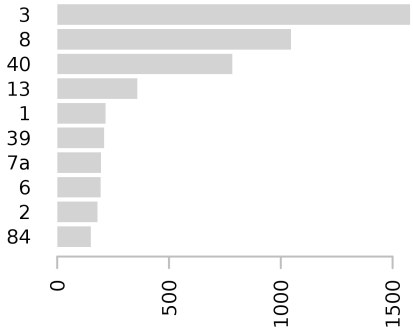



H201538 - FFPE
C>A
C>G
C>T
T>A
T>C
T>G
L18905 substitutions of signatures
37
46
40
32
39
22
8
1
TMZ
87
0
5000
10000
15000
20000
25000
30000
35000
1360 indels
complex
insertion
deletion other
deletion repeat
deletion
m-homology
0
100
200
300
400
500
600
700
karyotype (purity 0.69, ploidy 2.2)
LOH
gain
86 rearrangements
t. duplication
deletion
inversion
translocation
0
5
10
15
20
25

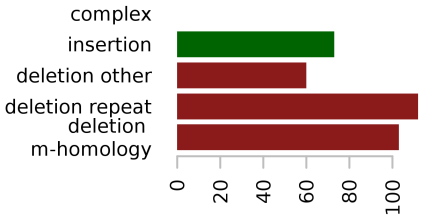
H201538 - Fresh Frozen



5904 substitutions of signatures:



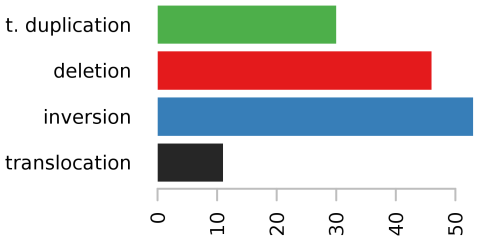
348 indels

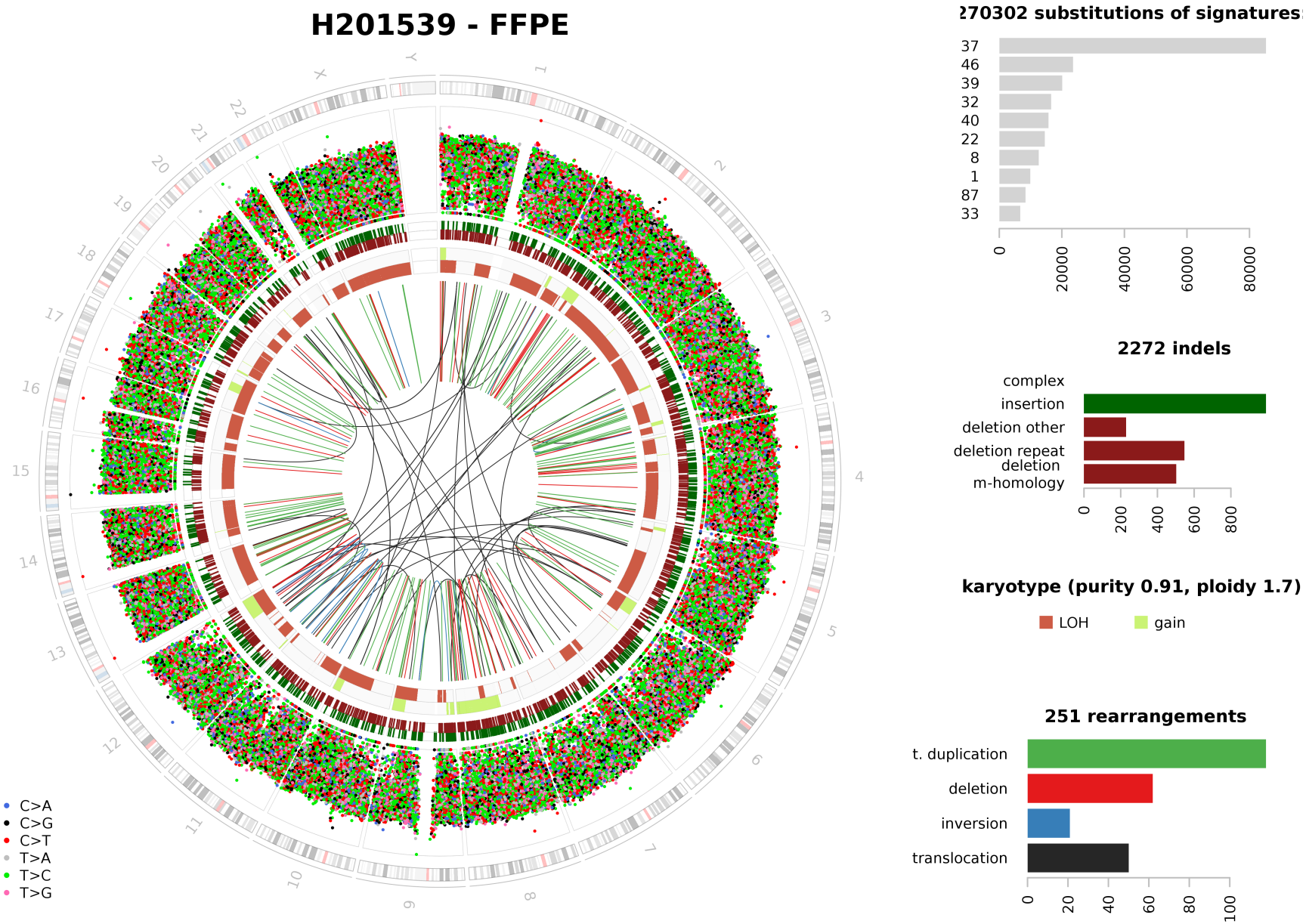


karyotype (purity 0.51, ploidy 1.9)

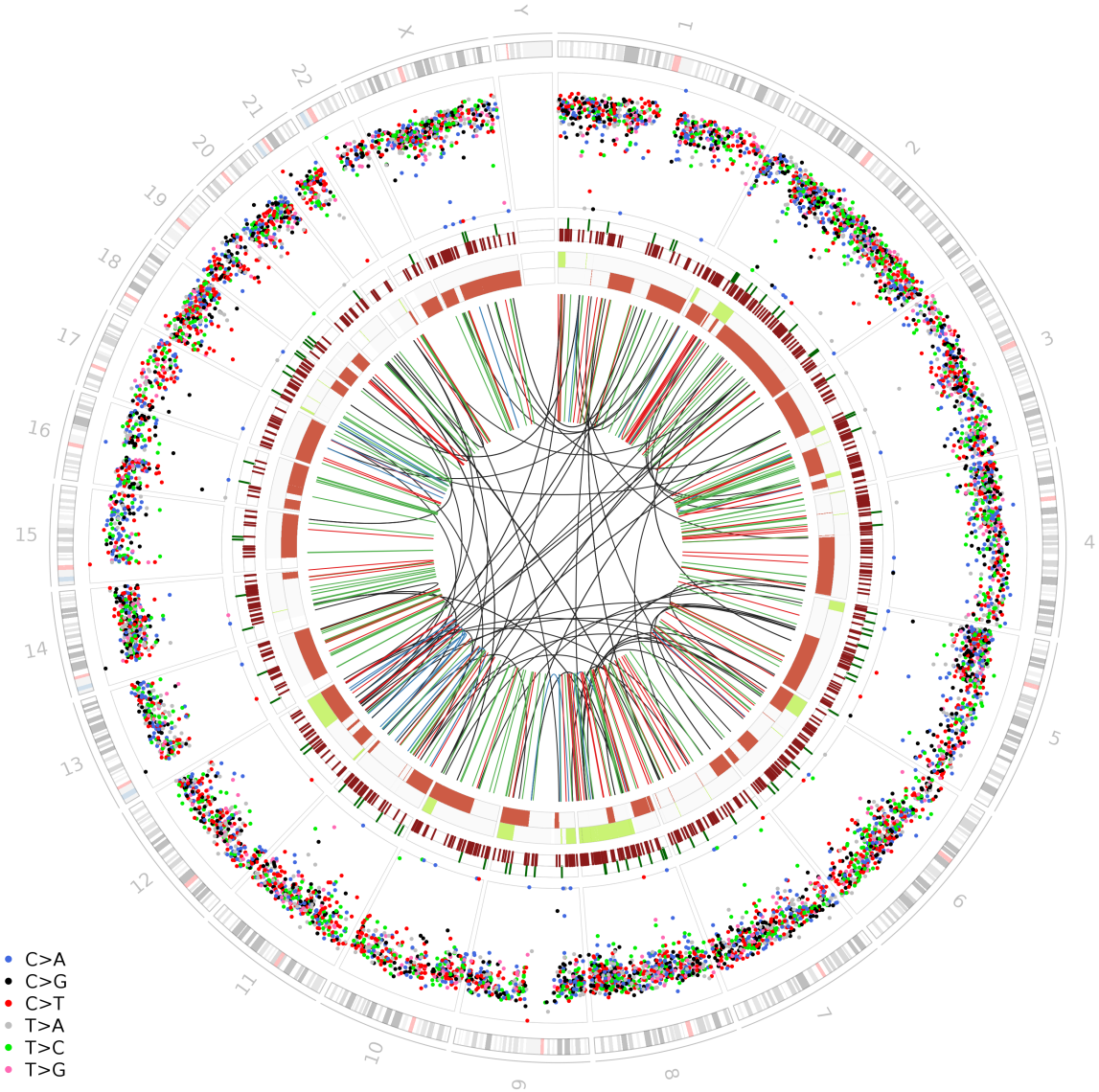


140 rearrangements

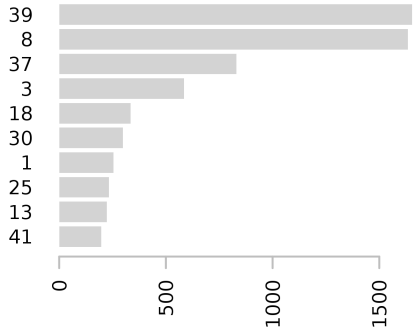



H201539 - FFPE
C>A
C>G
C>T
T>A
T>C
T>G
270302 substitutions of signatures
37
46
39
32
40
22
8
1
87
33
0
20000
40000
60000
80000
2272 indels
complex
insertion
deletion other
deletion repeat
deletion
m-homology
0
200
400
600
800
karyotype (purity 0.91, ploidy 1.7)
LOH
gain
251 rearrangements
t. duplication
deletion
inversion
translocation
0
20
40
60
80
100

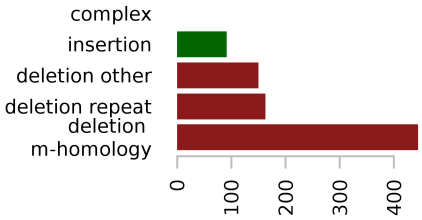
H201539 - Fresh Frozen



7783 substitutions of signatures:



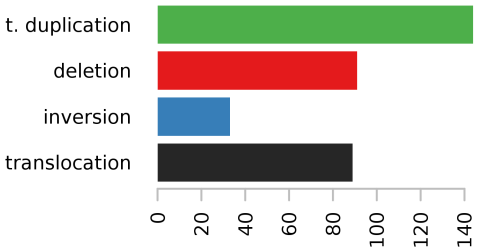
849 indels

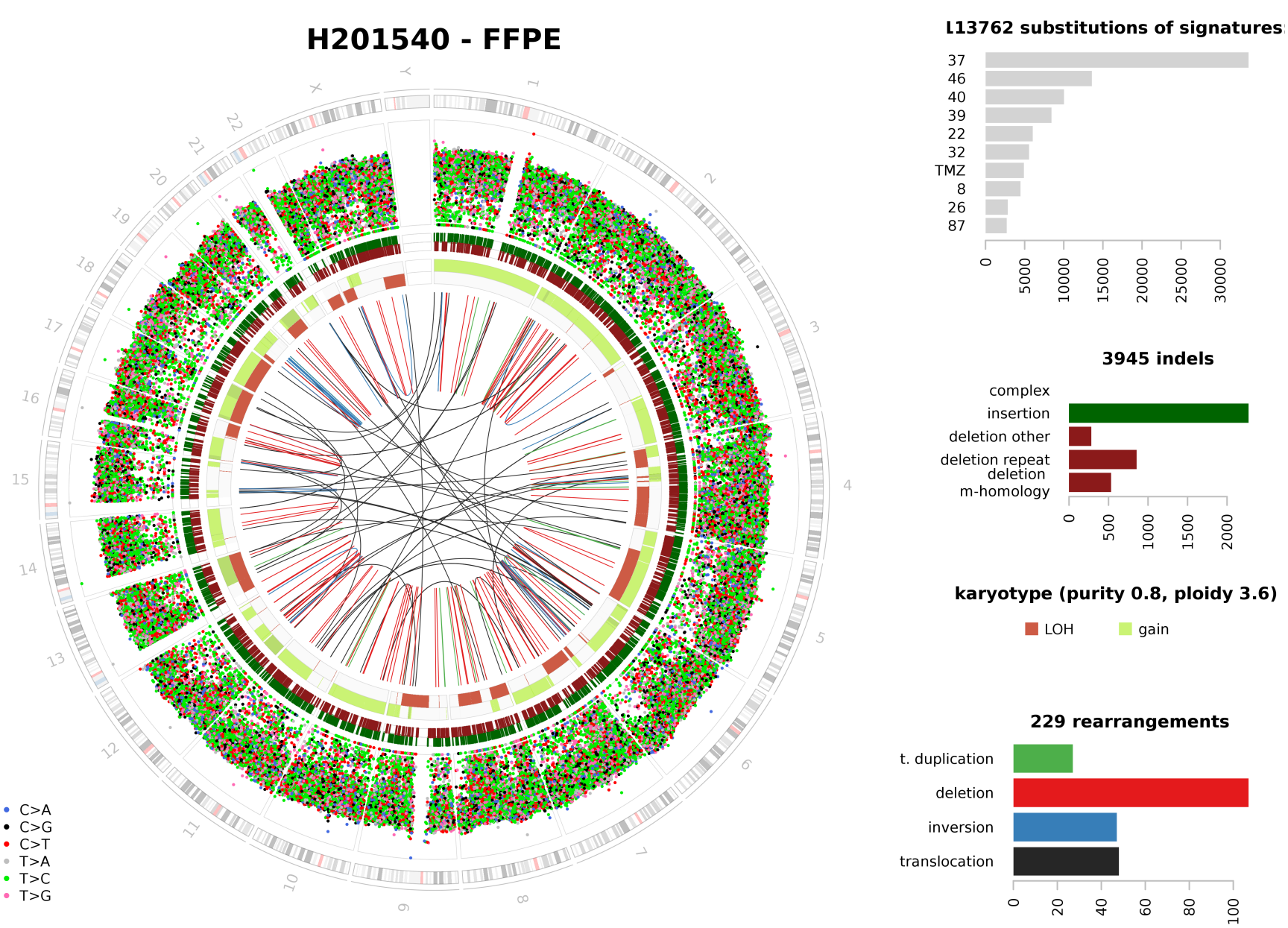


karyotype (purity 0.96, ploidy 1.8)

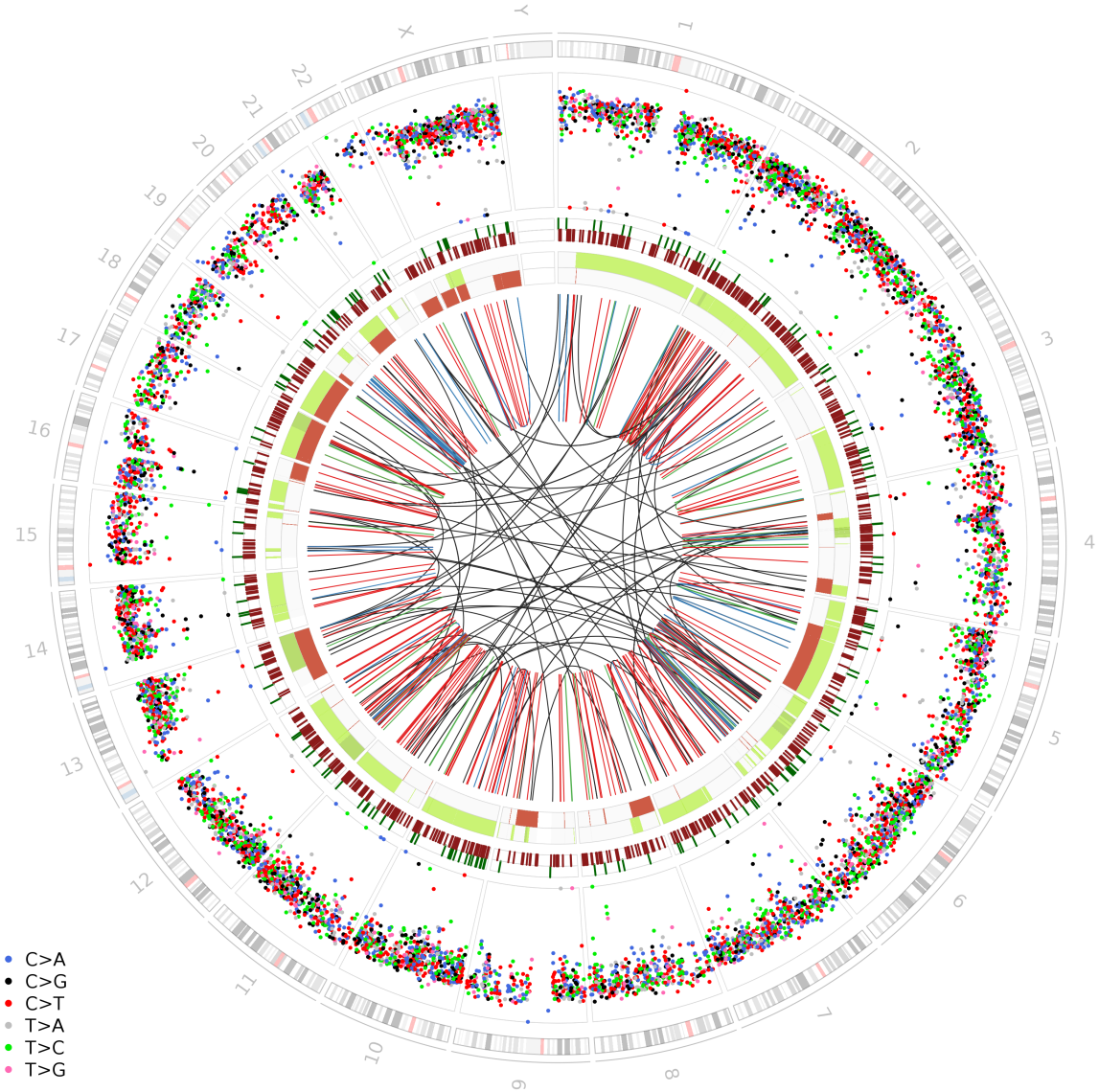


357 rearrangements

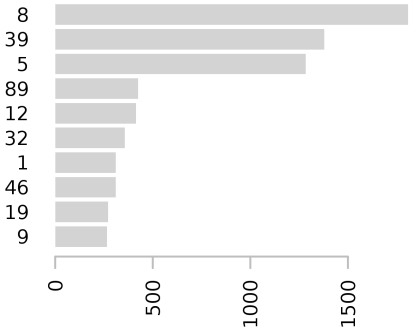



H201540 - FFPE
C>A
C>G
C>T
T>A
T>C
T>G
L13762 substitutions of signatures
37
46
40
39
22
32
TMZ
8
26
87
0
5000
10000
15000
20000
25000
30000
3945 indels
complex
insertion
deletion other
deletion repeat
deletion
m-homology
0
500
1000
1500
2000
karyotype (purity 0.8, ploidy 3.6)
LOH
gain
229 rearrangements
t. duplication
deletion
inversion
translocation
0
20
40
60
80
100

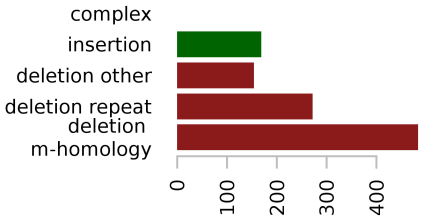
H201540 - Fresh Frozen



9041 substitutions of signatures:



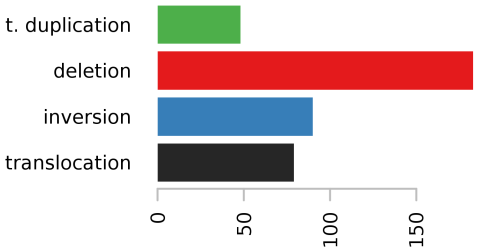
1079 indels



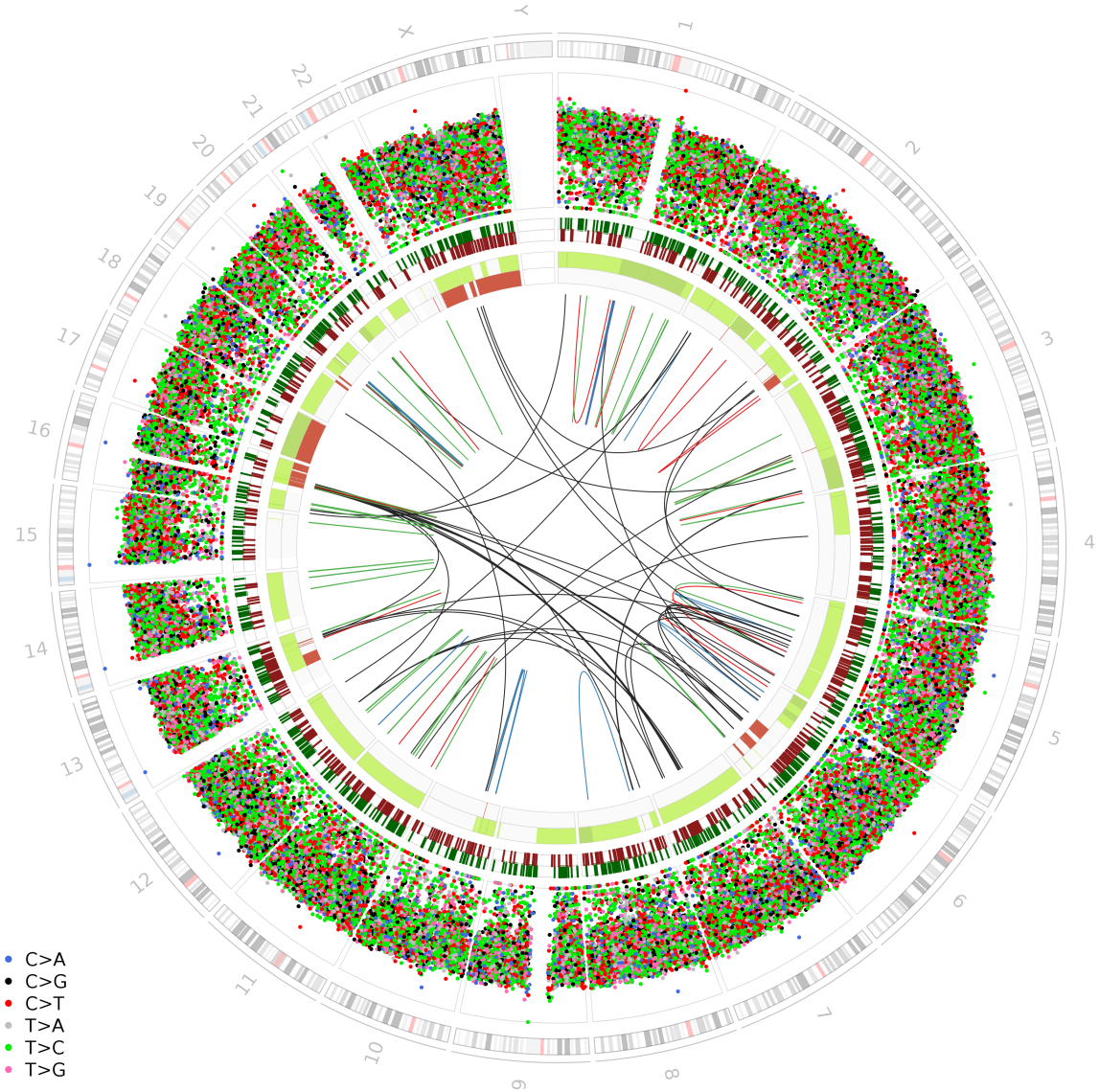
karyotype (purity 0.84, ploidy 2.9)

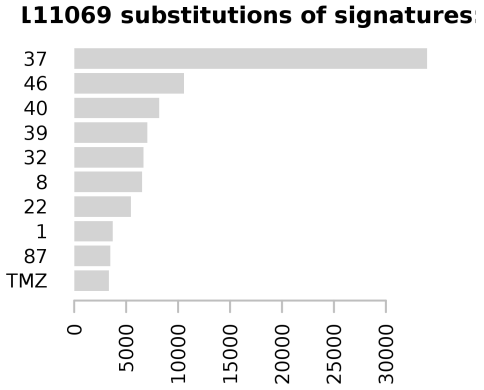


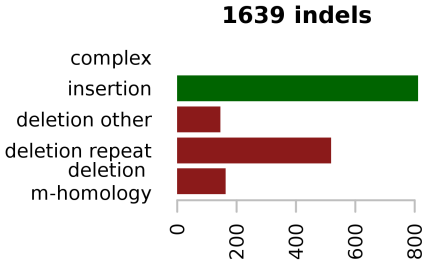
400 rearrangements



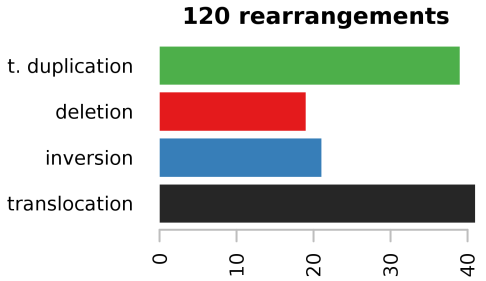
H201542 - FFPE



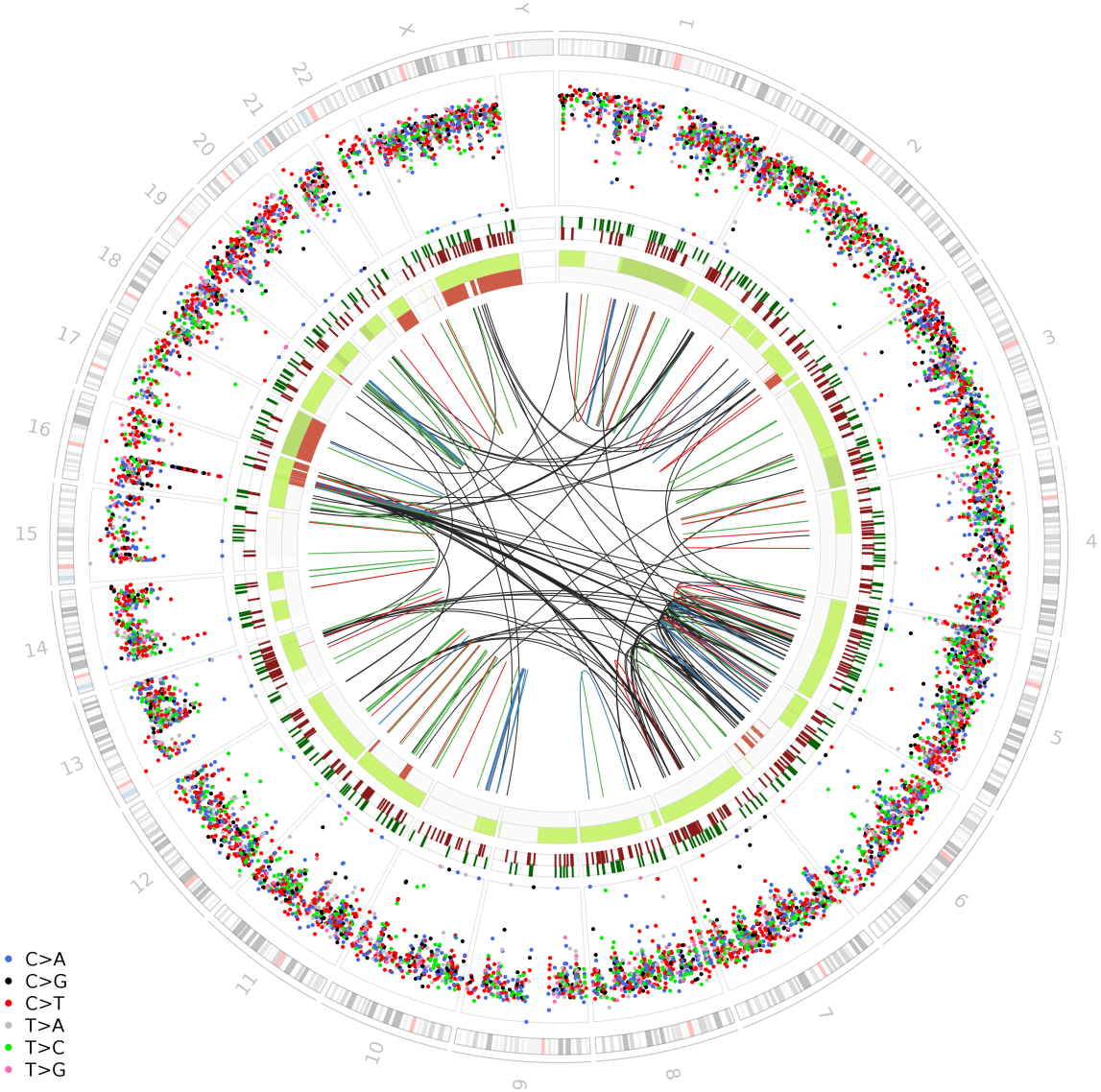




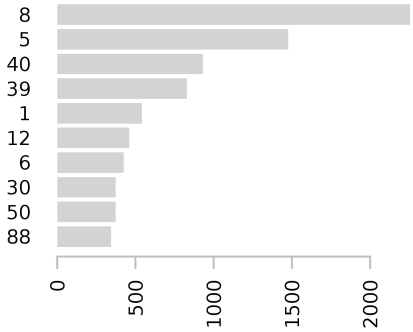




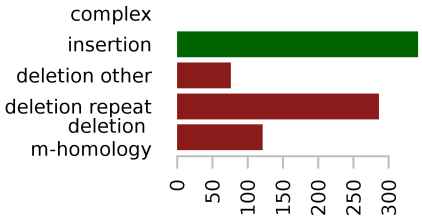
H201542 - Fresh Frozen



9694 substitutions of signatures:



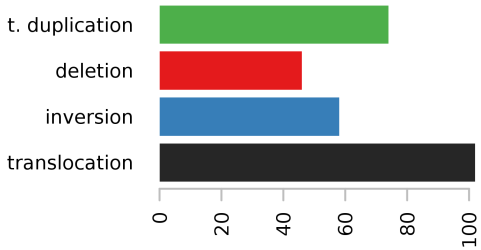
825 indels



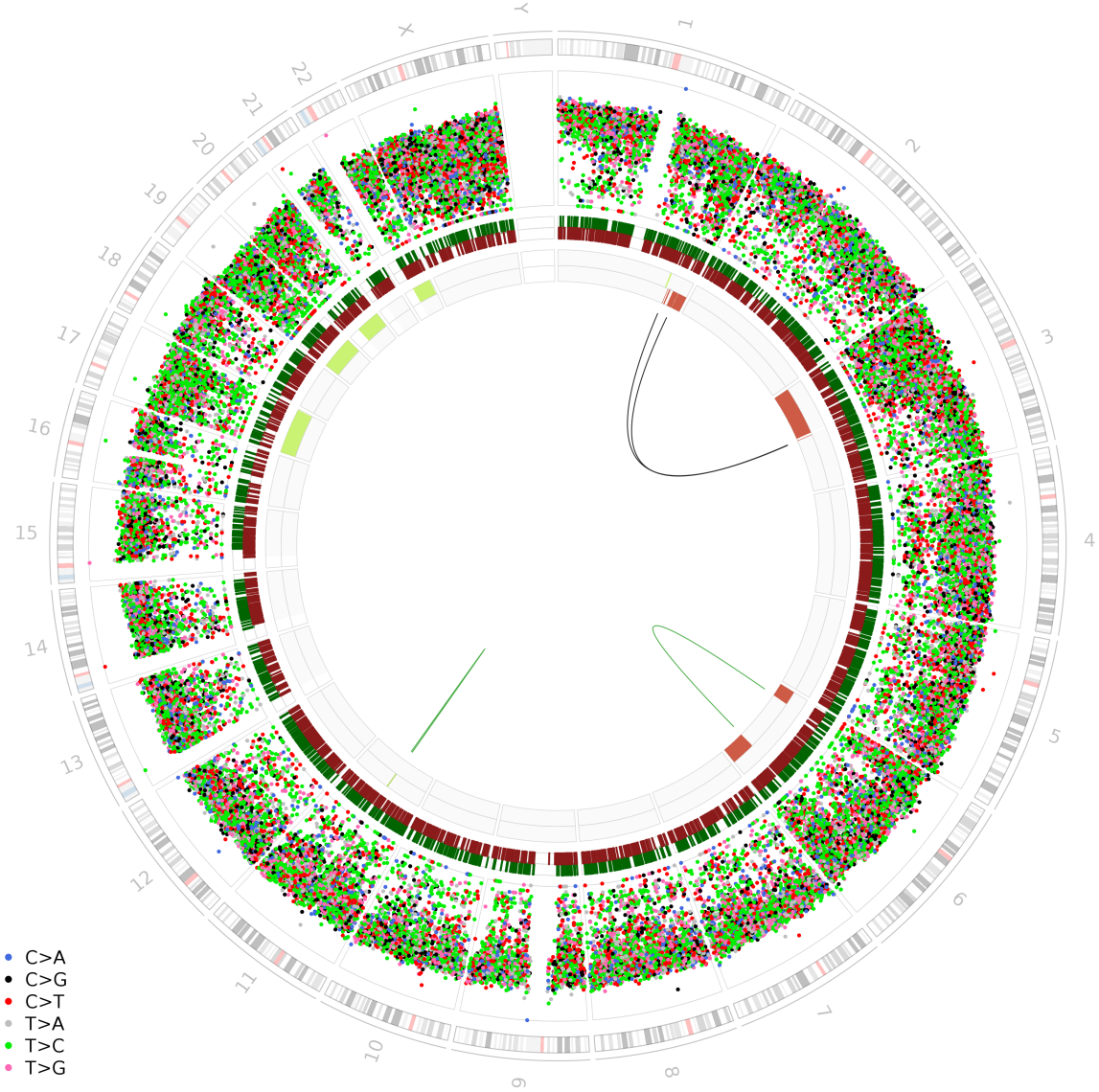
karyotype (purity 0.63, ploidy 3.8)



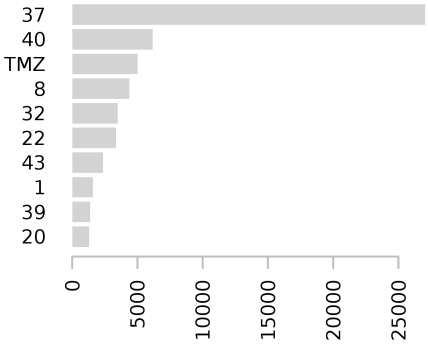
280 rearrangements



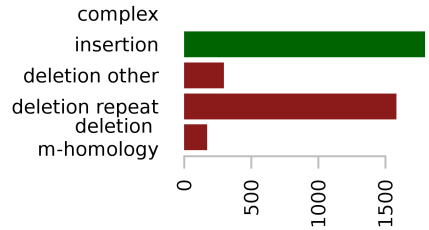
H203495 - FFPE



66971 substitutions of signatures:



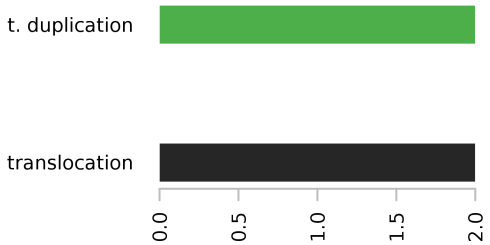
3848 indels



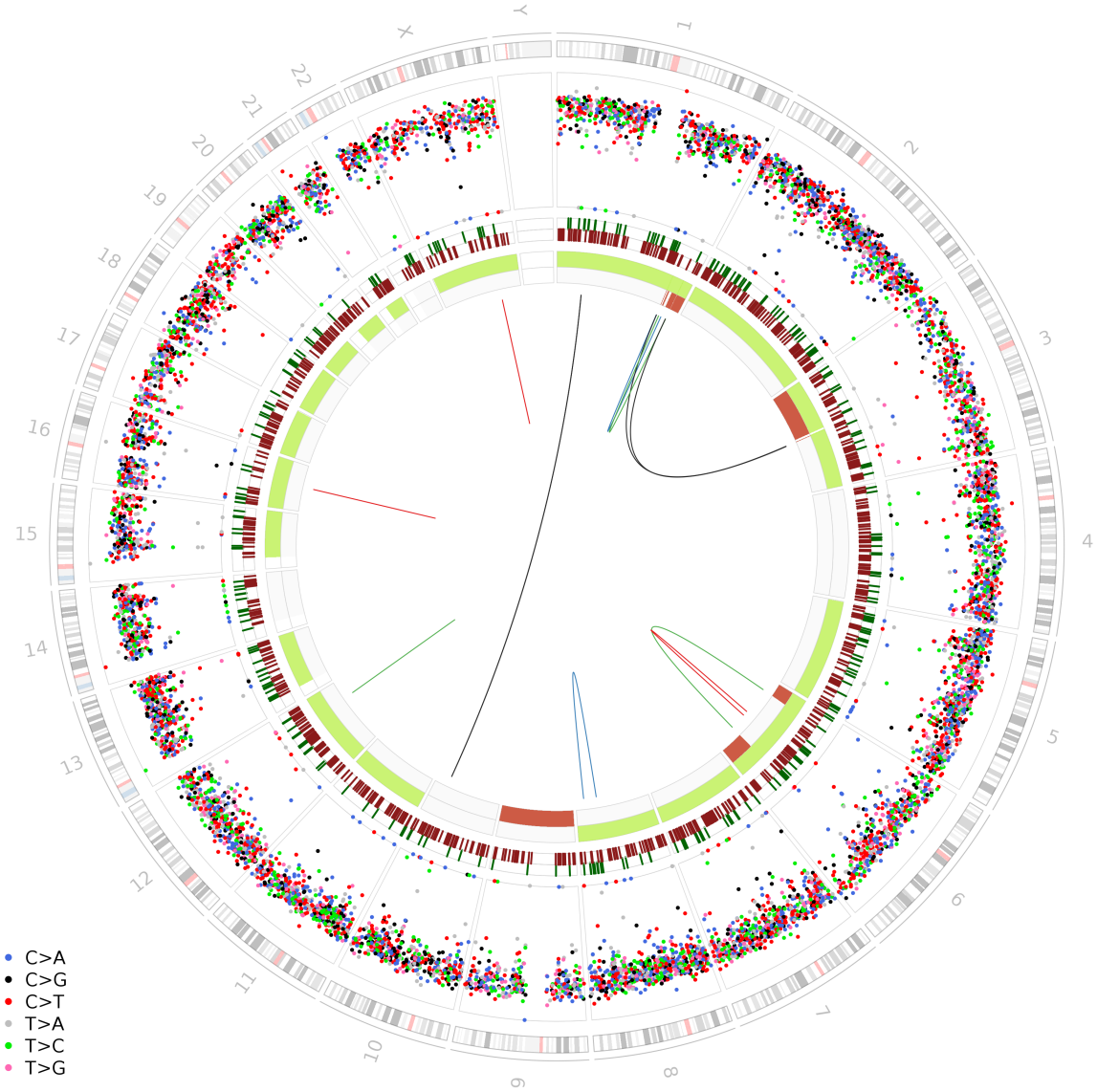
karyotype (purity 0.79, ploidy 3.8)



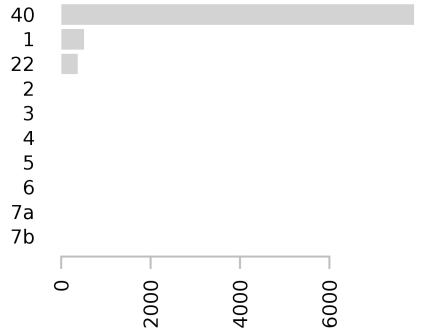
4 rearrangements



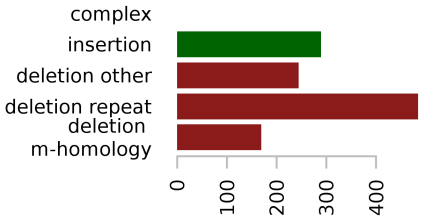
H203495 - Fresh Frozen



8985 substitutions of signatures:



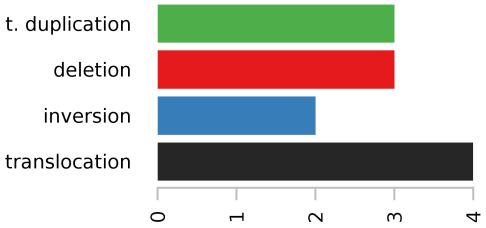
1187 indels



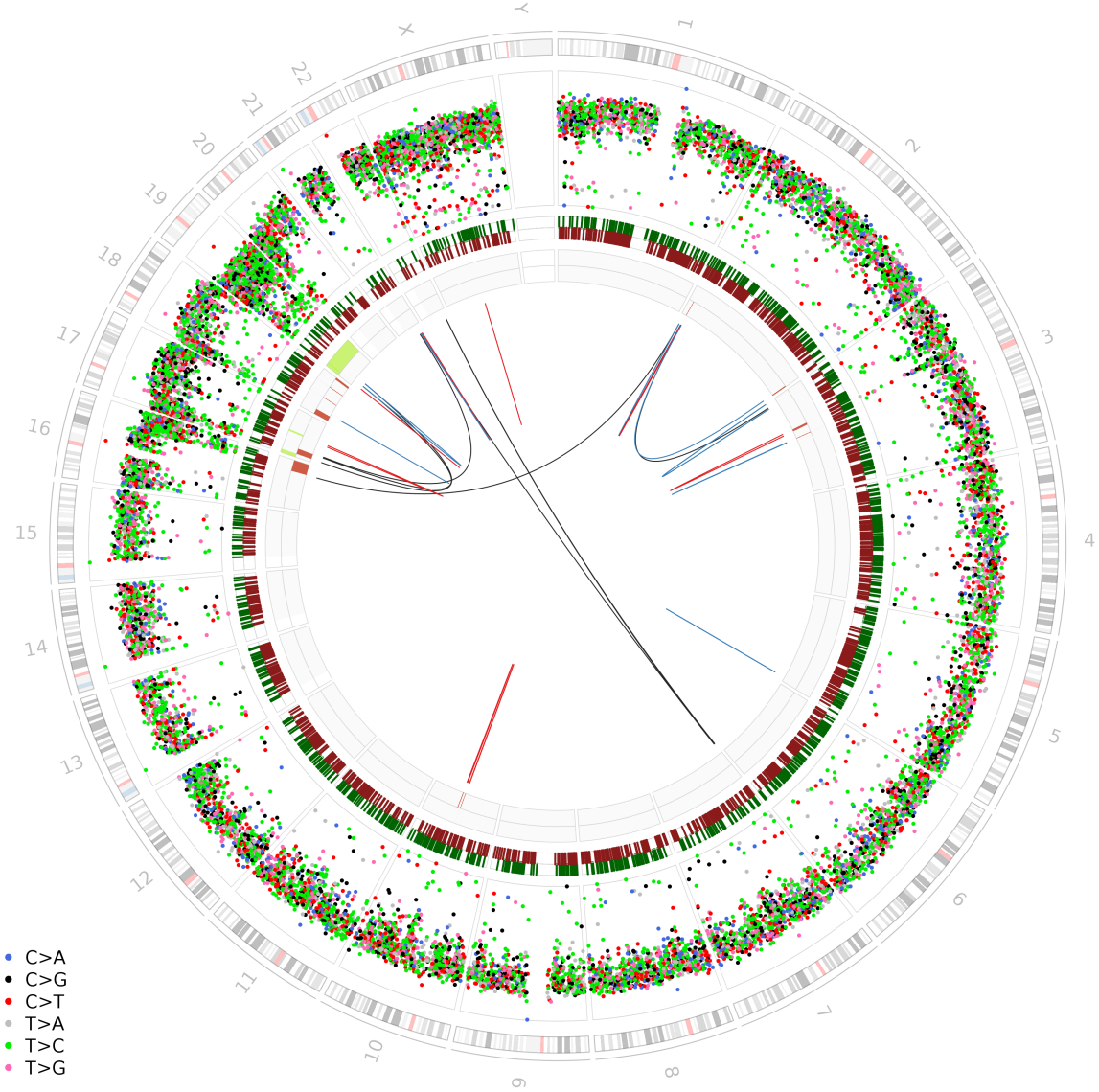
karyotype (purity 0.54, ploidy 3.4)



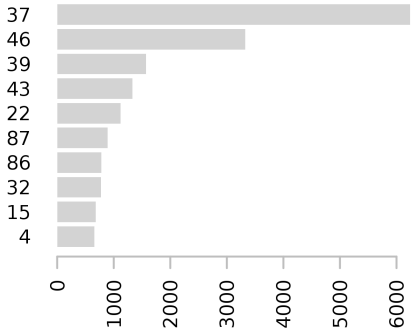
12 rearrangements



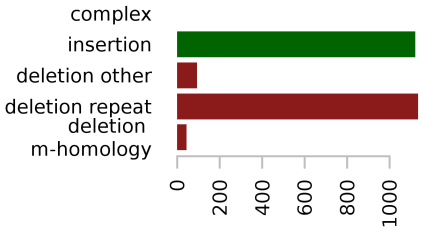
H203498 - FFPE



20788 substitutions of signatures:



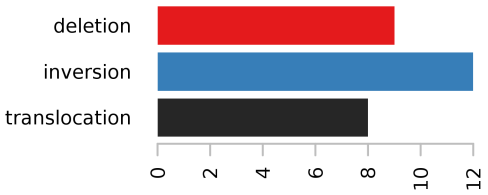
2393 indels



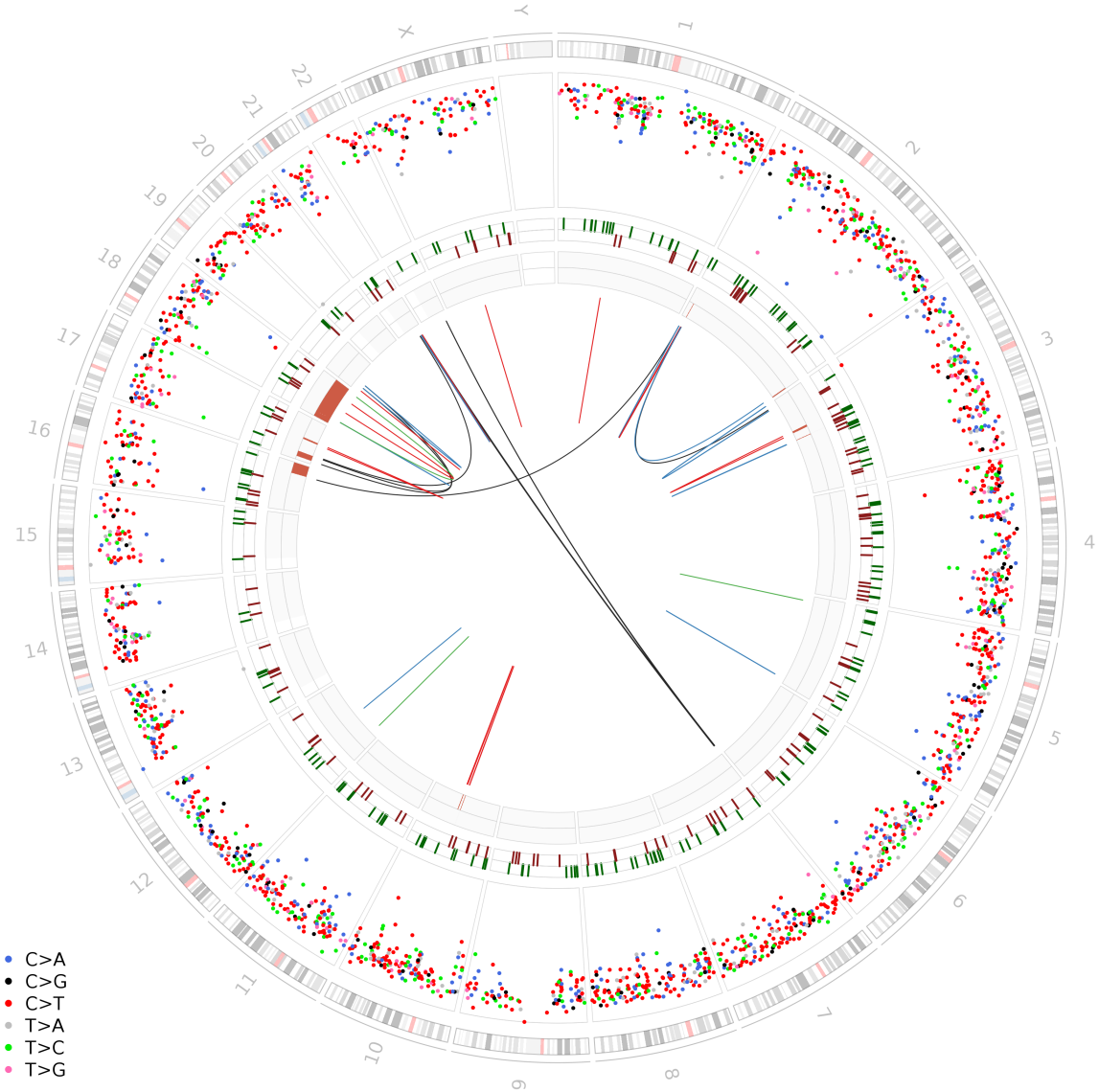
karyotype (purity 0.99, ploidy 2.6)



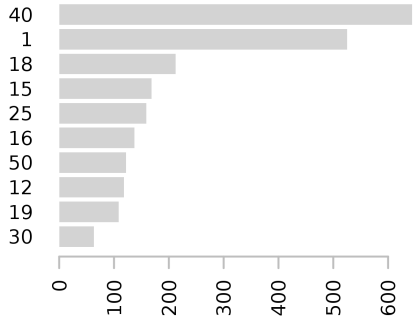
29 rearrangements



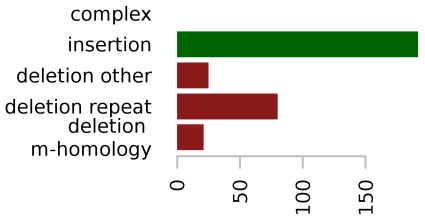
H203498 - Fresh Frozen



2358 substitutions of signatures:



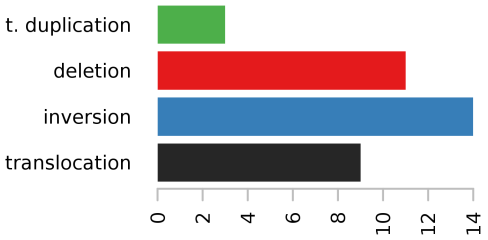
318 indels



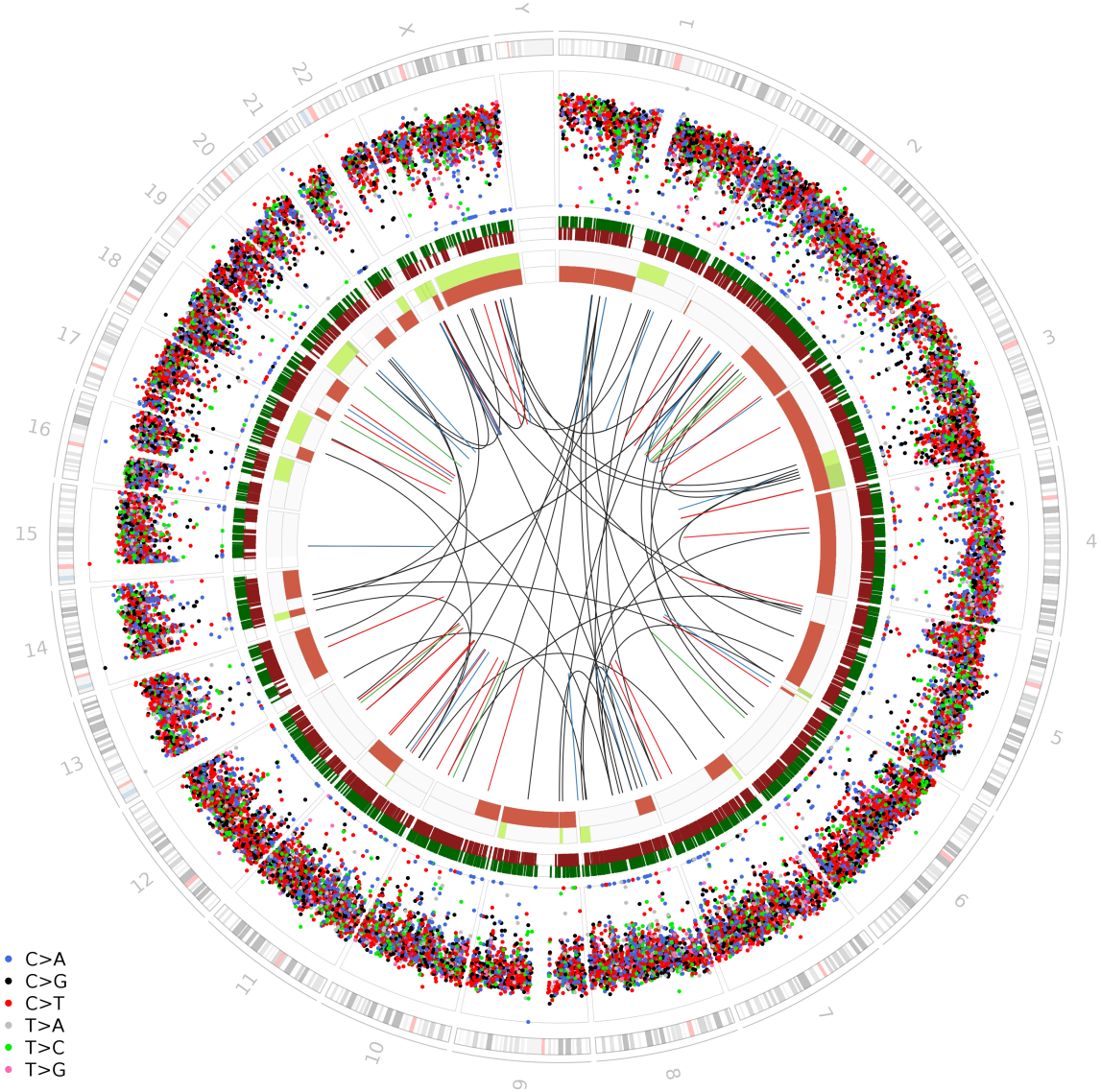
karyotype (purity 0.48, ploidy 2.1)



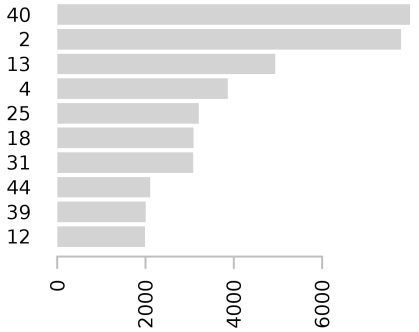
37 rearrangements



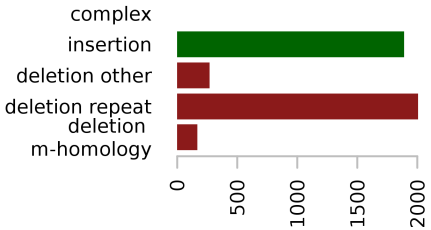
H203499 - FFPE



41045 substitutions of signatures:



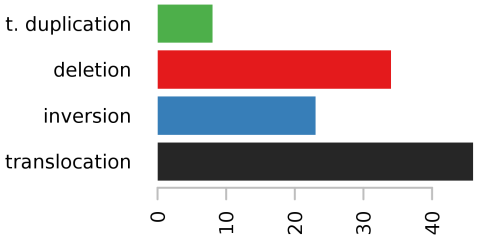
4335 indels



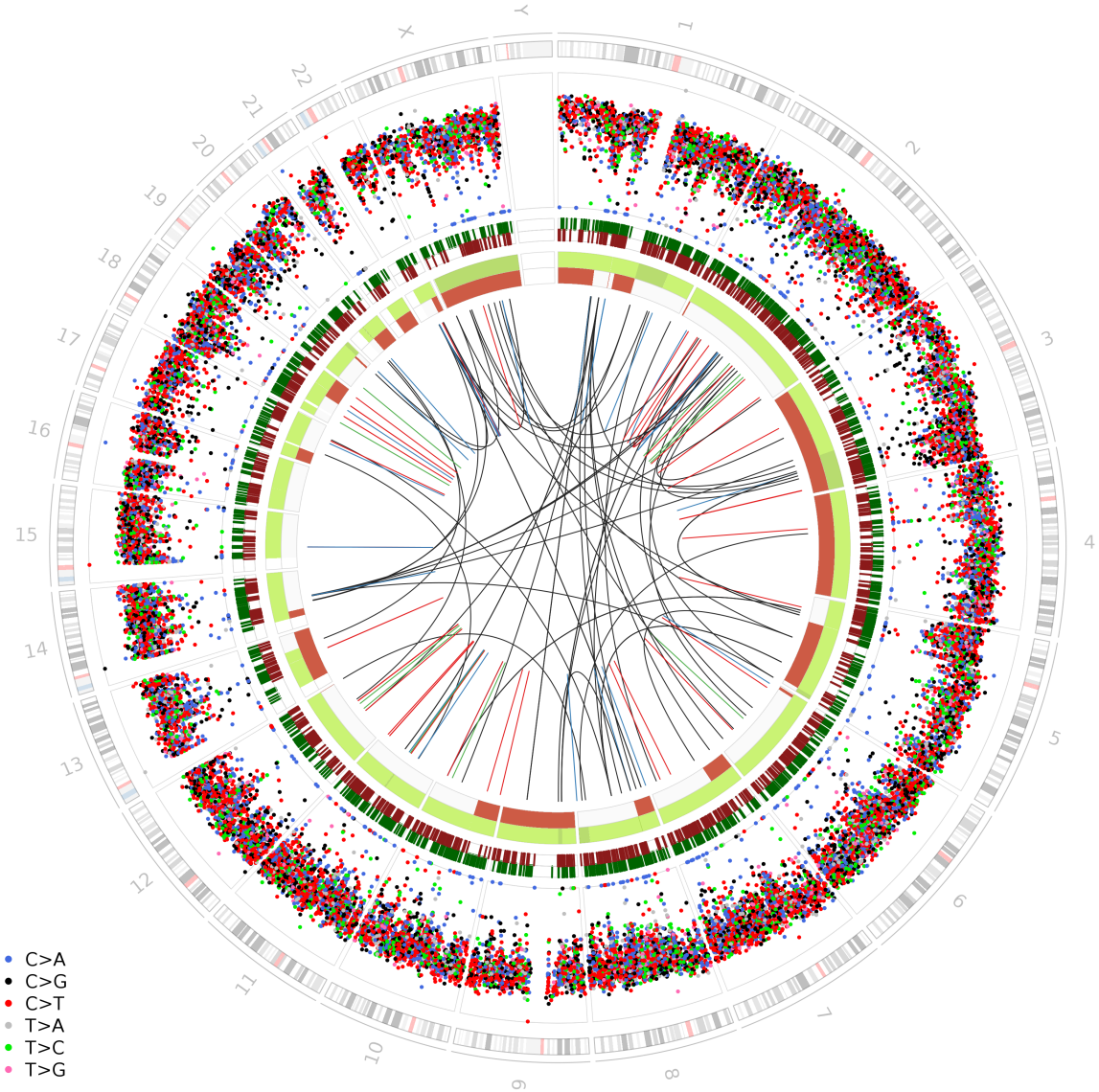
karyotype (purity 0.7, ploidy 2.1)



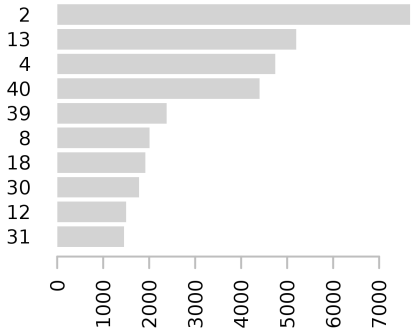
111 rearrangements



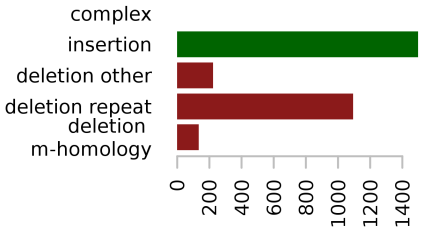
H203499 - Fresh Frozen



41115 substitutions of signatures:



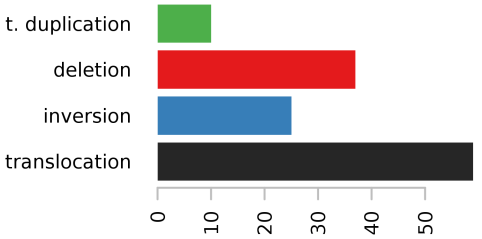
2948 indels



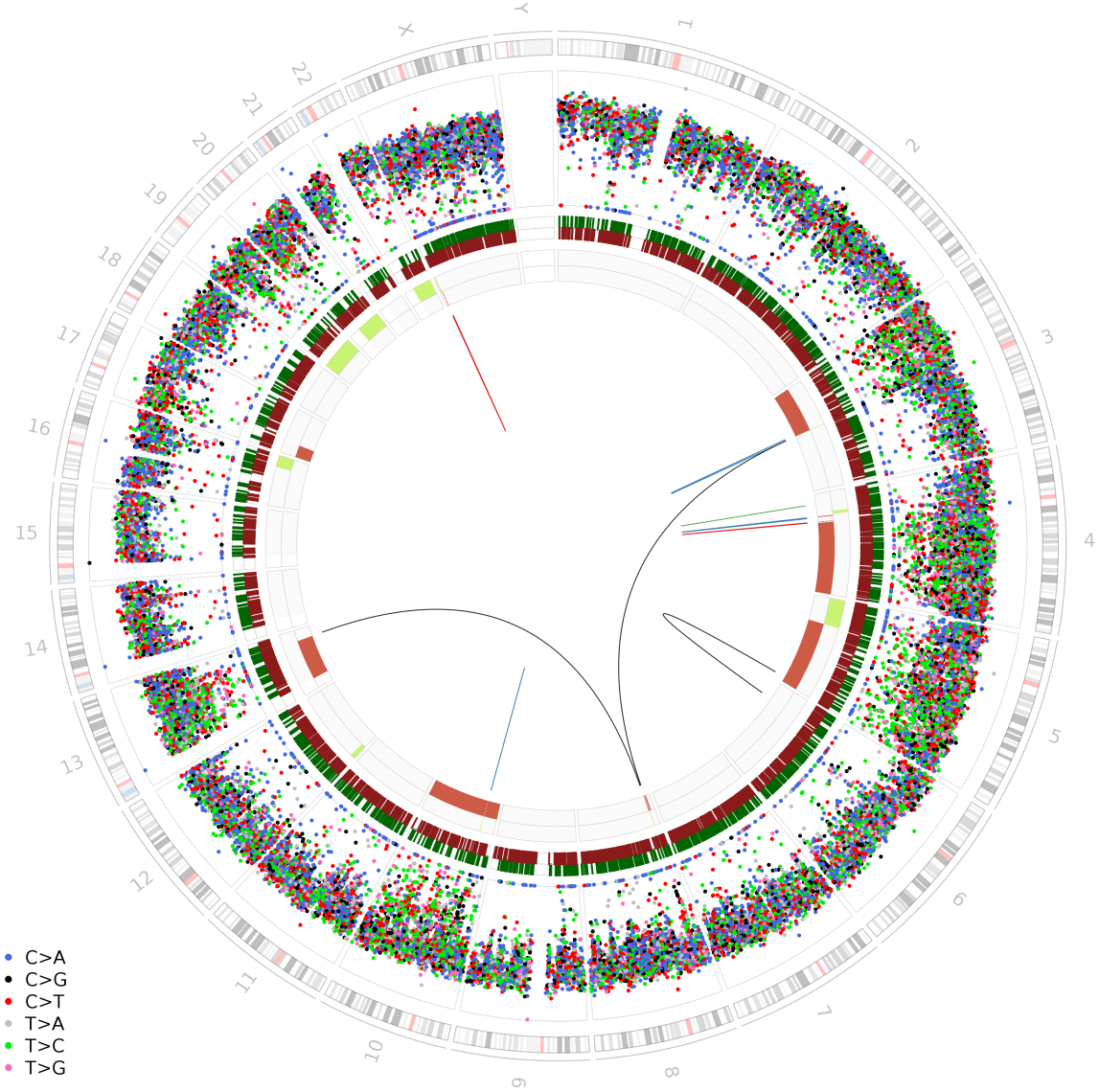
karyotype (purity 0.65, ploidy 3.4)



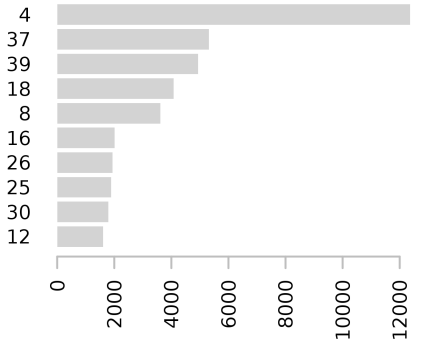
131 rearrangements



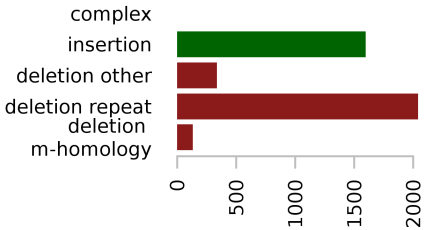
H203500 - FFPE



50011 substitutions of signatures:



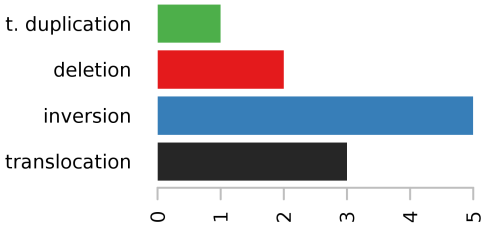
4107 indels



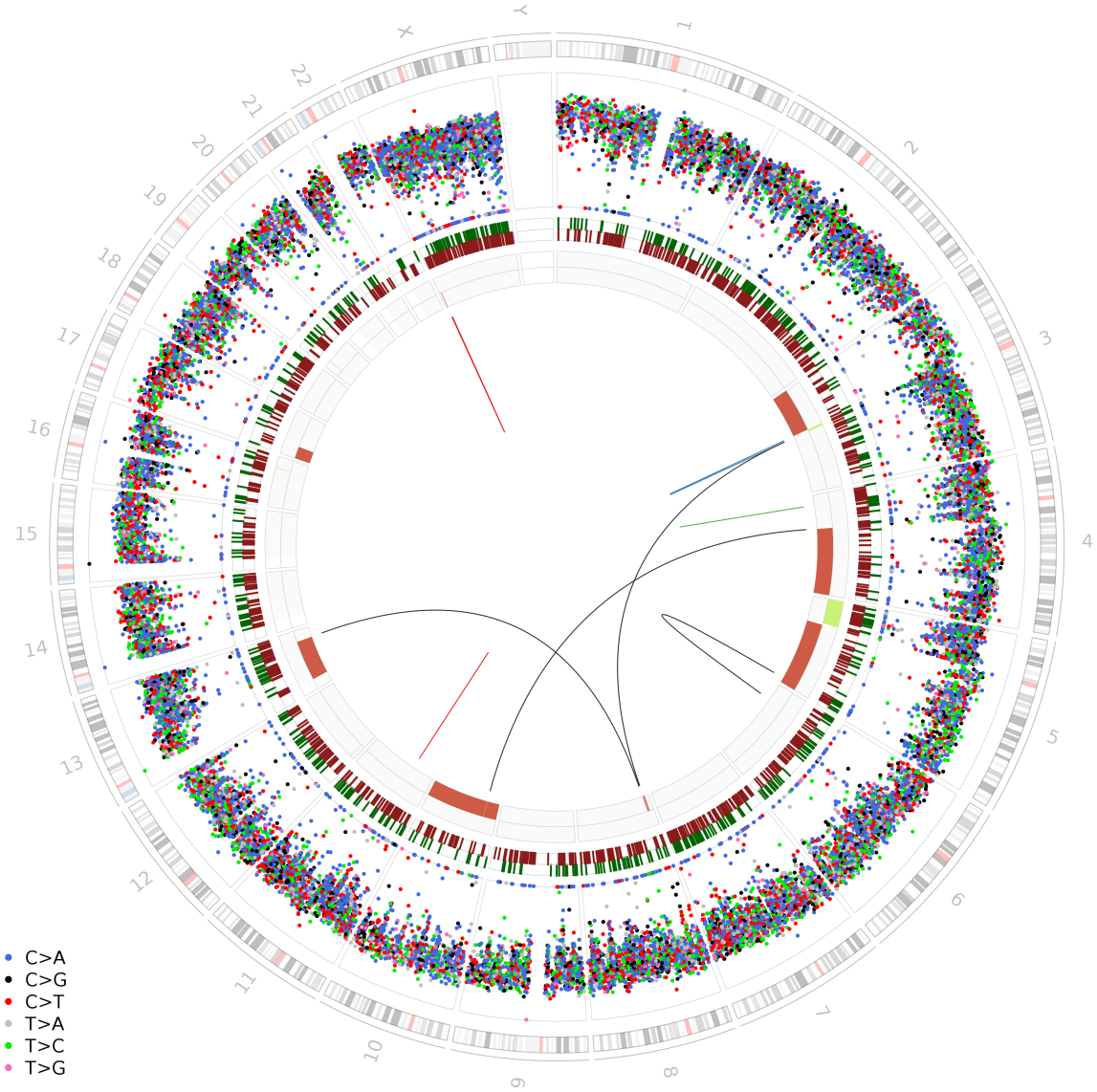
karyotype (purity 0.95, ploidy 2.7)



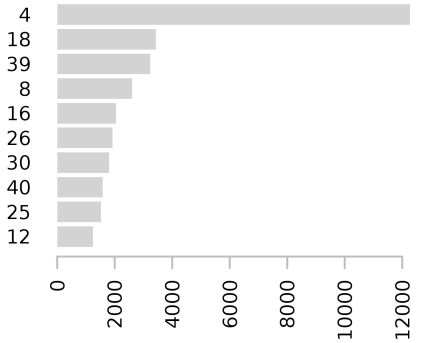
11 rearrangements



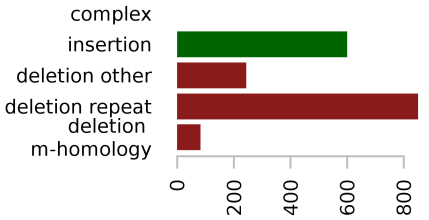
H203500 - Fresh Frozen



37039 substitutions of signatures:



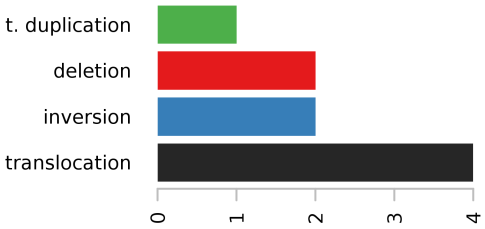
1777 indels



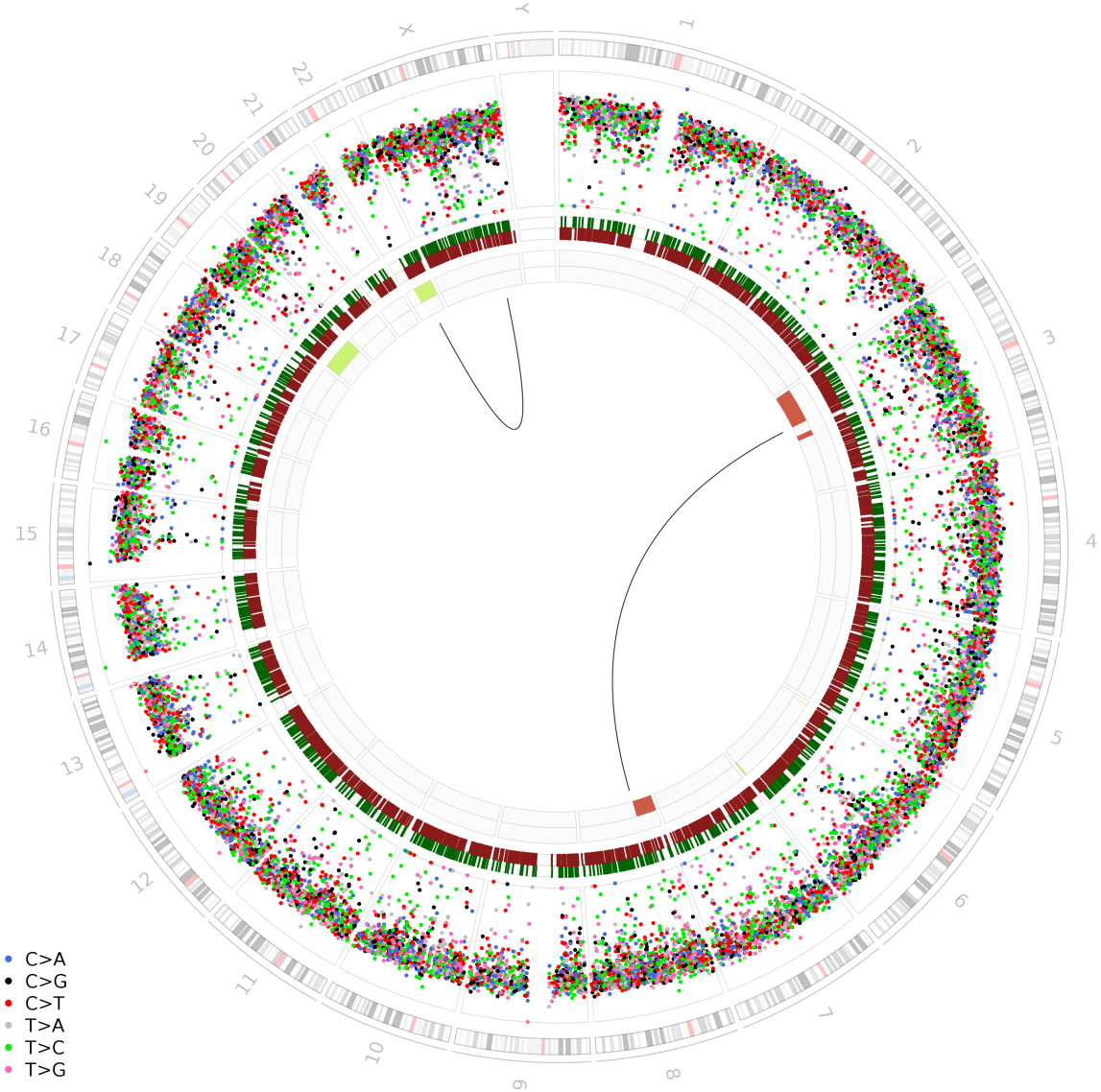
karyotype (purity 0.98, ploidy 1.9)



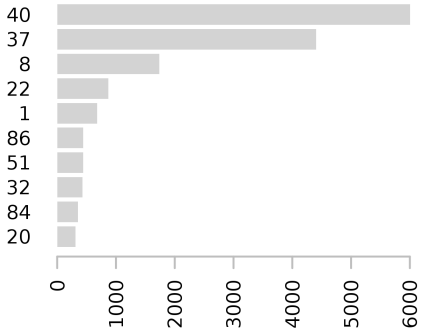
9 rearrangements



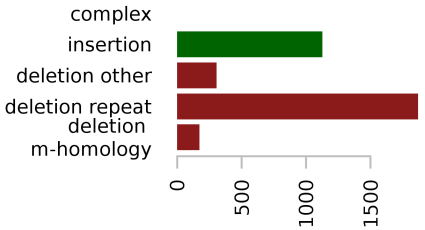
H203501 - FFPE



18259 substitutions of signatures:



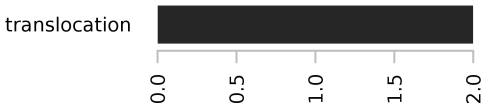
3476 indels



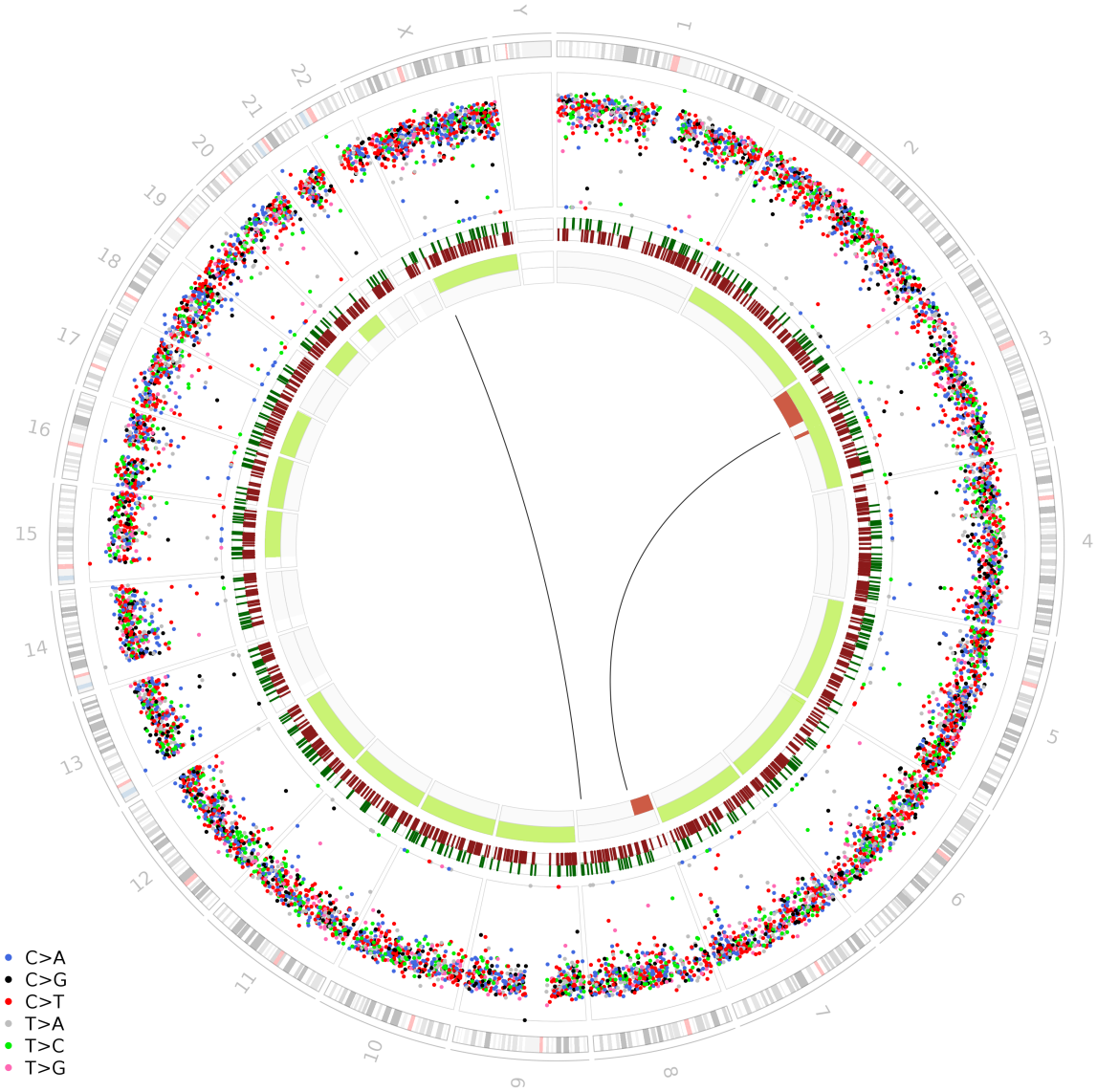
karyotype (purity 0.66, ploidy 2.7)

LOH gain

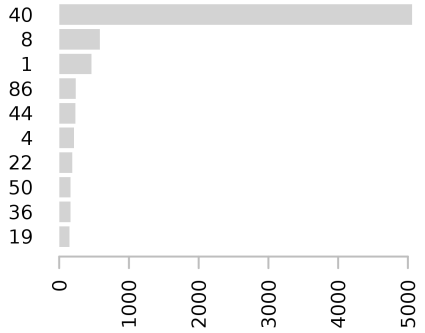
2 rearrangements



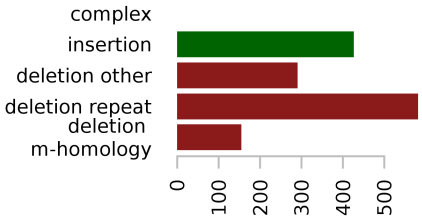
H203501 - Fresh Frozen



8458 substitutions of signatures:



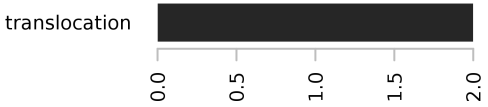
1454 indels



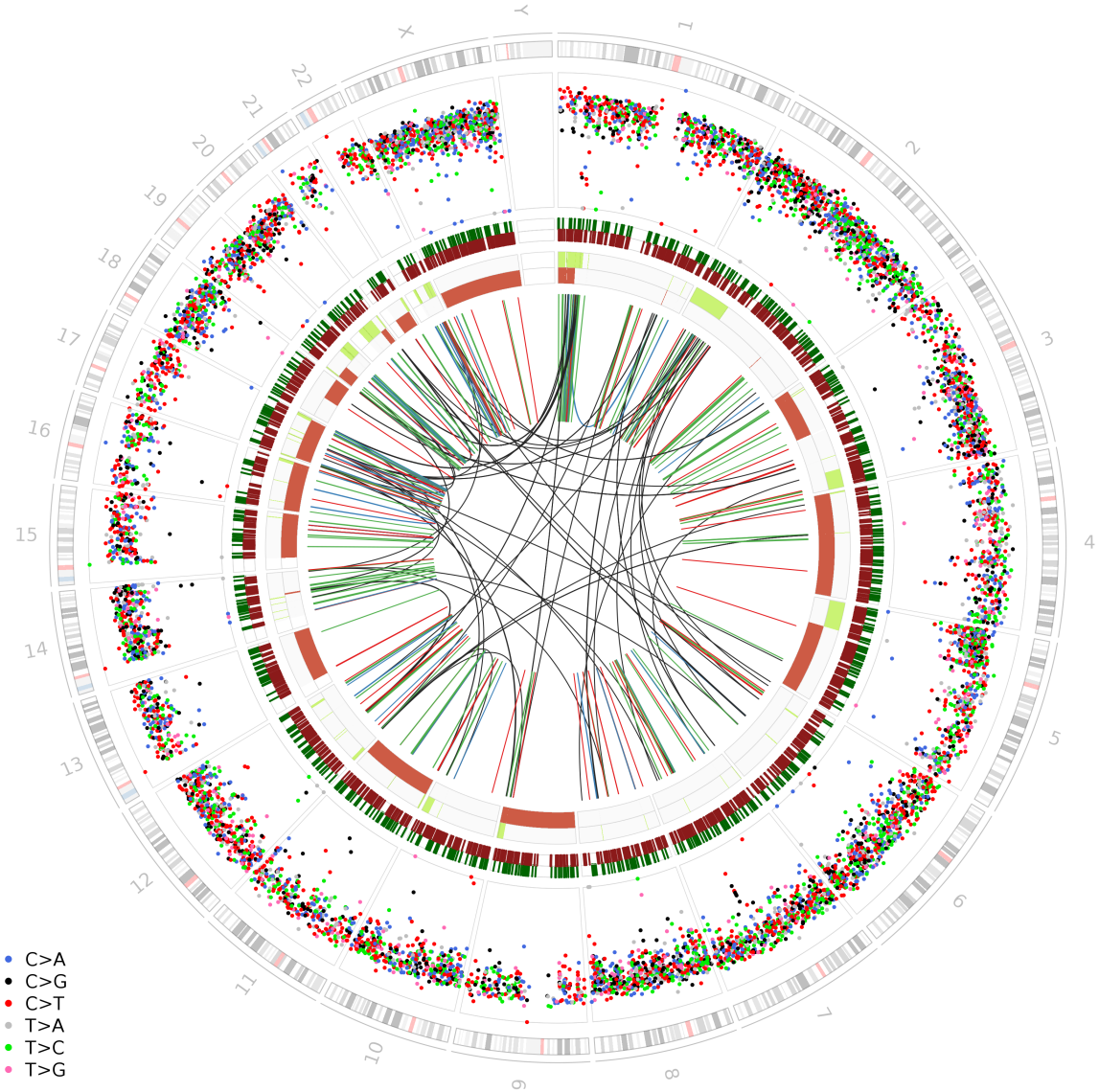
karyotype (purity 0.64, ploidy 3)



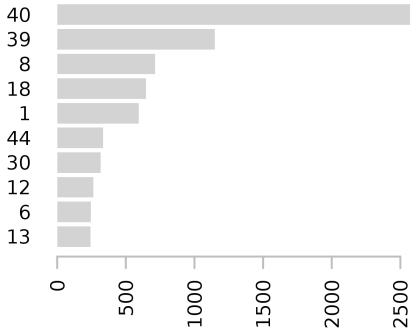
2 rearrangements



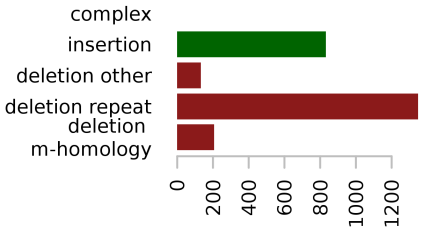
H203502 - FFPE



9305 substitutions of signatures:



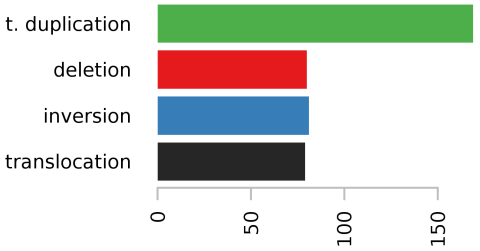
2519 indels



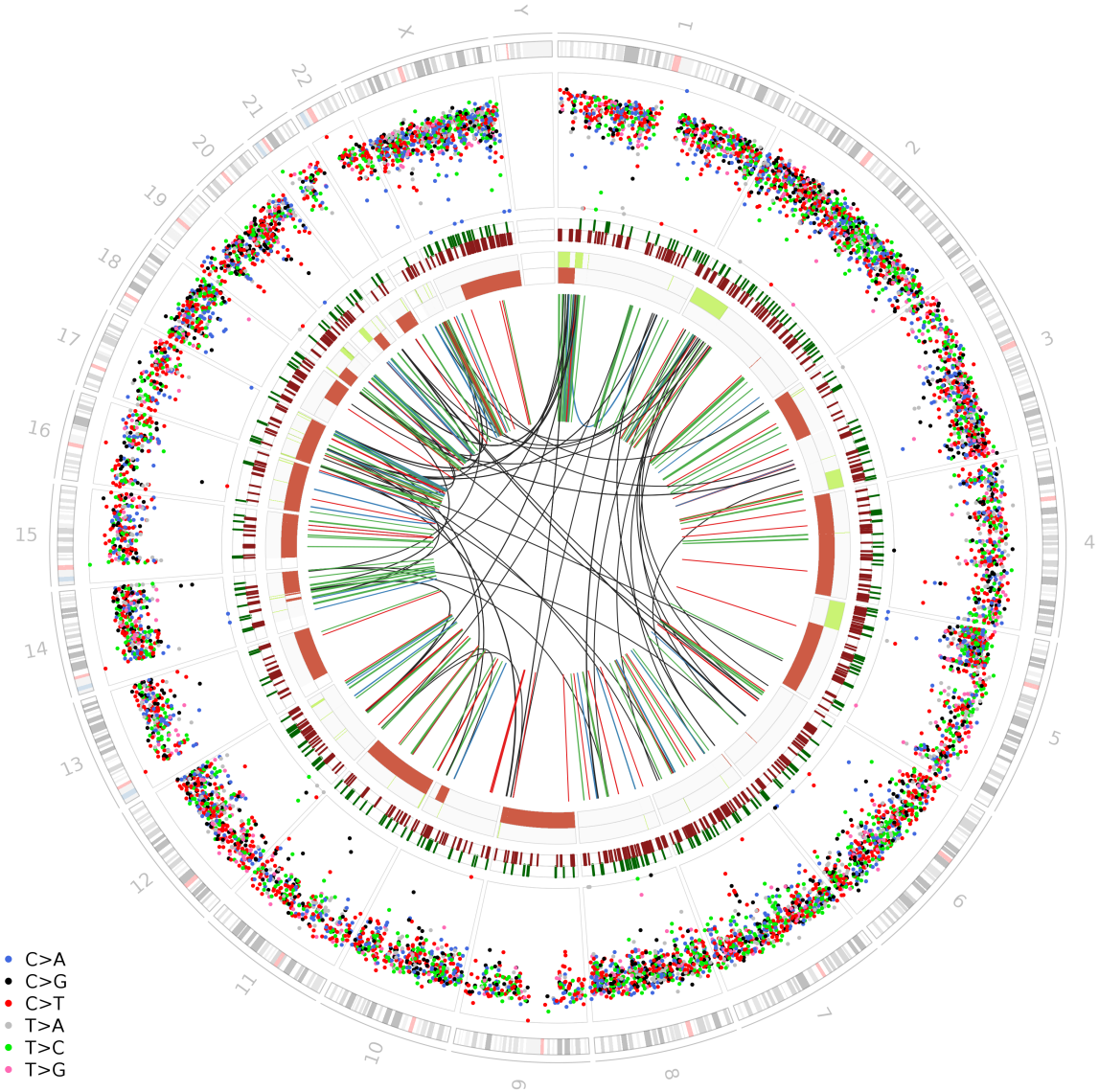
karyotype (purity 0.52, ploidy 1.9)



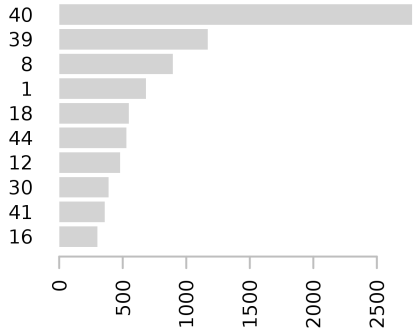
409 rearrangements



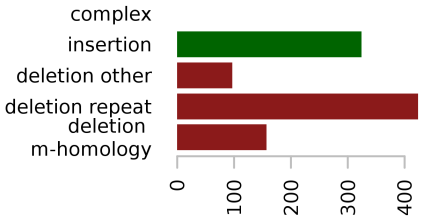
H203502 - Fresh Frozen



9228 substitutions of signatures:



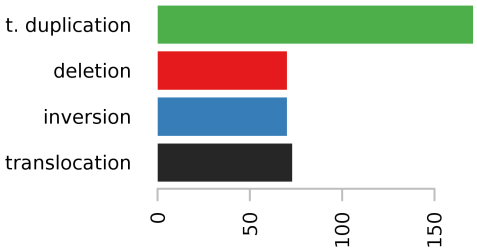
1002 indels



karyotype (purity 0.47, ploidy 1.9)



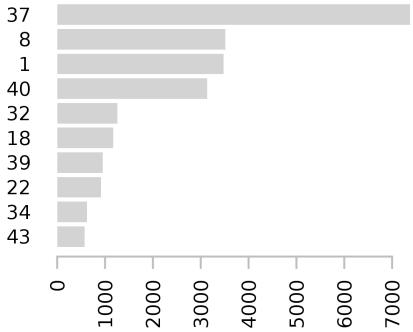
384 rearrangements



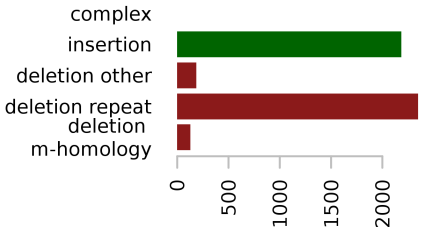
H203503 - FFPE



29612 substitutions of signatures:



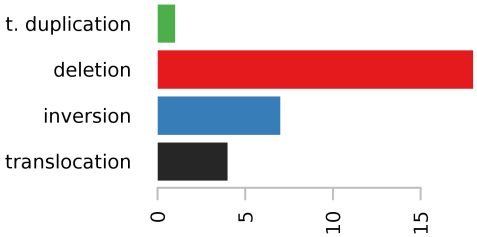
4851 indels



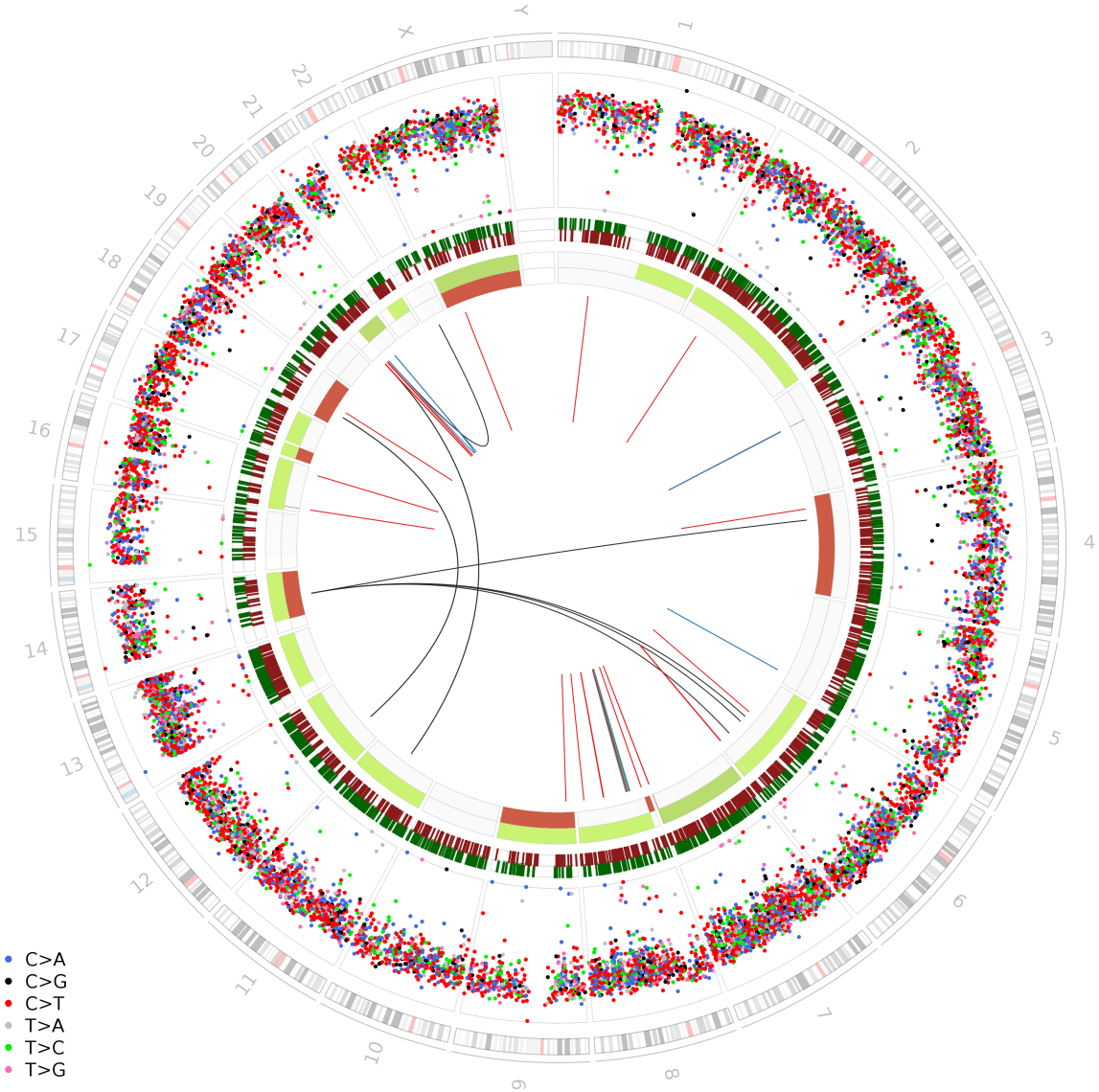
karyotype (purity 0.75, ploidy 4.5)



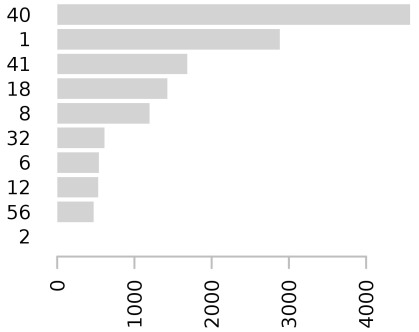
30 rearrangements



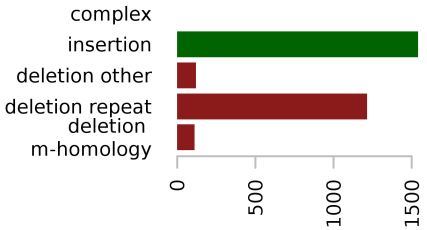
H203503 - Fresh Frozen



14074 substitutions of signatures:



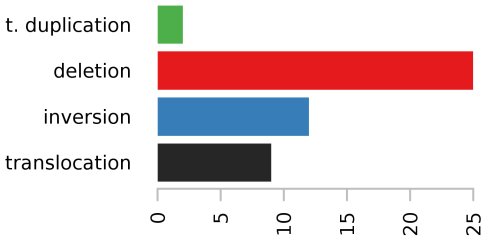
2985 indels



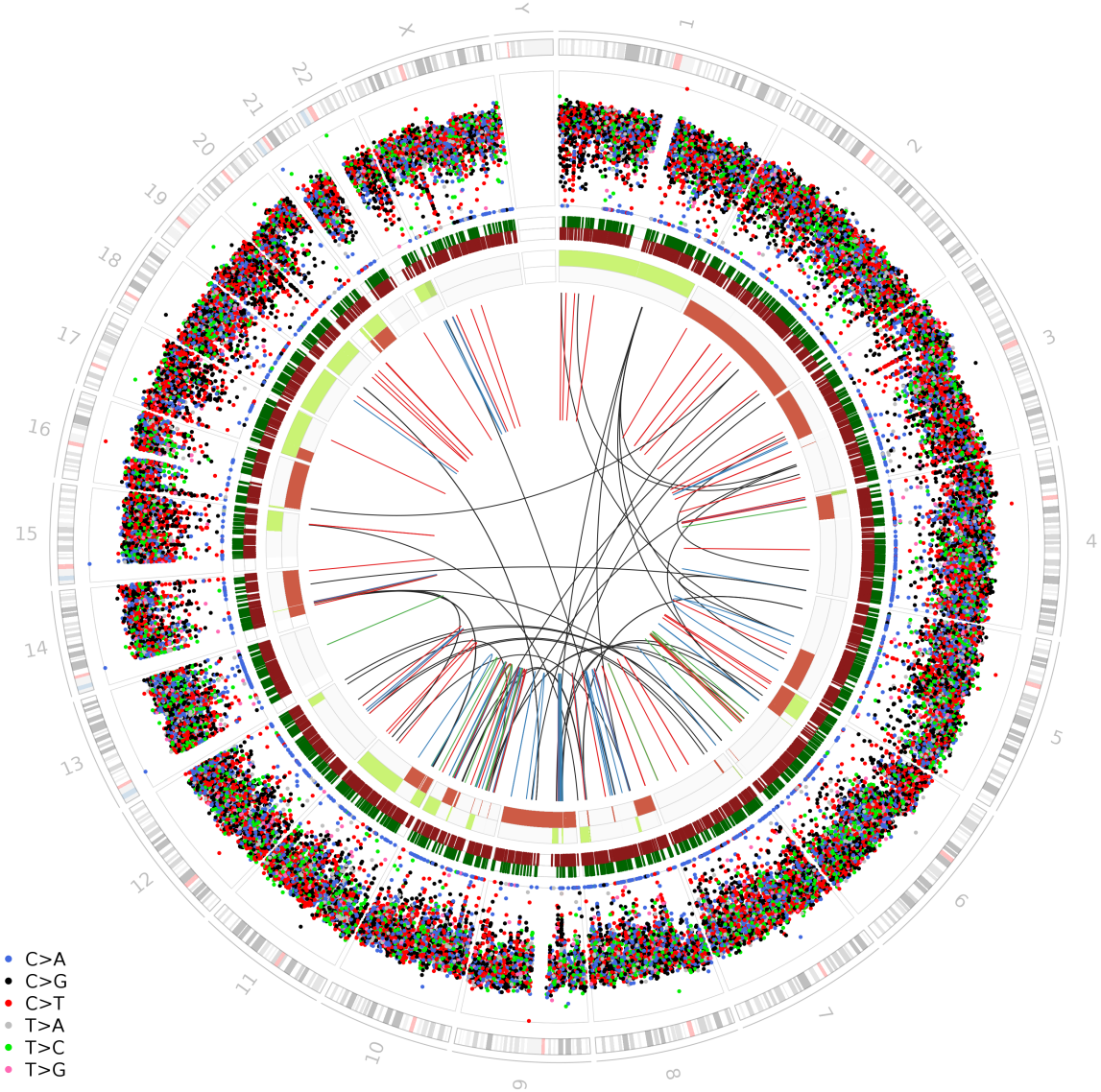
karyotype (purity 0.5, ploidy 3)



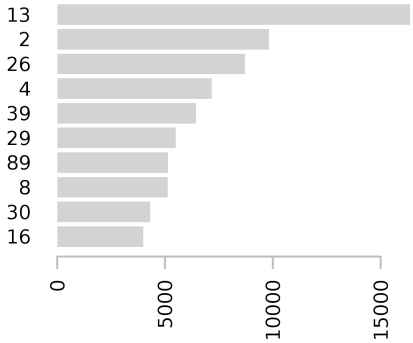
48 rearrangements



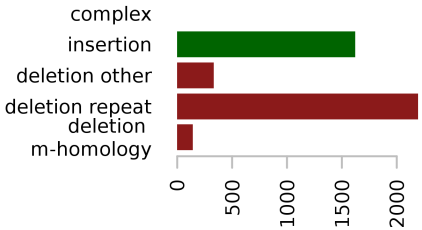
H203504 - FFPE



77275 substitutions of signatures:



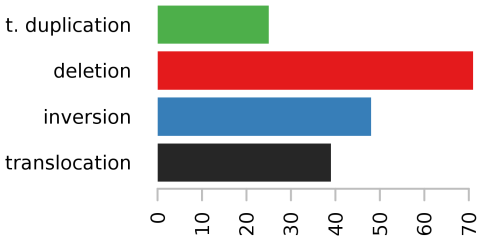
4295 indels



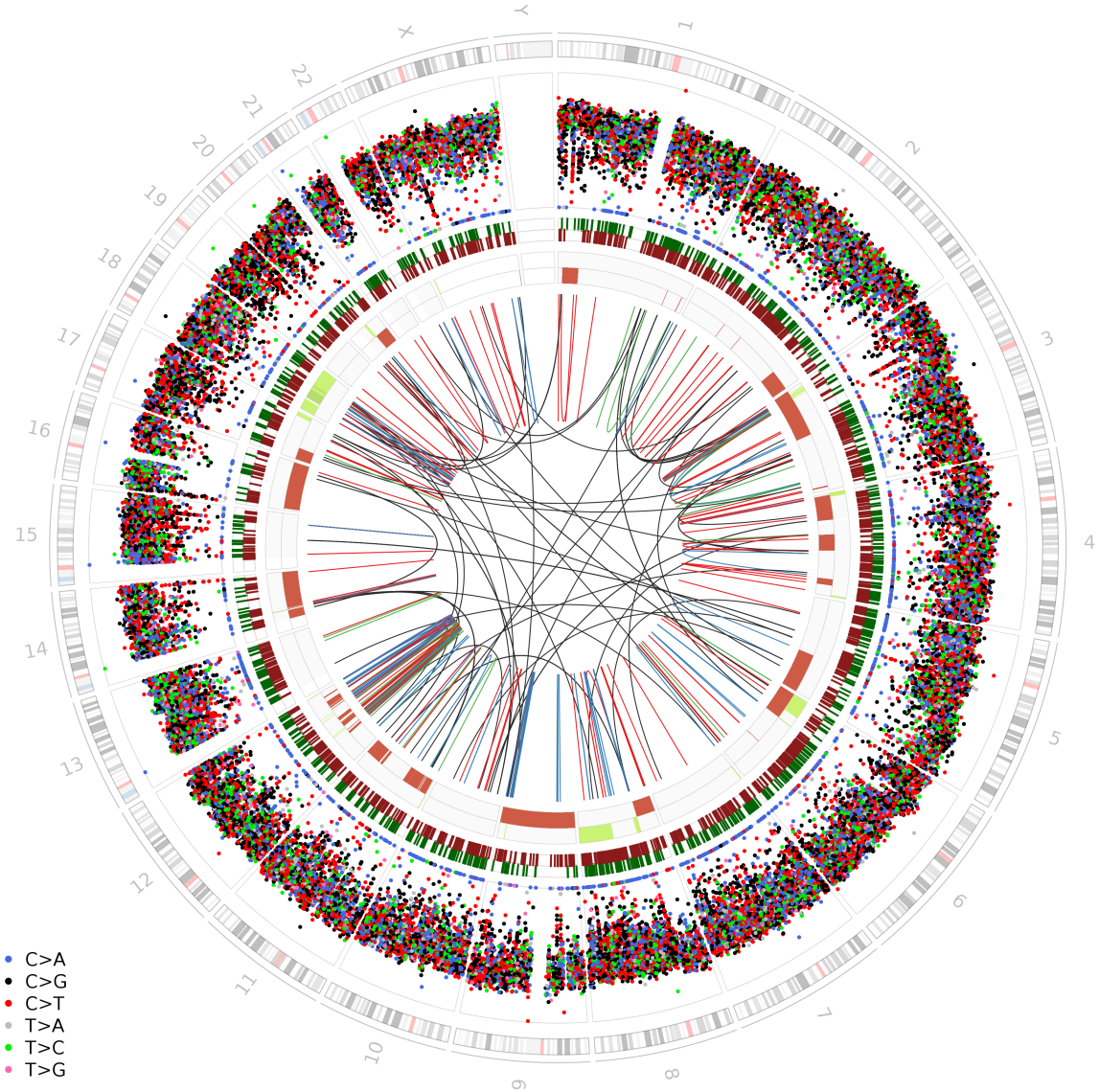
karyotype (purity 0.67, ploidy 2.7)



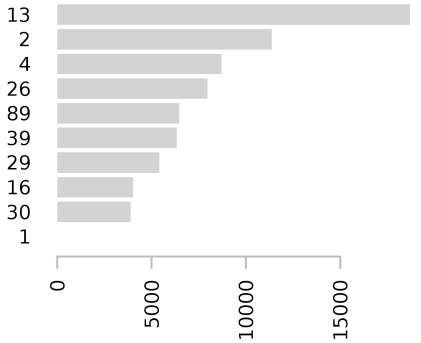
183 rearrangements



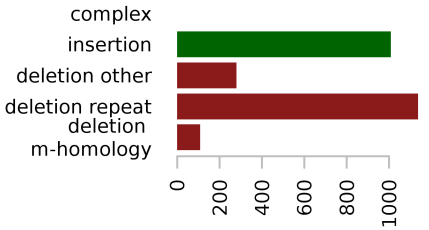
H203504 - Fresh Frozen



77453 substitutions of signatures:



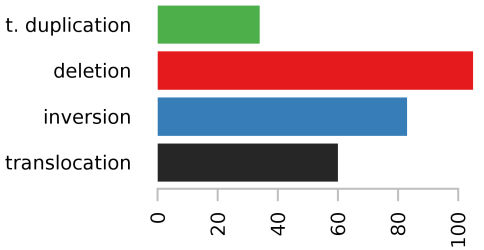
2536 indels



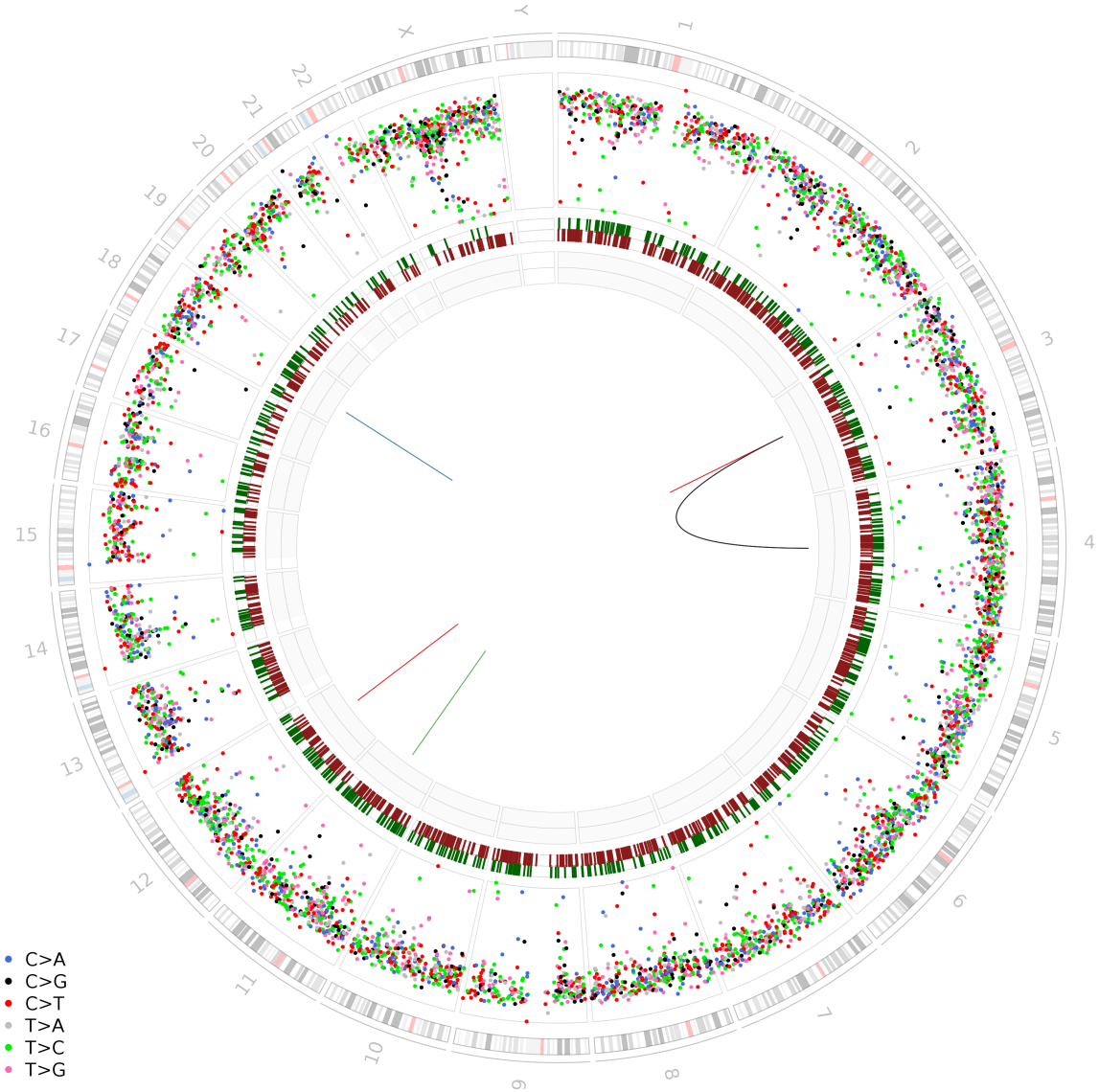
karyotype (purity 0.53, ploidy 2.1)

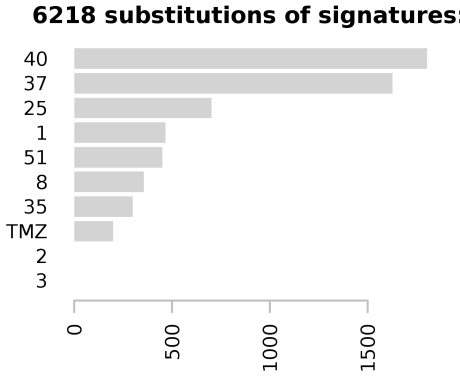


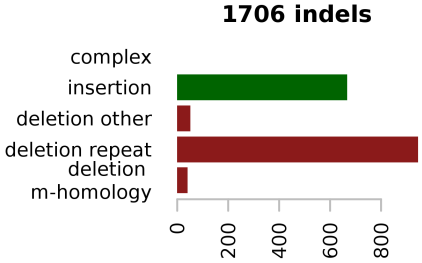
282 rearrangements

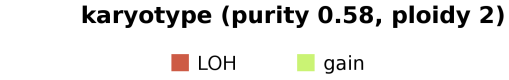


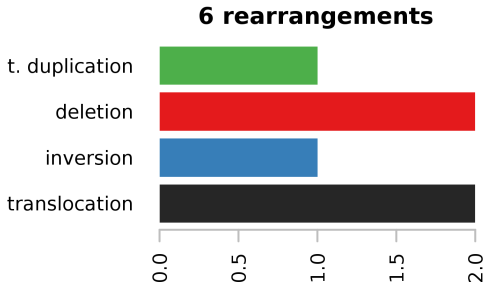
H203505 - FFPE



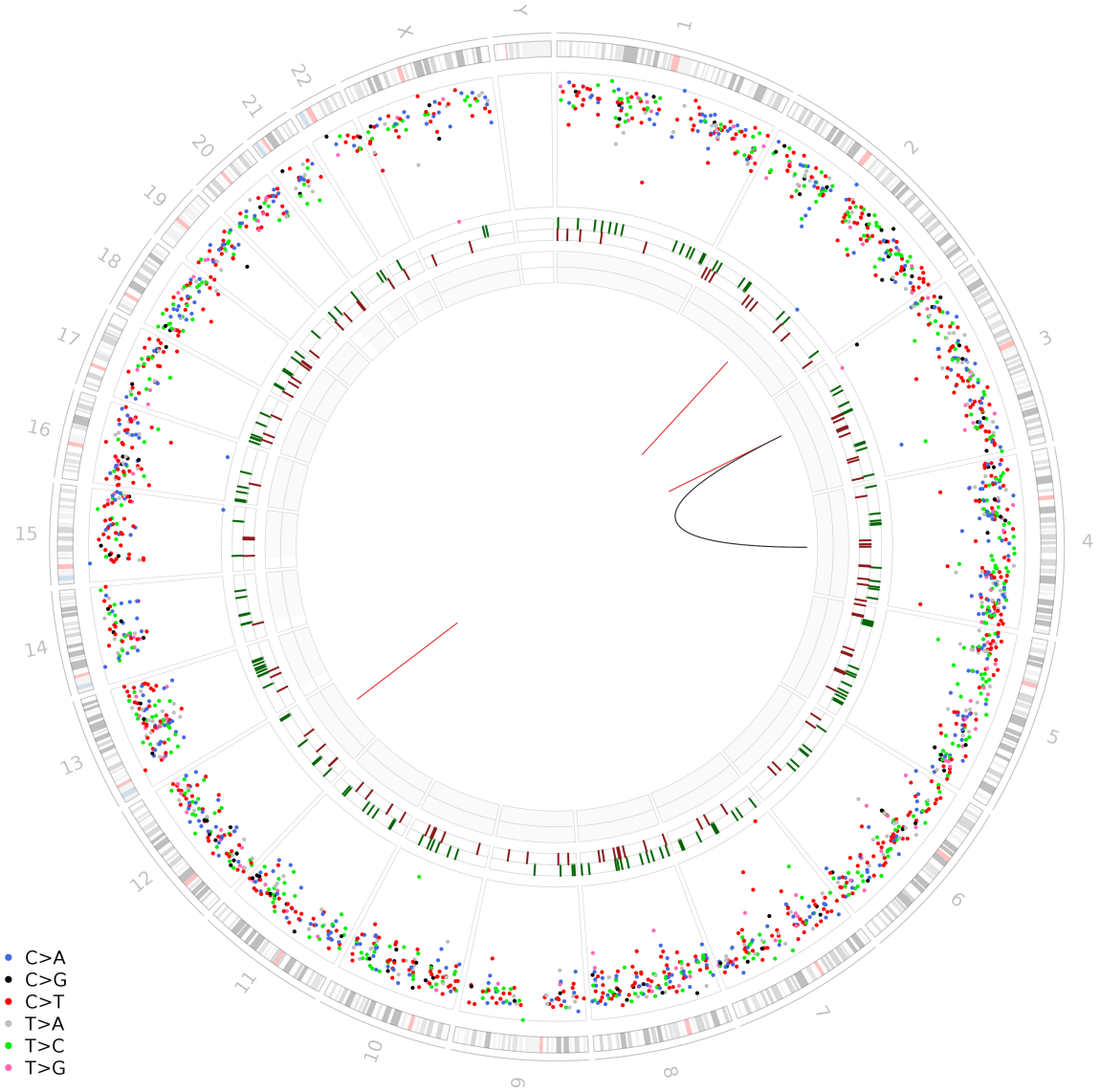




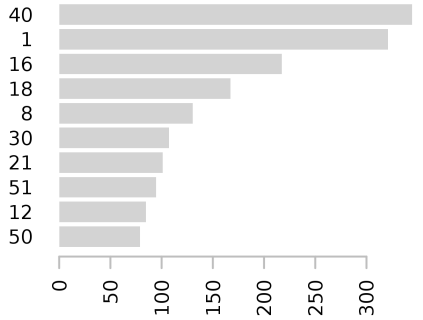




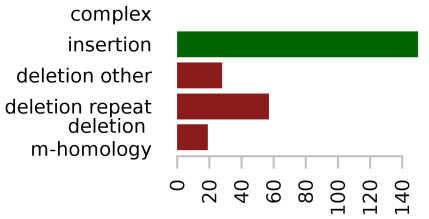
H203505 - Fresh Frozen



2036 substitutions of signatures:



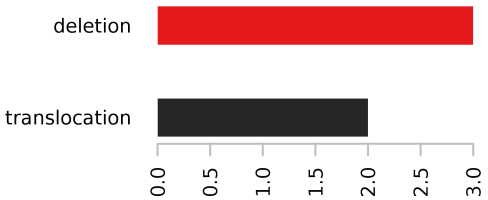
254 indels



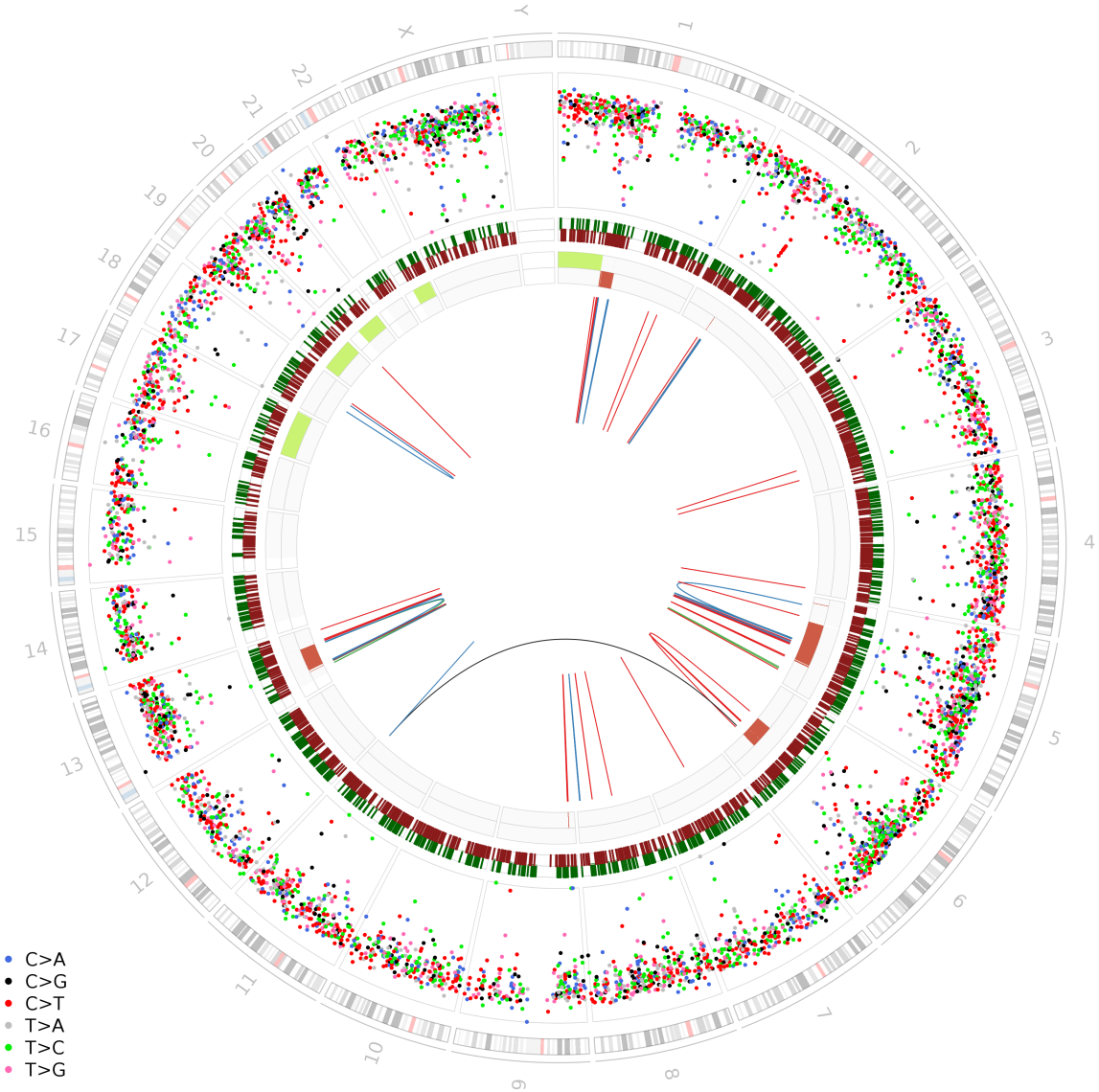
karyotype (purity 0.54, ploidy 2)



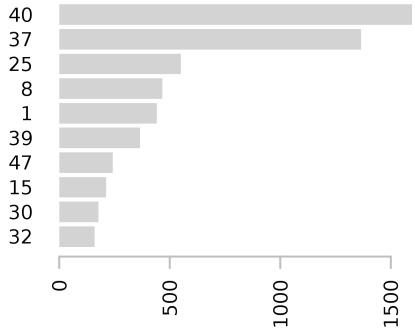
5 rearrangements



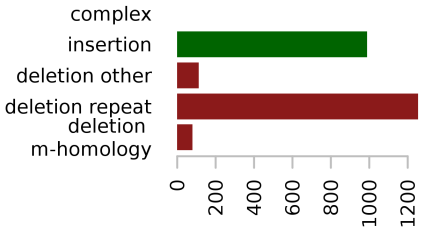
H203506 - FFPE



6012 substitutions of signatures:



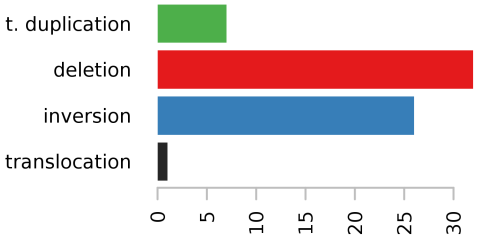
2436 indels



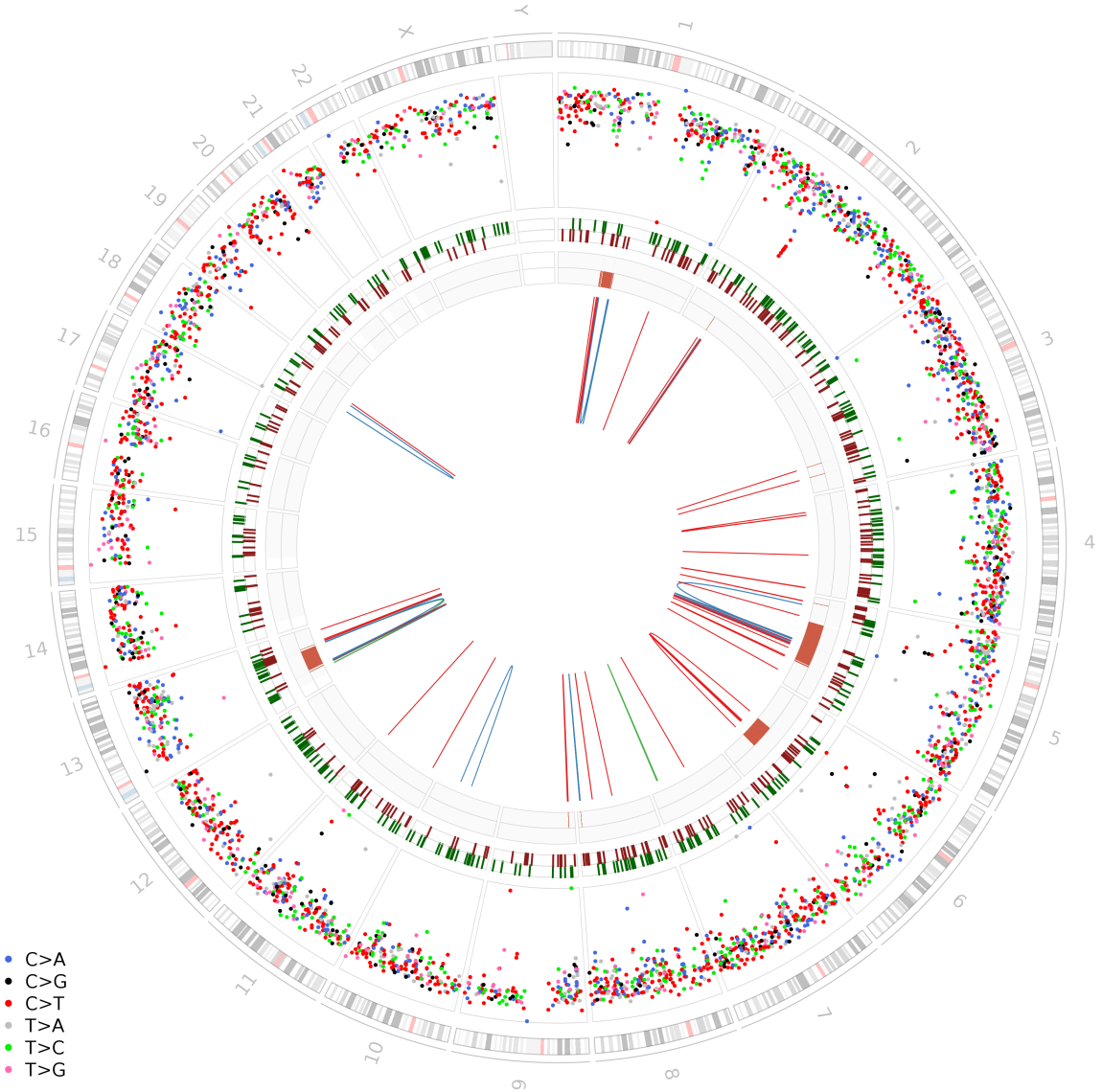
karyotype (purity 0.64, ploidy 3)



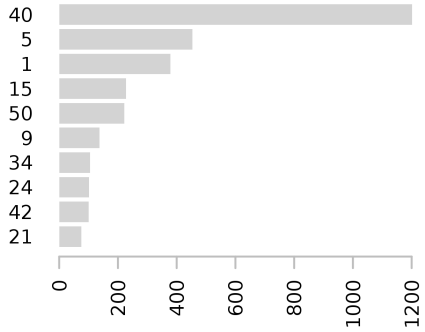
66 rearrangements



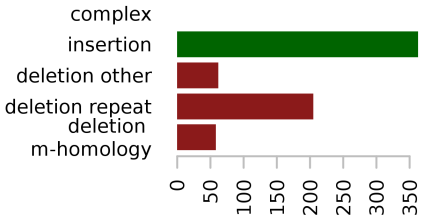
H203506 - Fresh Frozen



3581 substitutions of signatures:



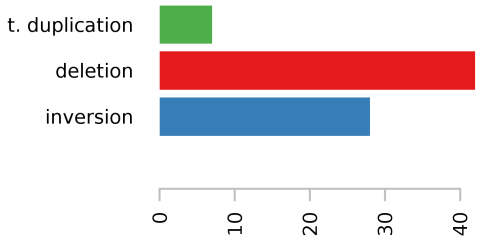
688 indels



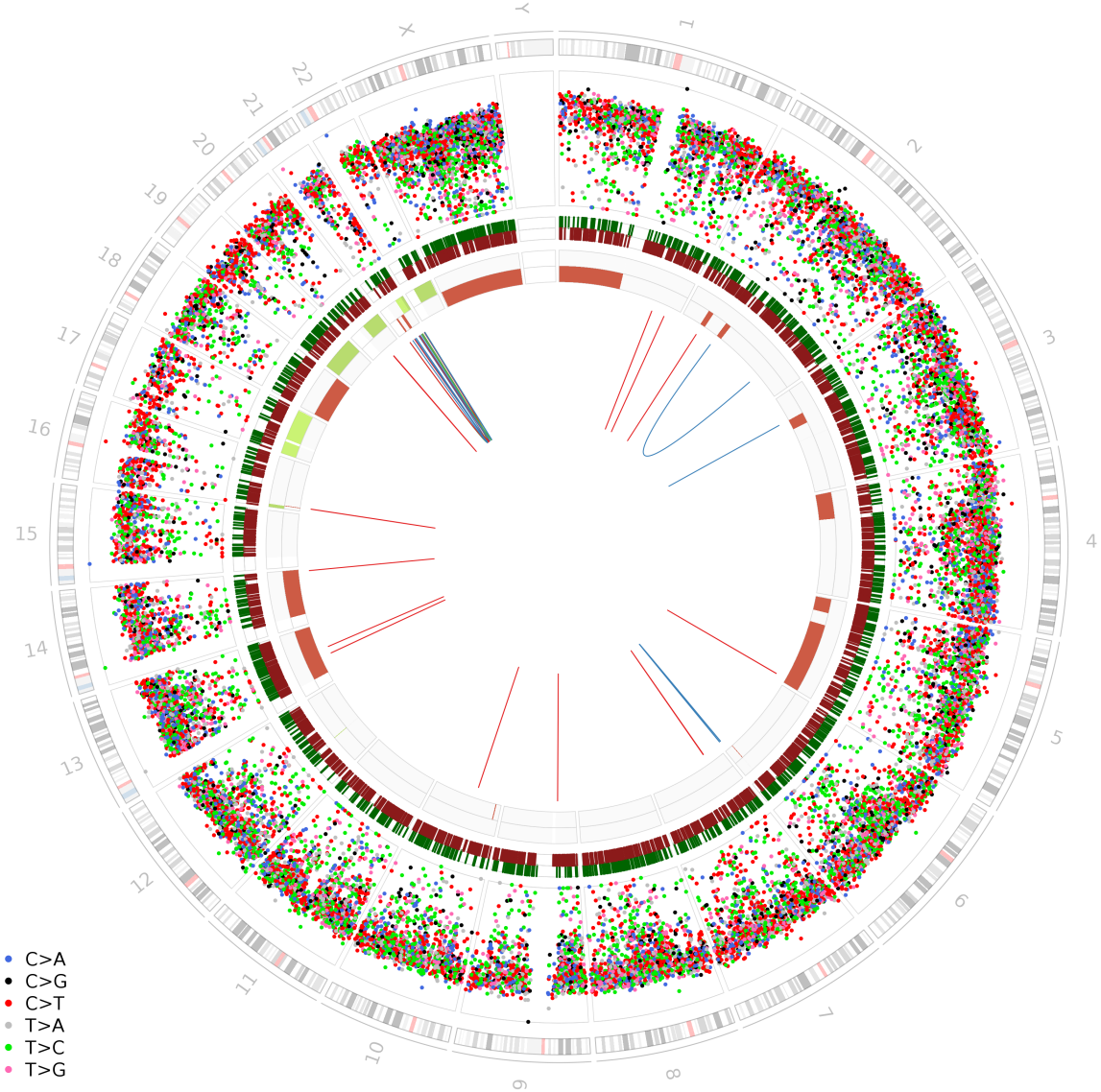
karyotype (purity 0.76, ploidy 2)



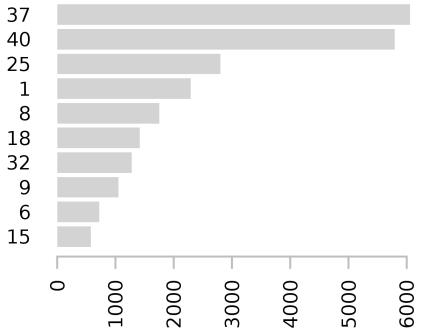
77 rearrangements



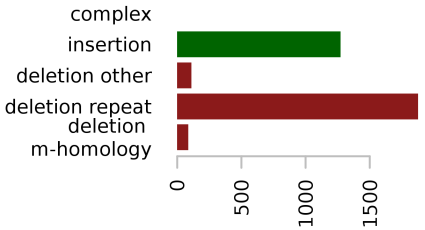
H203507 - FFPE



26457 substitutions of signatures:



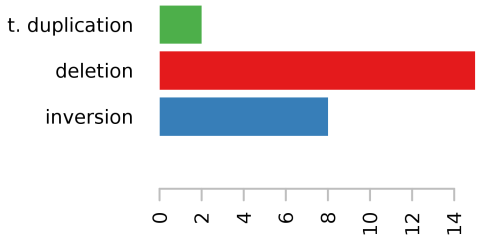
3345 indels



karyotype (purity 0.42, ploidy 2)



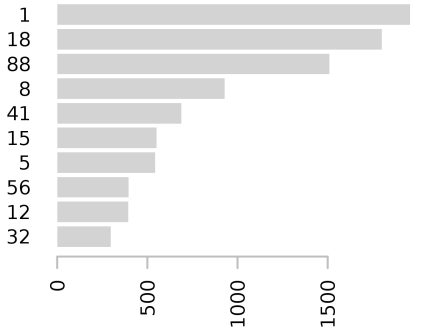
25 rearrangements



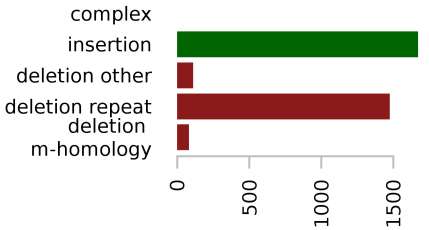
H203507 - Fresh Frozen



12553 substitutions of signatures:



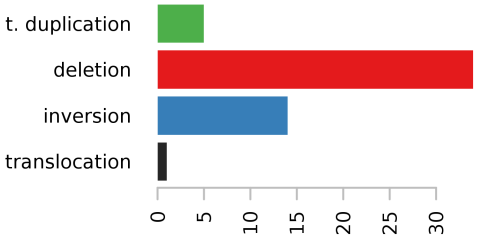
3340 indels



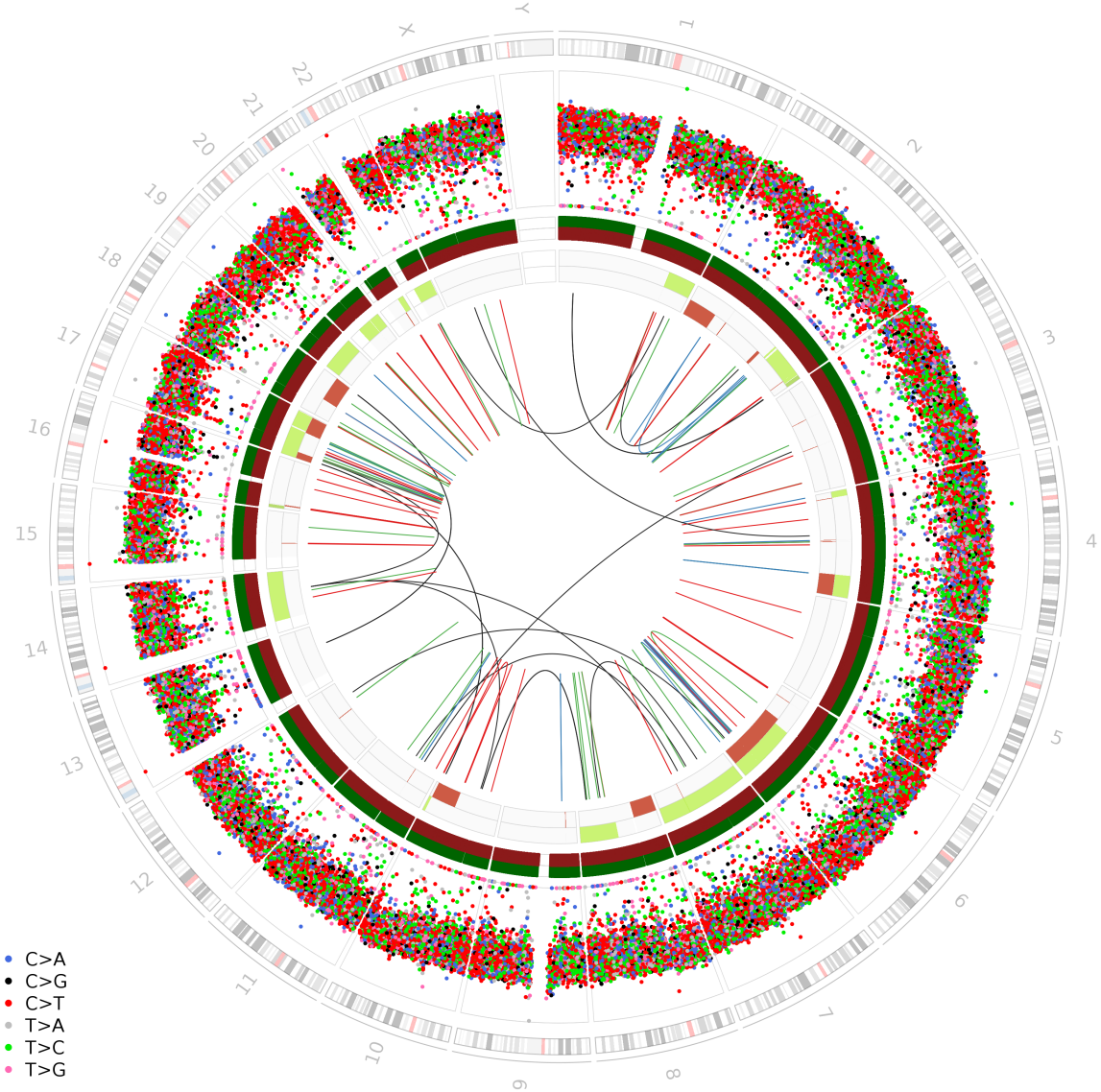
karyotype (purity 0.43, ploidy 2.1)



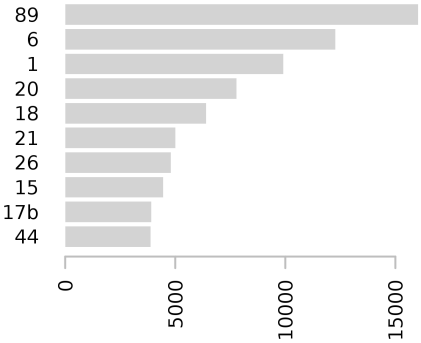
54 rearrangements



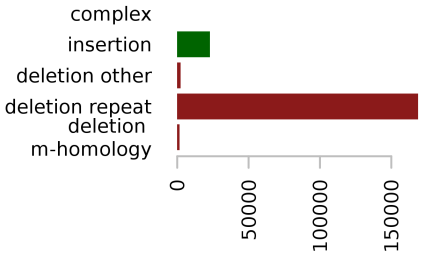
H203508 - FFPE



99861 substitutions of signatures:



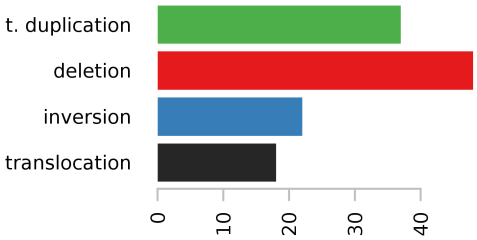
196058 indels



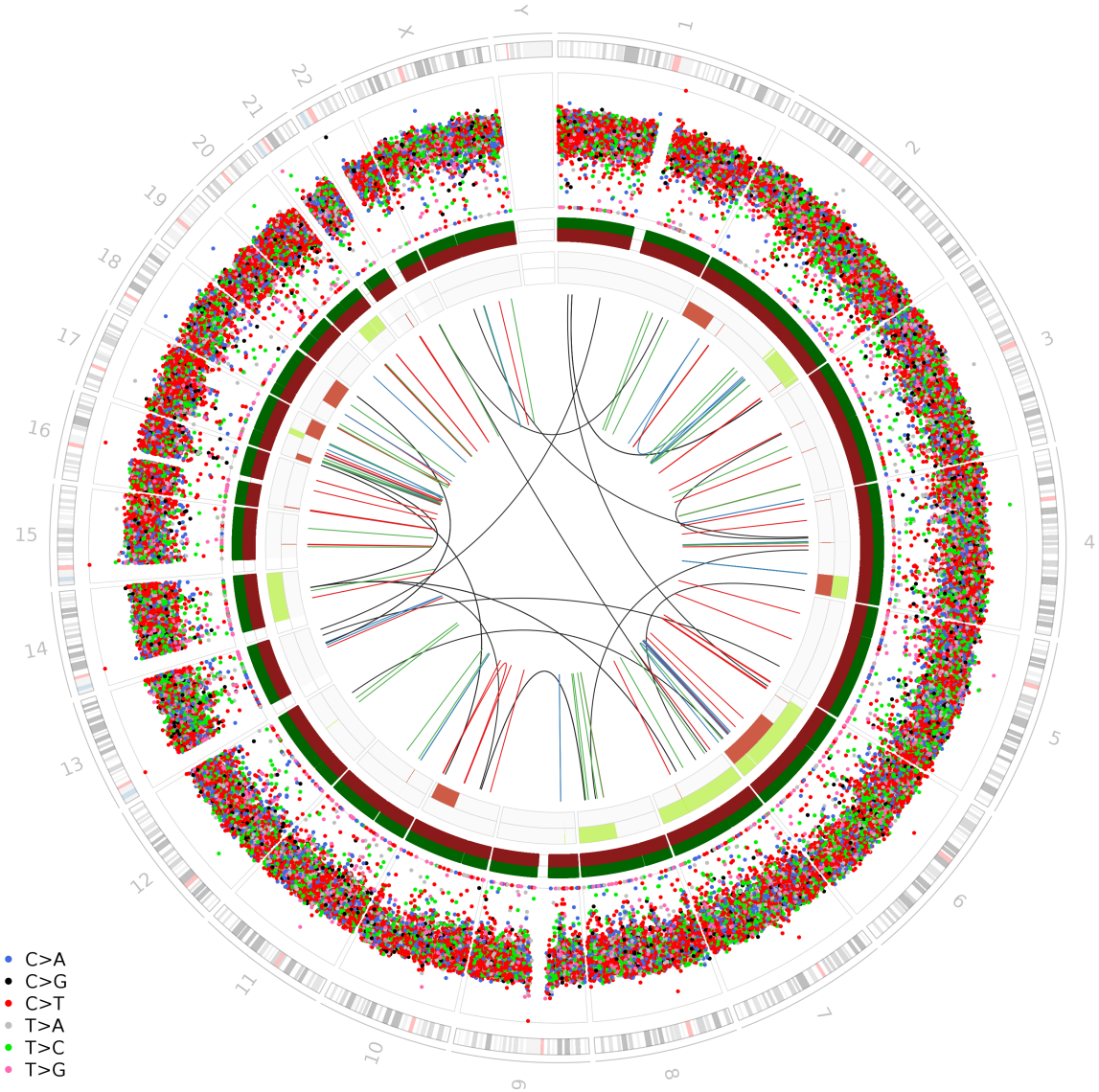
karyotype (purity 0.63, ploidy 3.1)

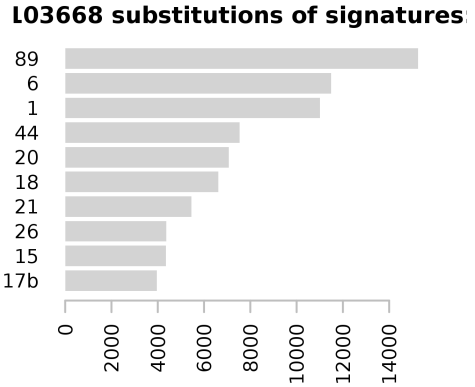


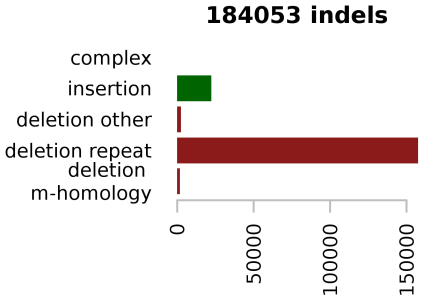
125 rearrangements



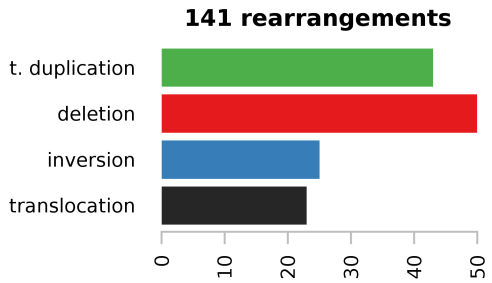
H203508 - Fresh Frozen



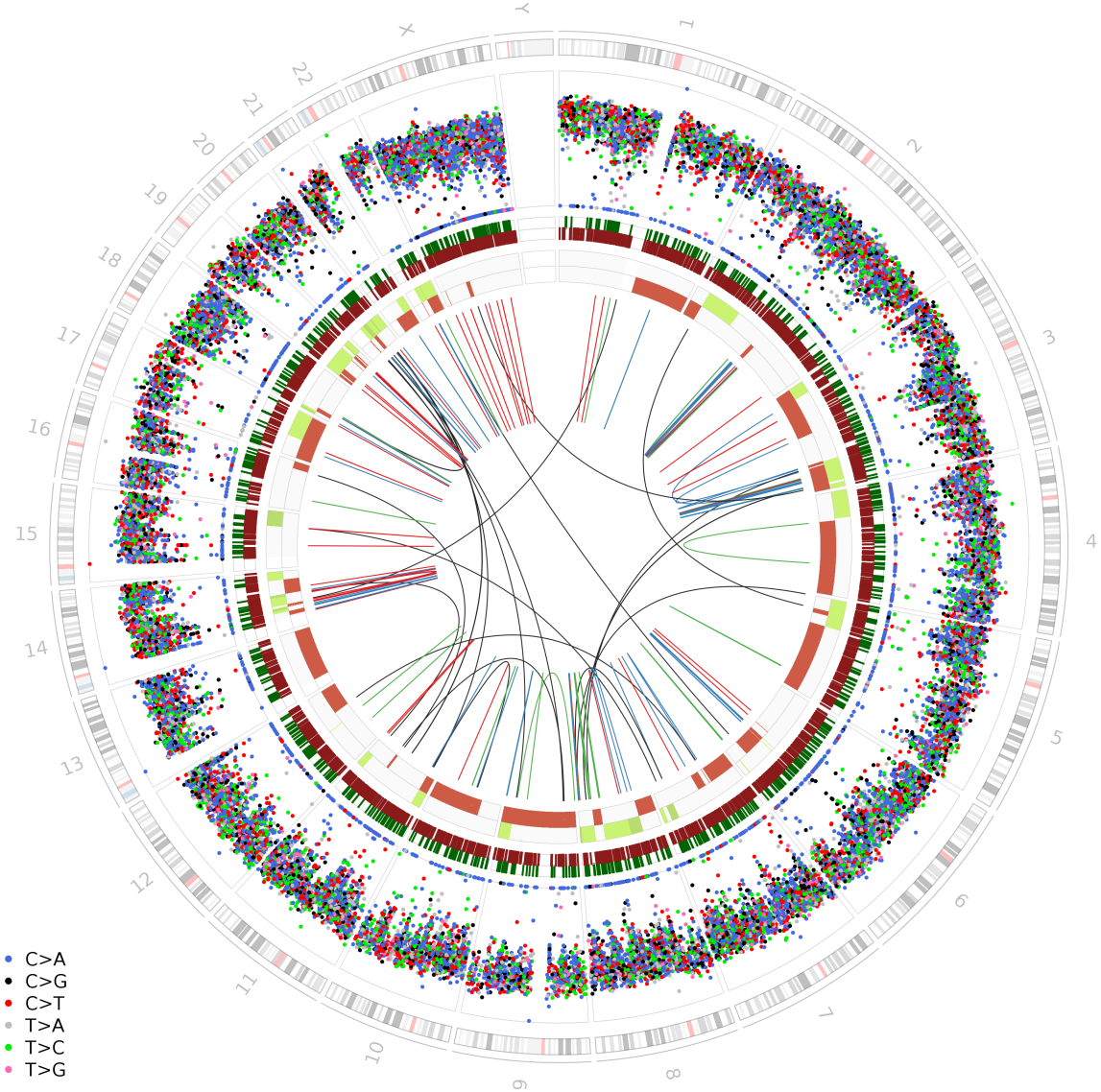




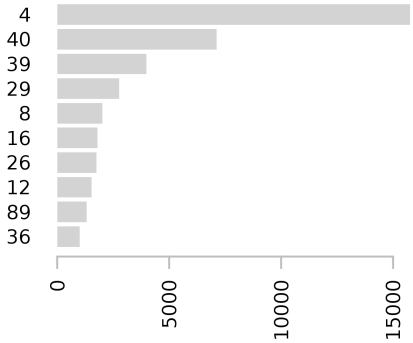




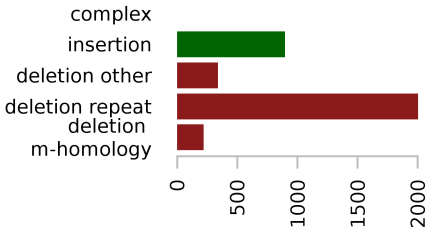
H203511 - FFPE



44504 substitutions of signatures:



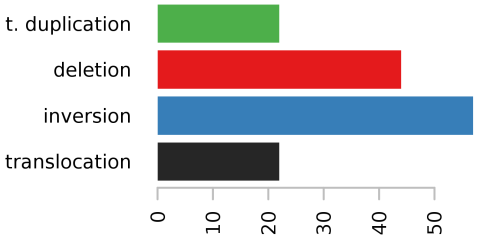
3460 indels



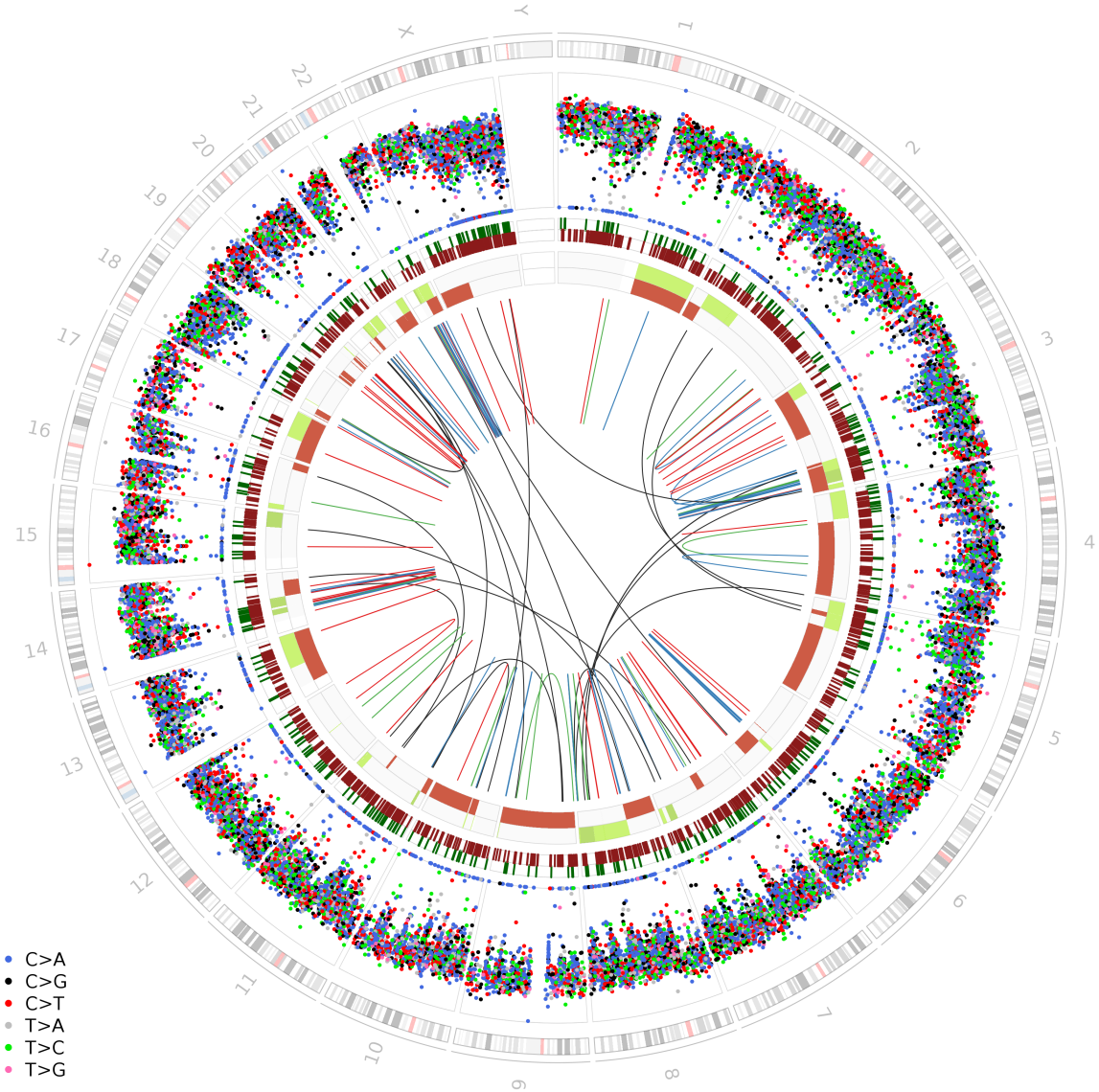
karyotype (purity 0.83, ploidy 2.5)



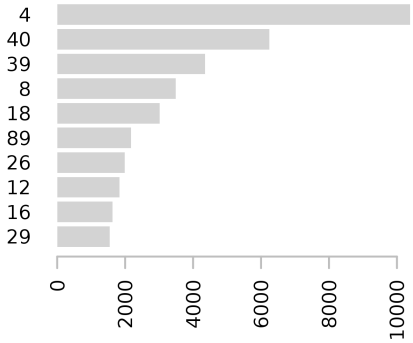
145 rearrangements



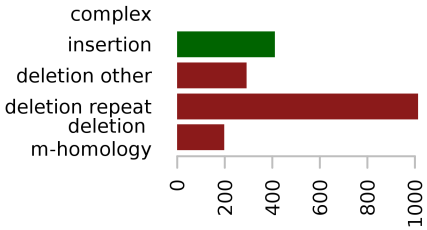
H203511 - Fresh Frozen



43303 substitutions of signatures:



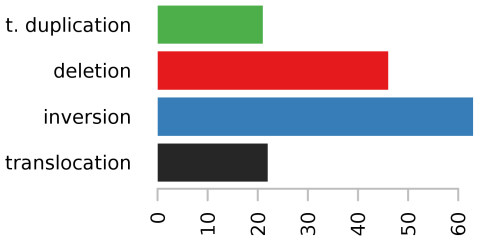
1915 indels



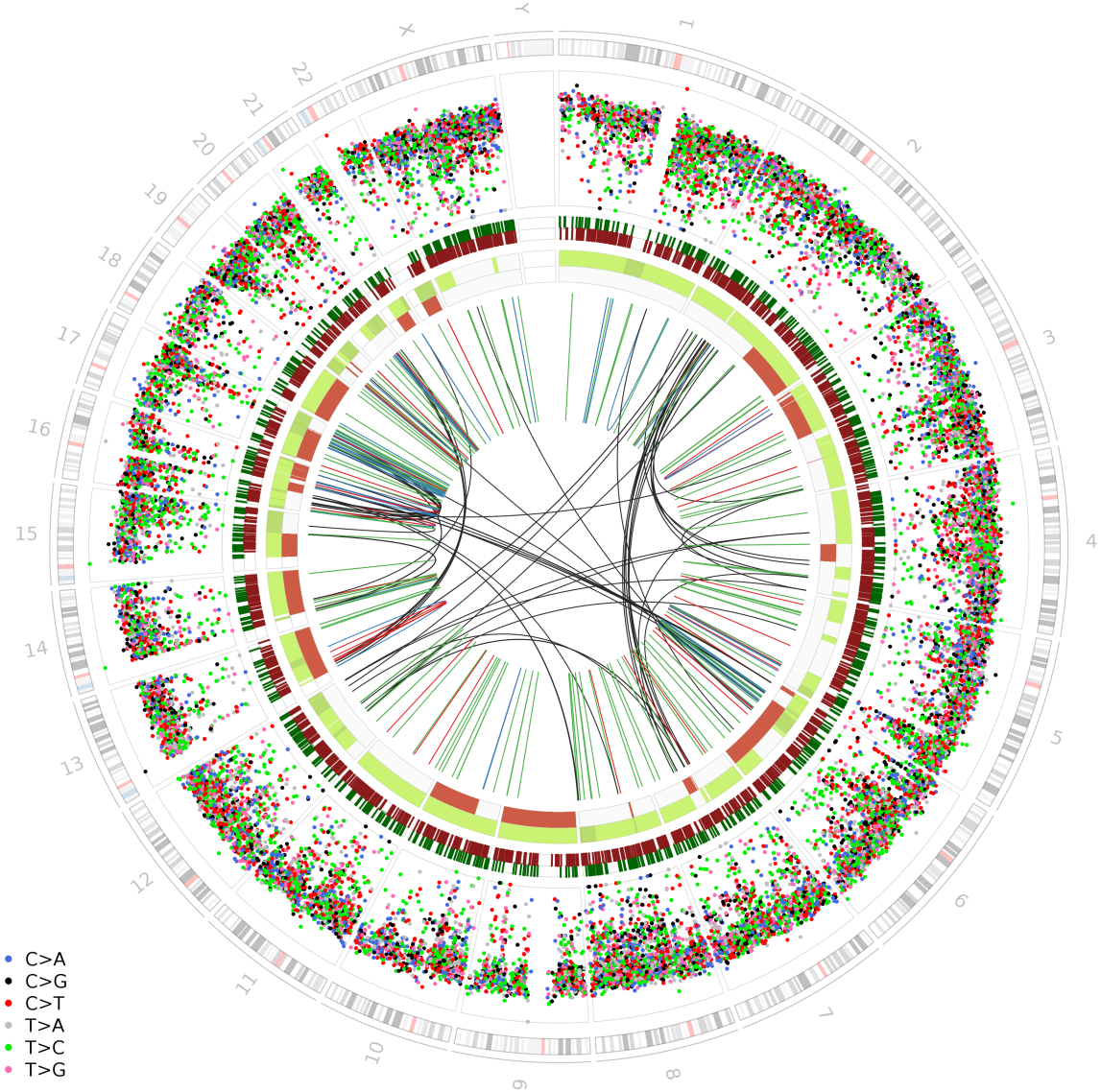
karyotype (purity 0.74, ploidy 2.2)



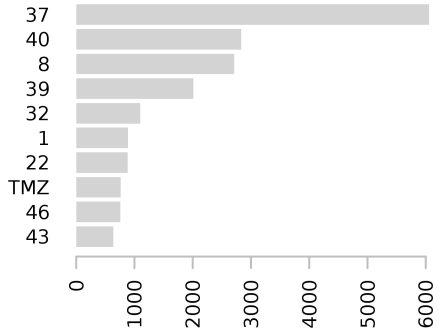
152 rearrangements



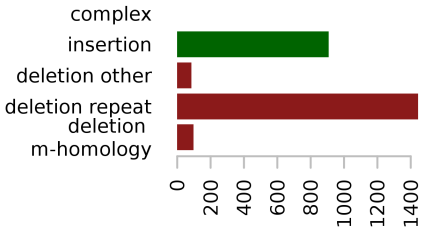
H203512 - FFPE



23586 substitutions of signatures:



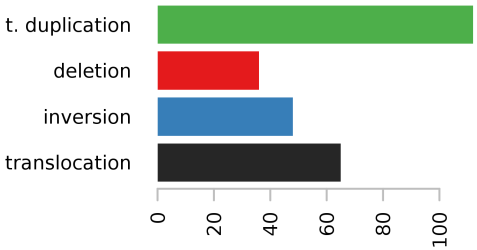
2535 indels



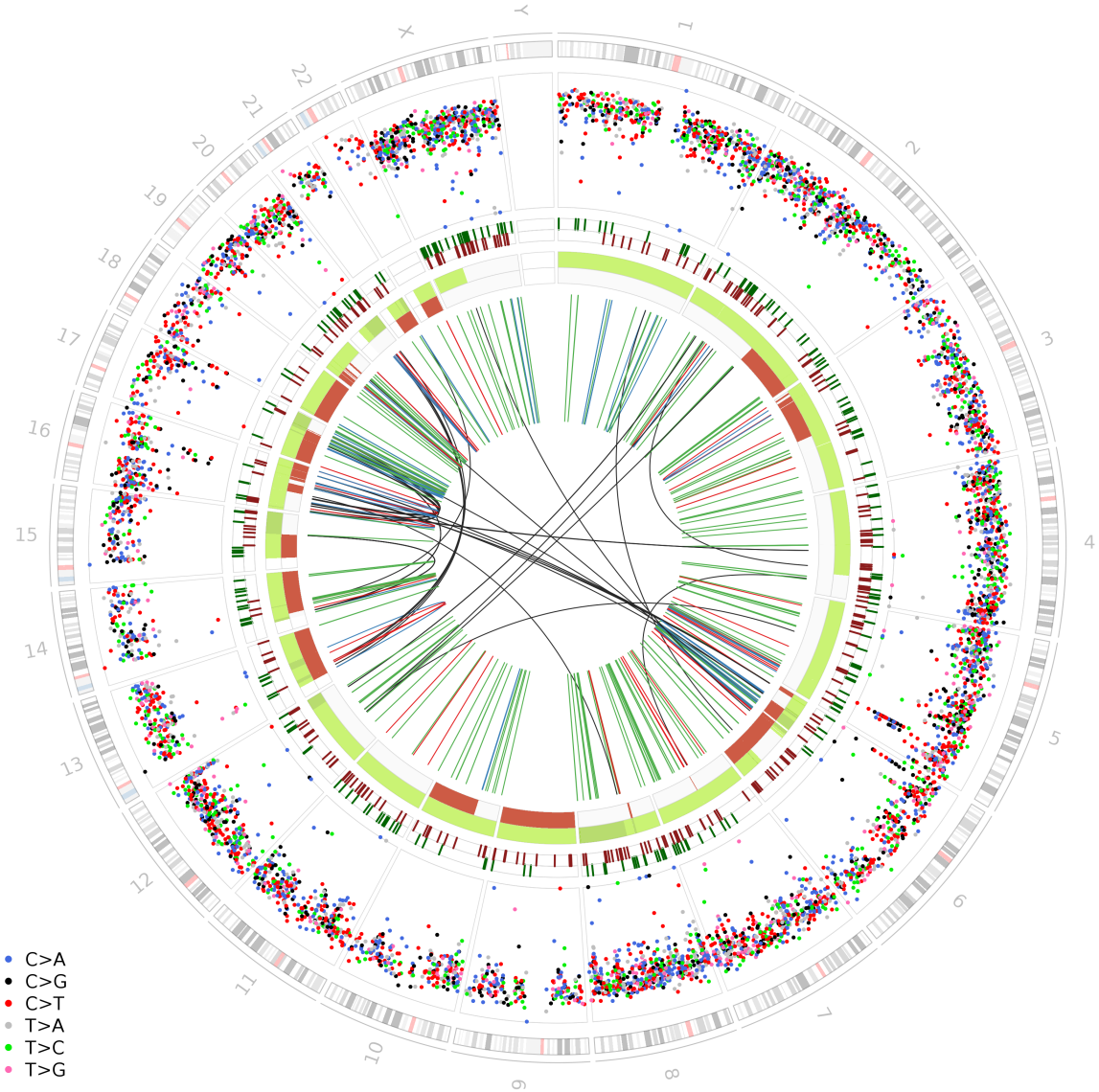
karyotype (purity 0.64, ploidy 4.9)



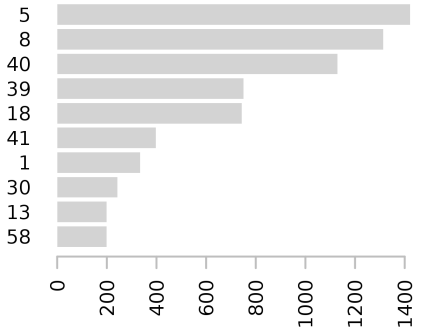
261 rearrangements



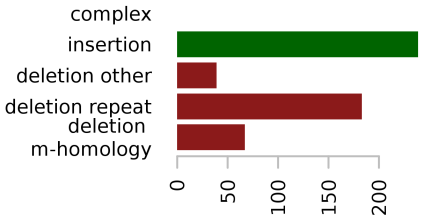
H203512 - Fresh Frozen



6806 substitutions of signatures:



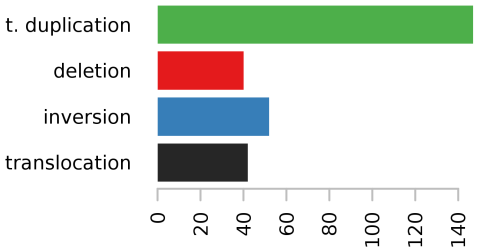
528 indels



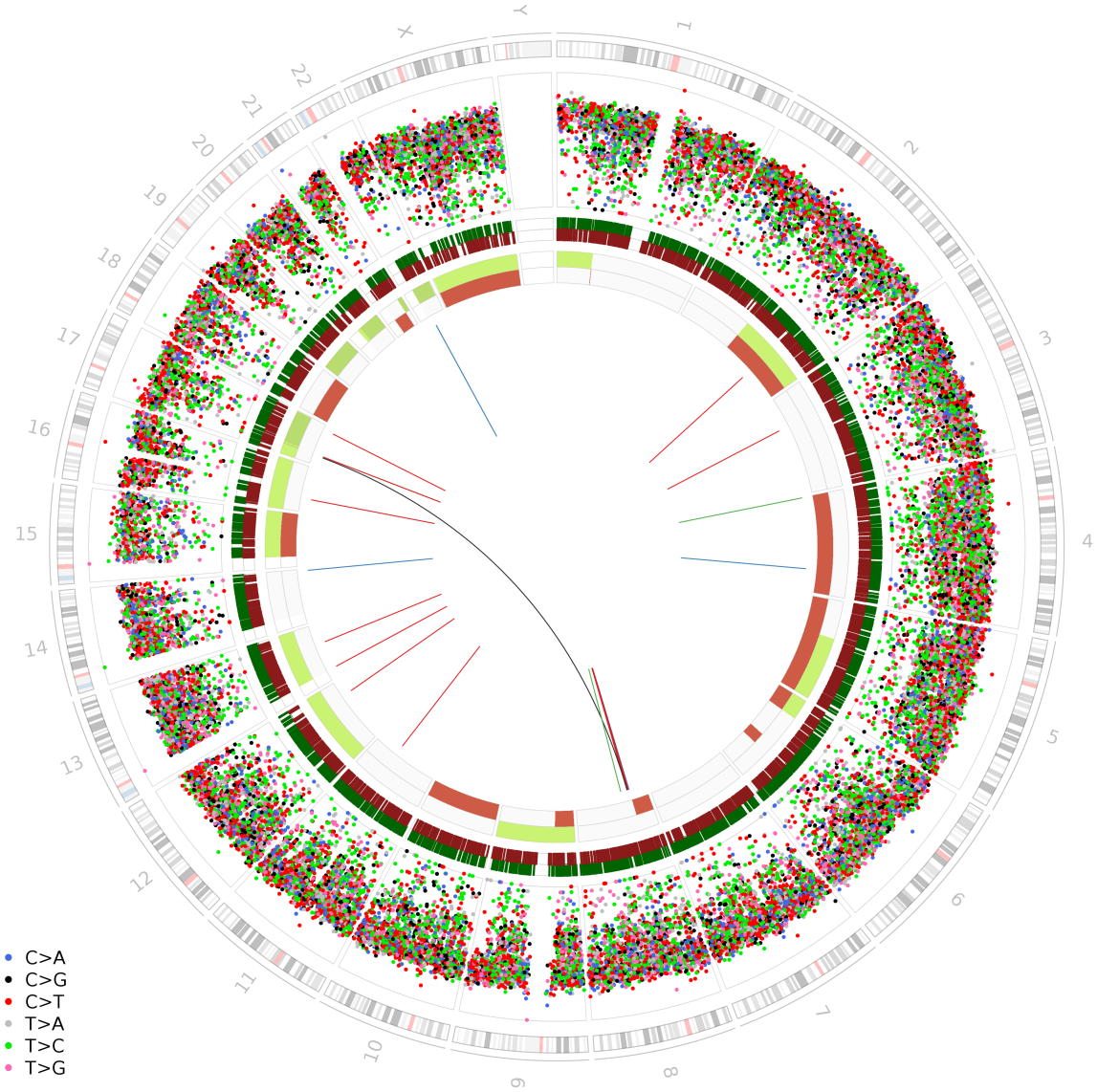
karyotype (purity 0.34, ploidy 3.4)



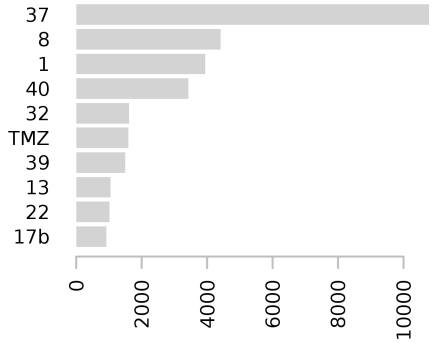
281 rearrangements



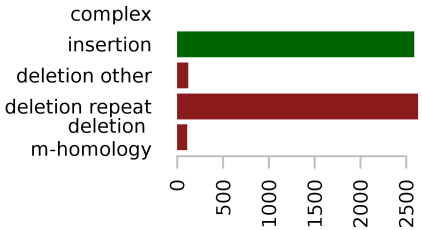
H203513 - FFPE



40593 substitutions of signatures:



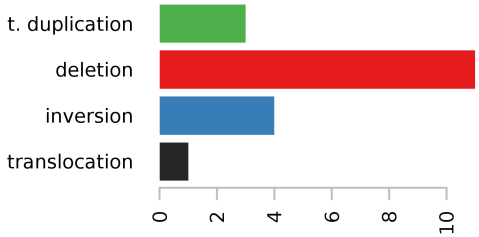
5449 indels



karyotype (purity 0.42, ploidy 4.5)



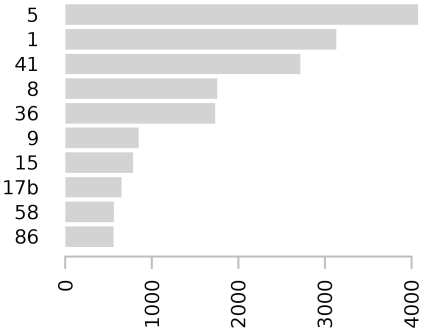
19 rearrangements



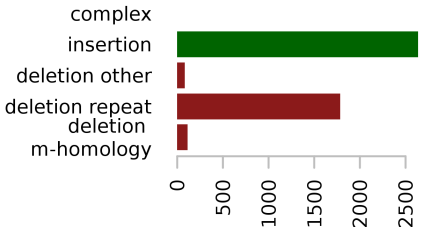
H203513 - Fresh Frozen



16931 substitutions of signatures:



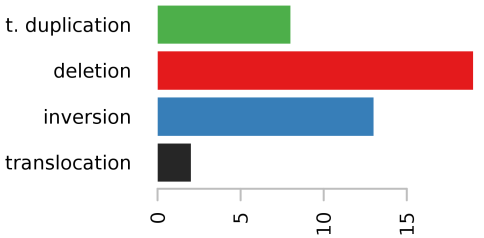
4616 indels



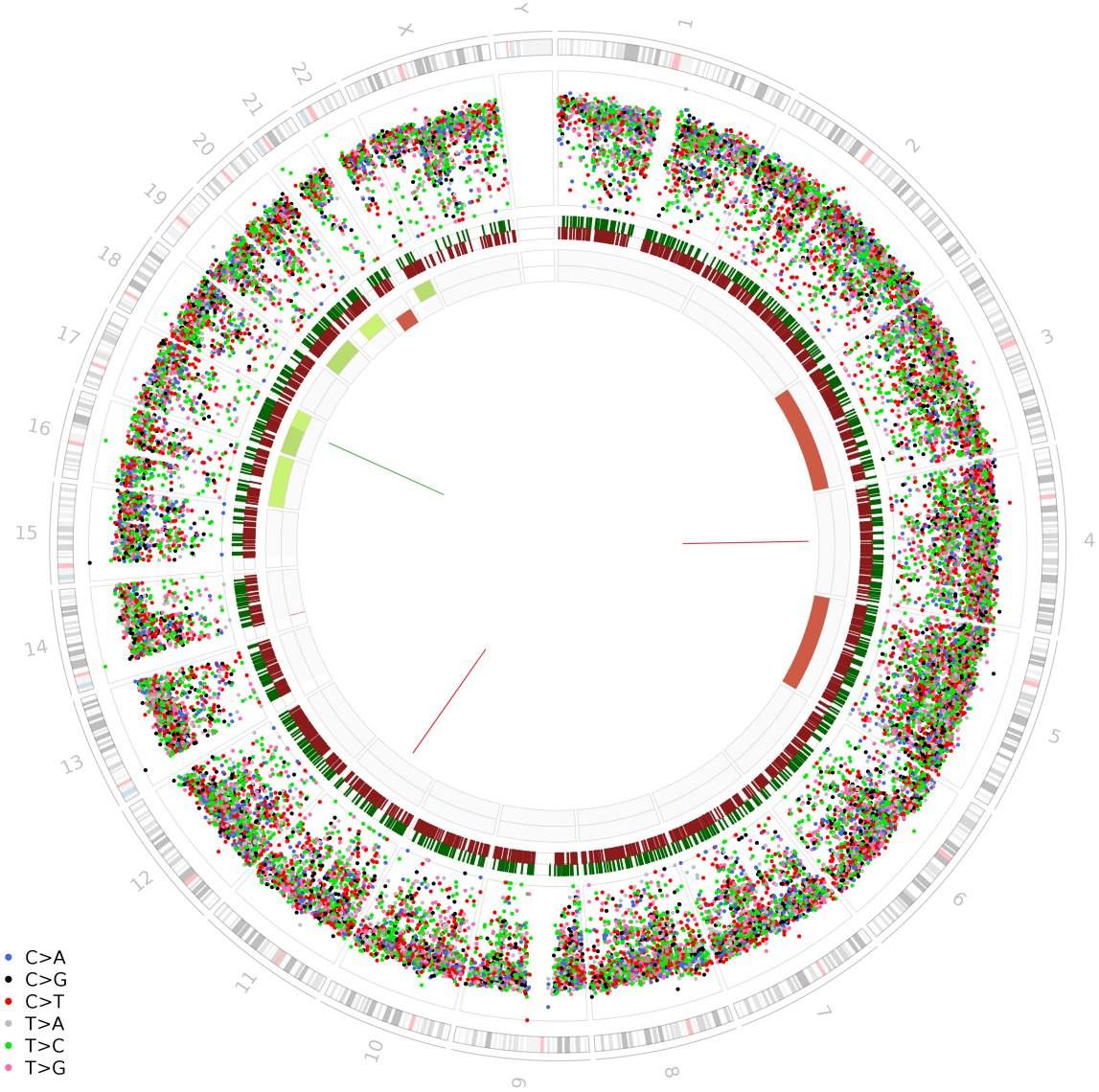
karyotype (purity 0.42, ploidy 2.2)



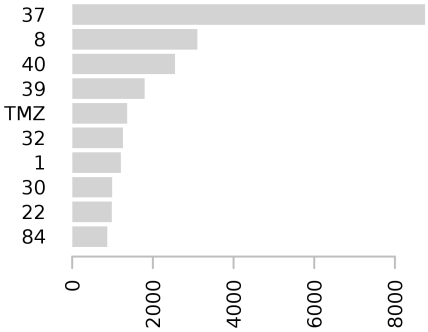
42 rearrangements



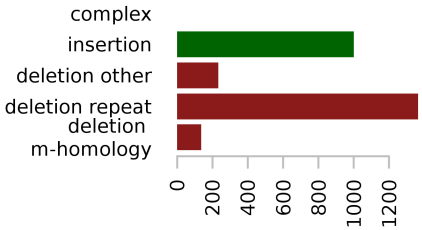
H203514 - FFPE



27203 substitutions of signatures:



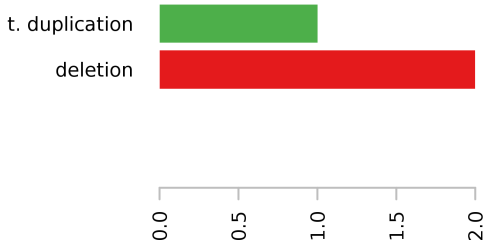
2735 indels



karyotype (purity 0.31, ploidy 2)



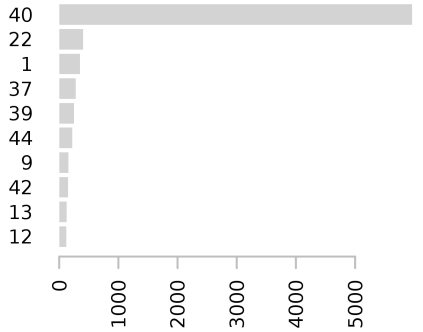
3 rearrangements



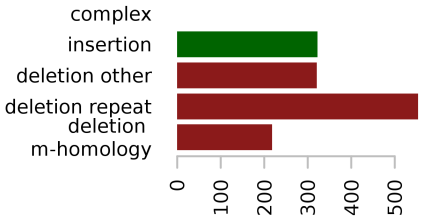
H203514 - Fresh Frozen



8727 substitutions of signatures:



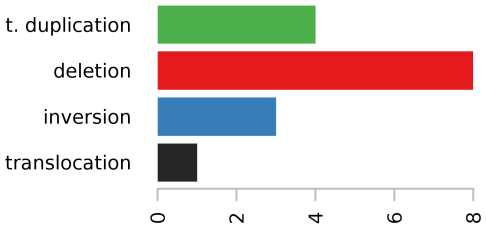
1415 indels



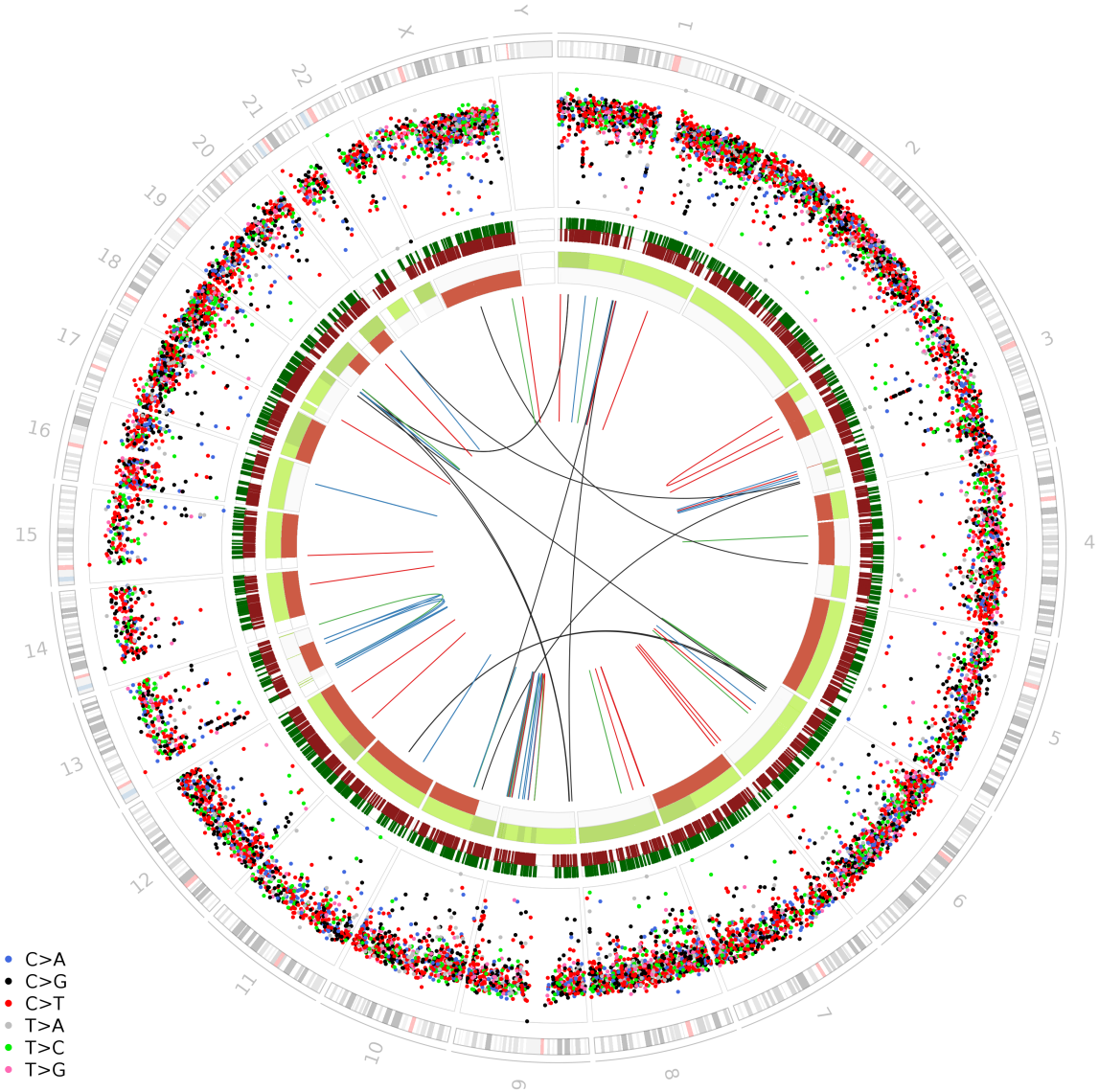
karyotype (purity 0.35, ploidy 3.4)



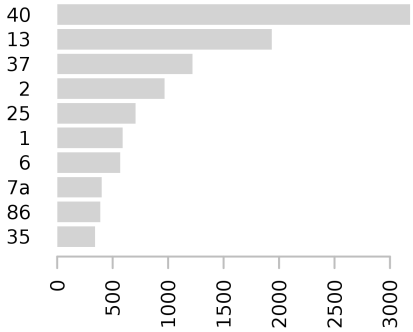
16 rearrangements



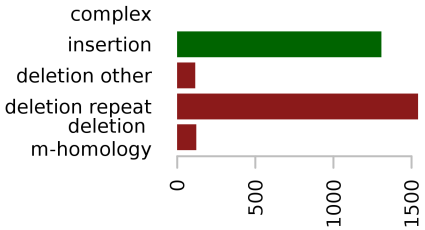
H203515 - FFPE



11249 substitutions of signatures:



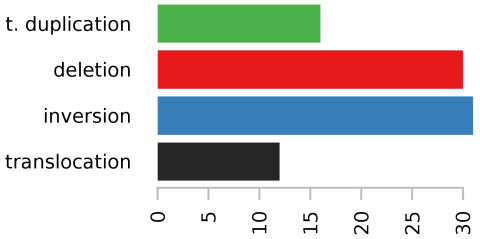
3087 indels



karyotype (purity 0.66, ploidy 4.7)



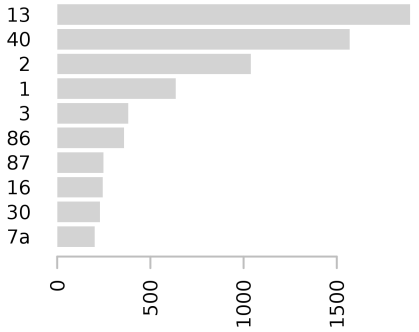
89 rearrangements



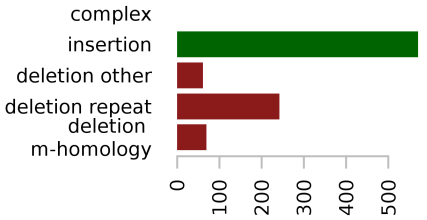
H203515 - Fresh Frozen



8888 substitutions of signatures:



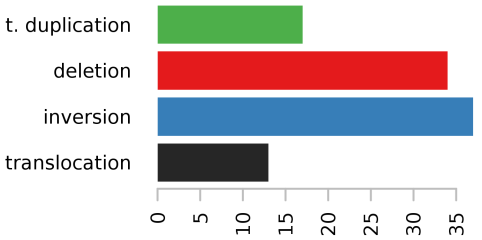
943 indels



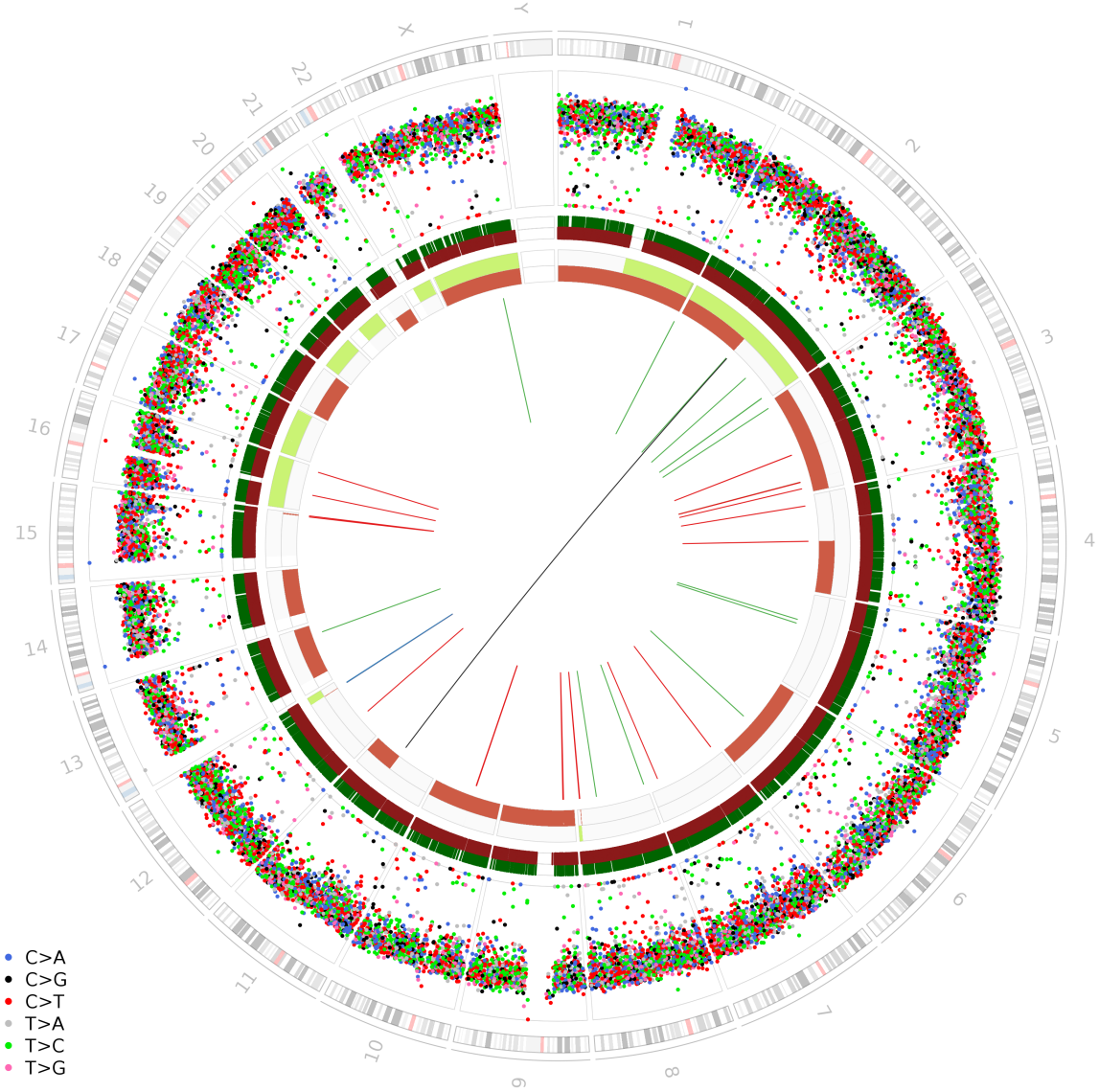
karyotype (purity 0.37, ploidy 3.4)



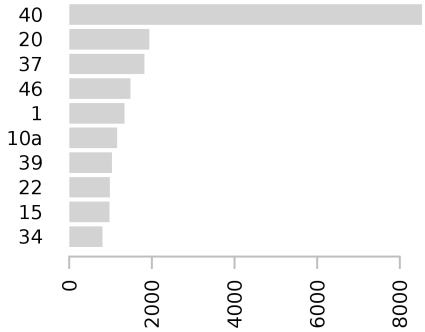
101 rearrangements



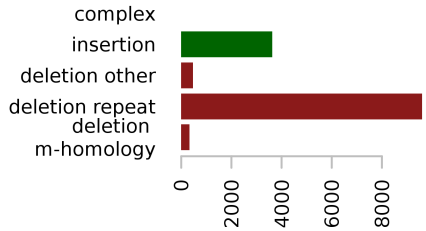
H203516 - FFPE



26464 substitutions of signatures:



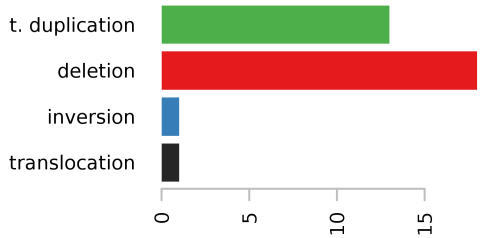
14039 indels



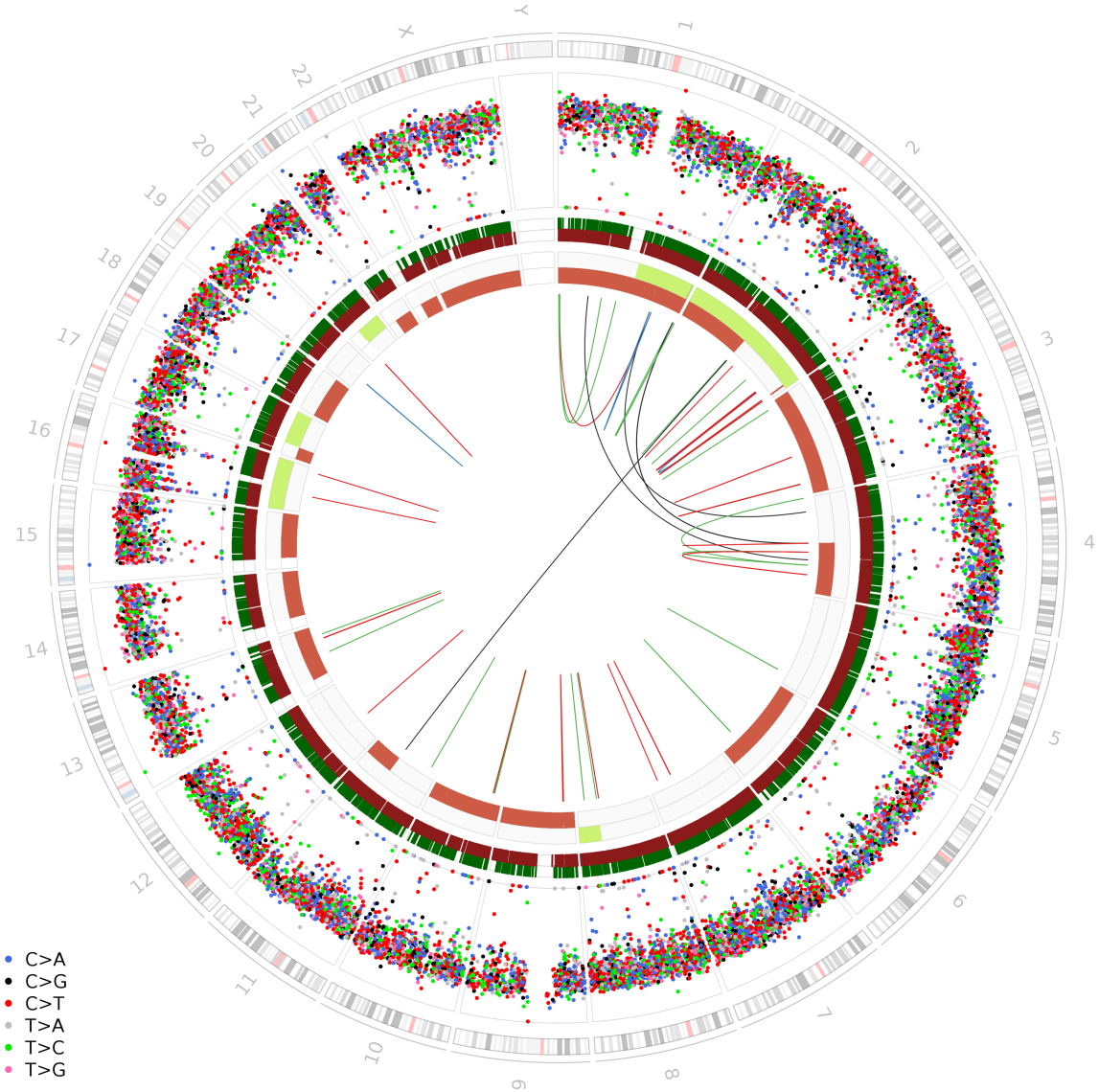
karyotype (purity 0.42, ploidy 2.4)



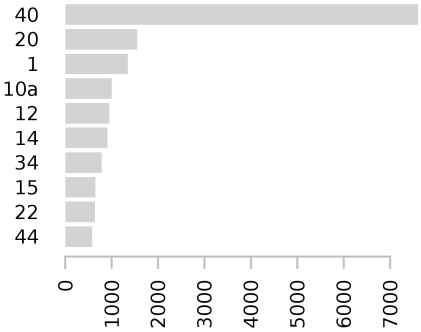
33 rearrangements



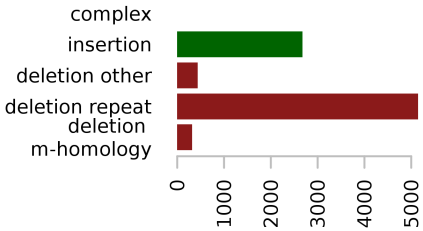
H203516 - Fresh Frozen



21859 substitutions of signatures:



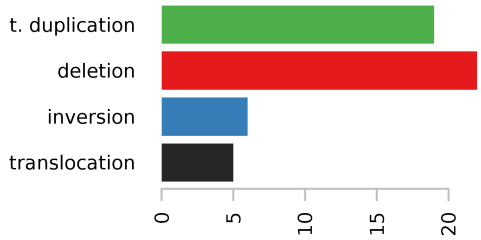
8579 indels



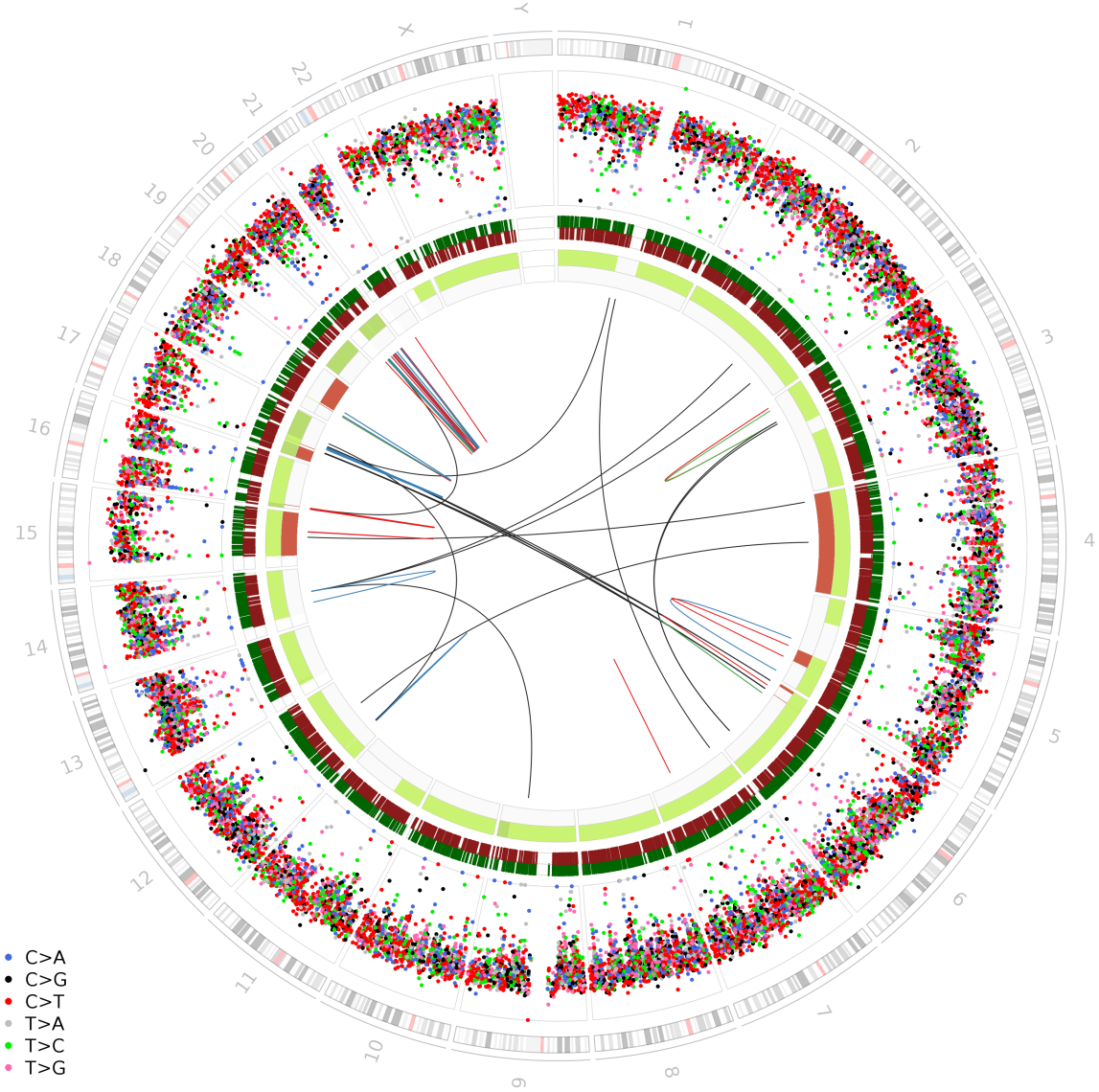
karyotype (purity 0.39, ploidy 1.8)



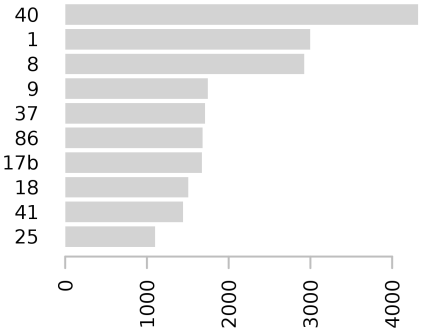
52 rearrangements



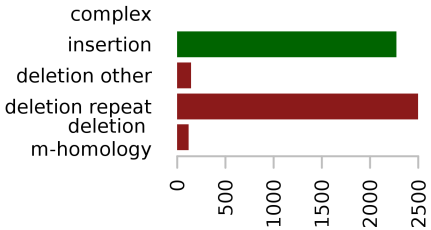
H203517 - FFPE



25690 substitutions of signatures:



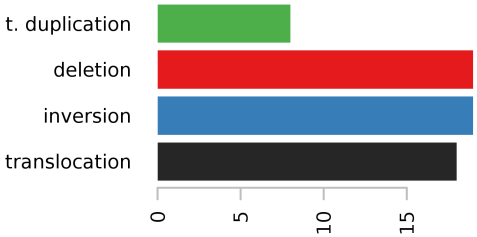
5036 indels



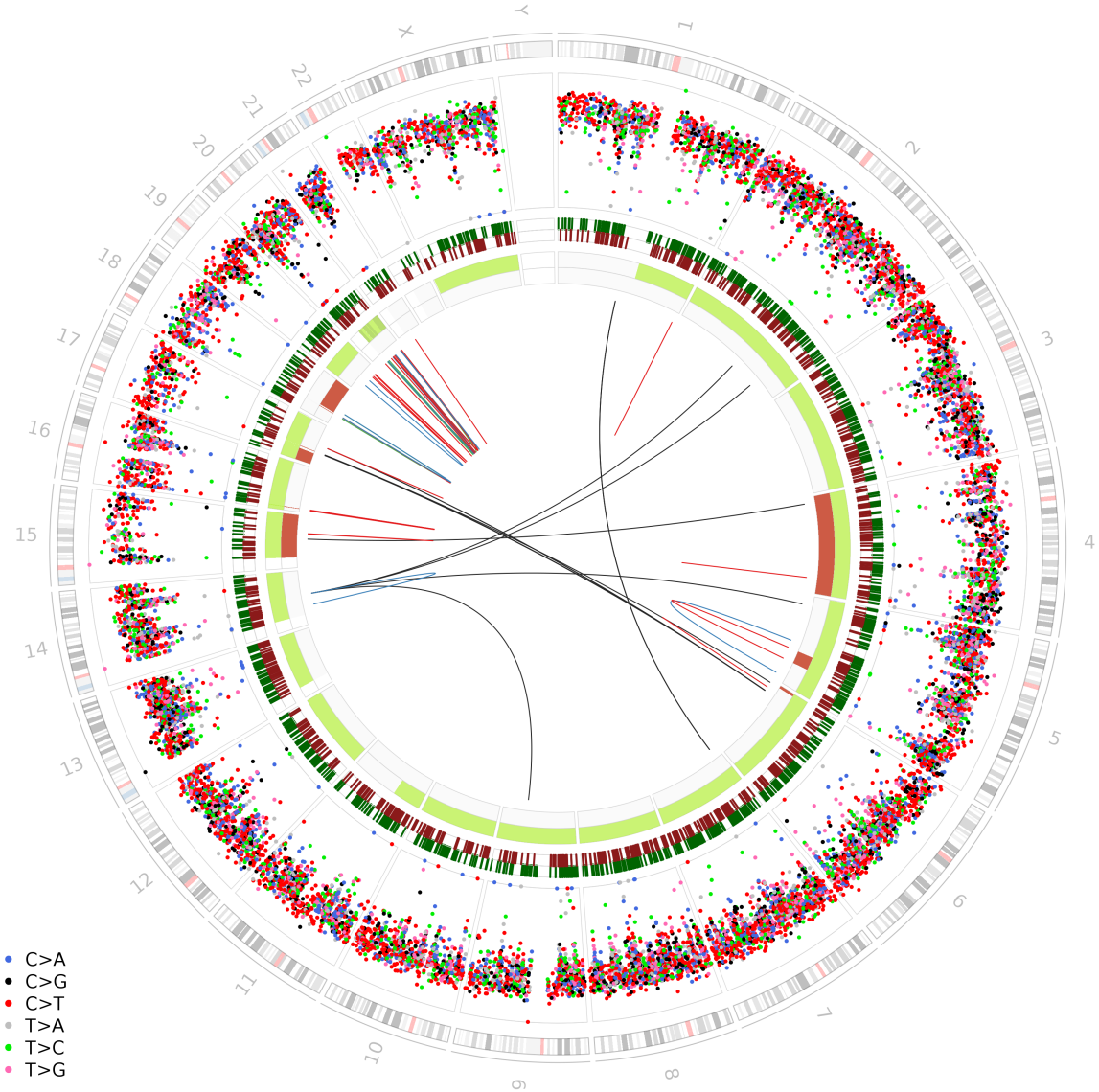
karyotype (purity 0.62, ploidy 5.2)



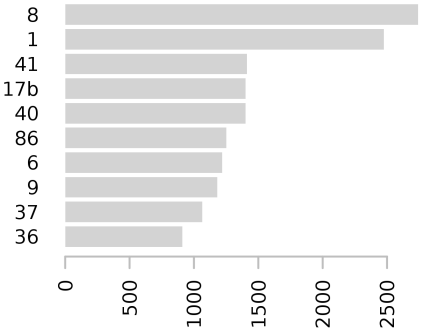
64 rearrangements



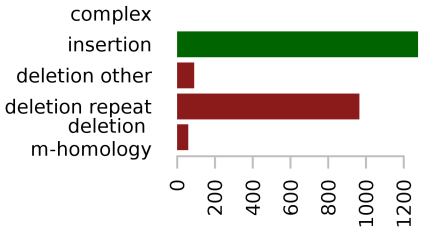
H203517 - Fresh Frozen



18101 substitutions of signatures:



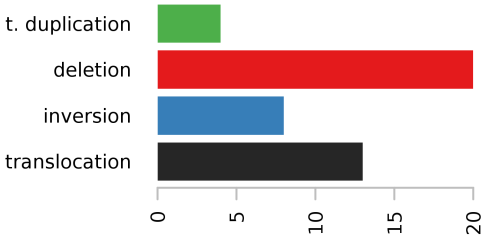
2391 indels



karyotype (purity 0.29, ploidy 3.1)



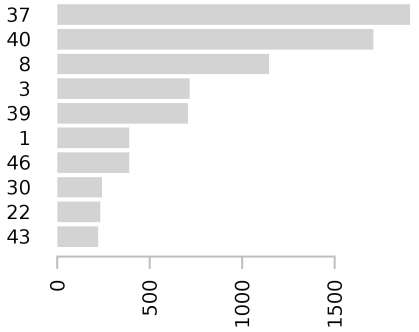
45 rearrangements



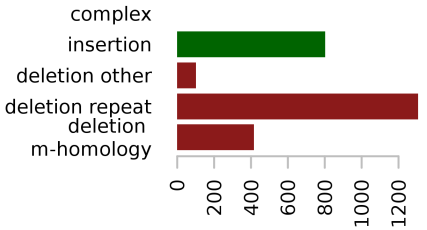
H203518 - FFPE



9509 substitutions of signatures:



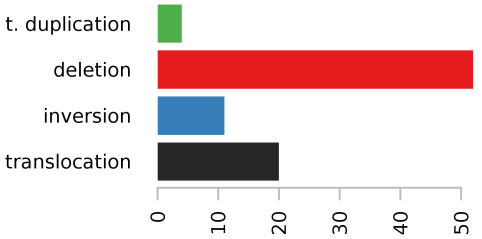
2627 indels



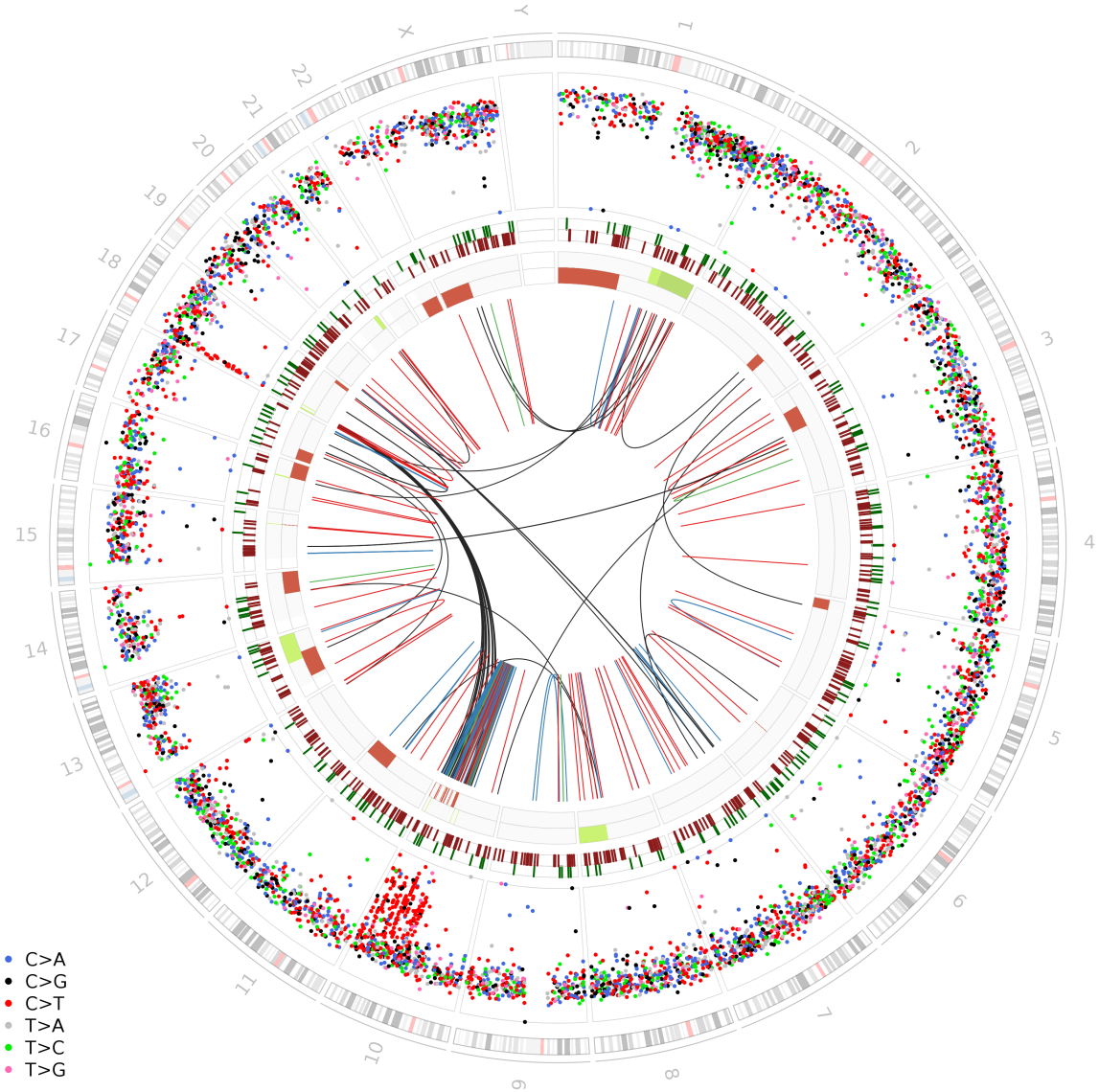
karyotype (purity 0.67, ploidy 2.6)



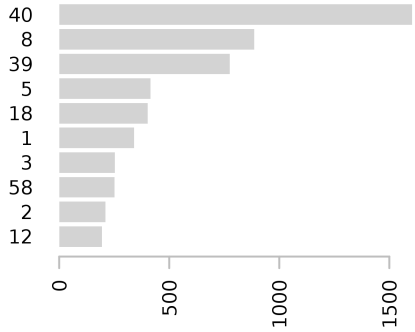
87 rearrangements



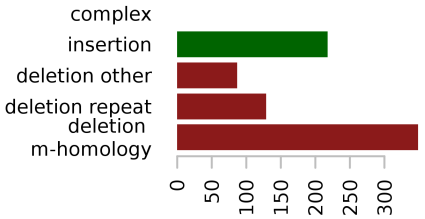
H203518 - Fresh Frozen



6367 substitutions of signatures:



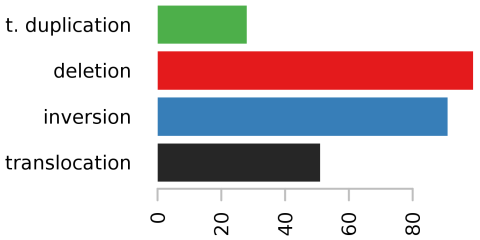
783 indels



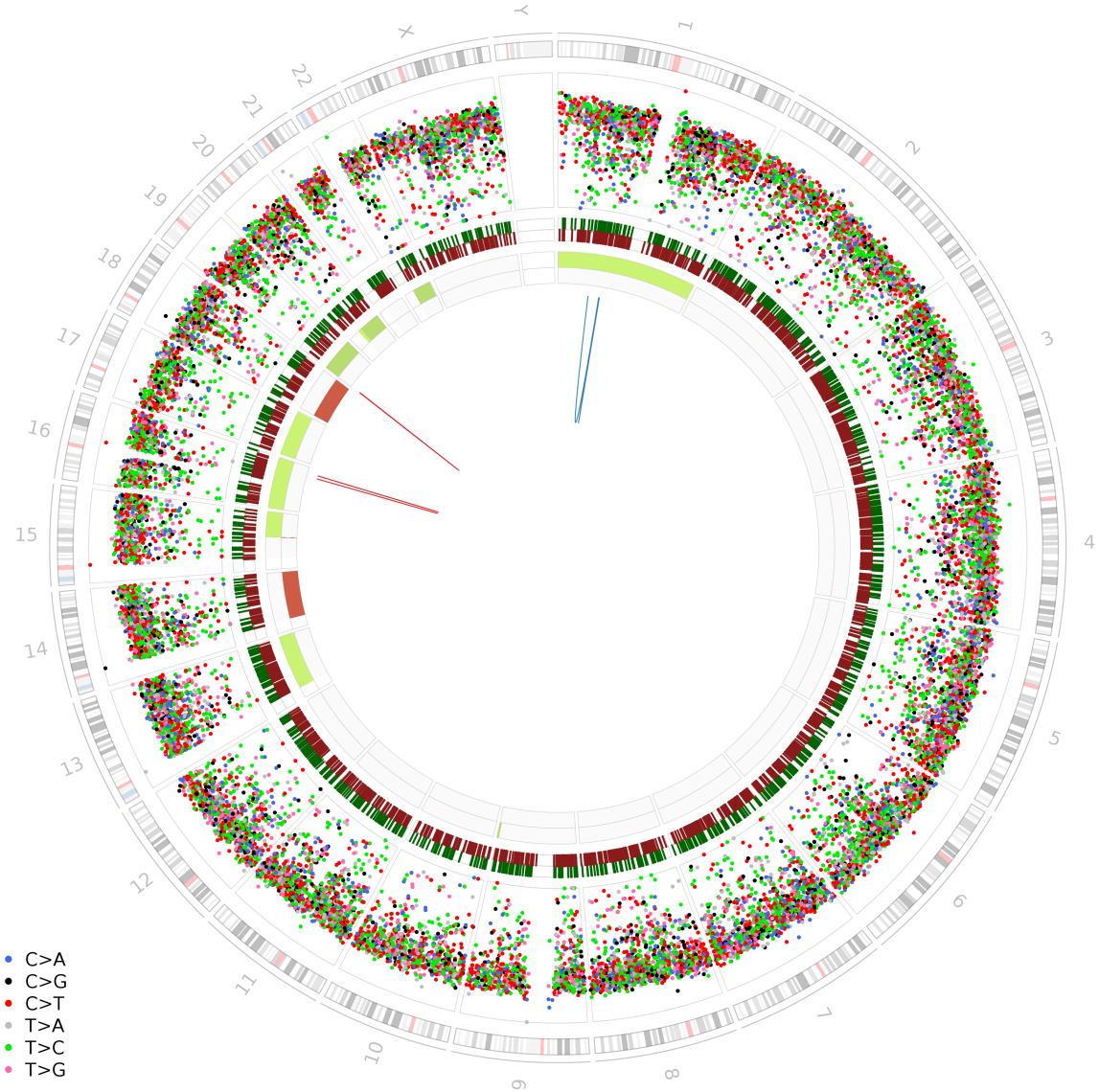
karyotype (purity 0.68, ploidy 2)

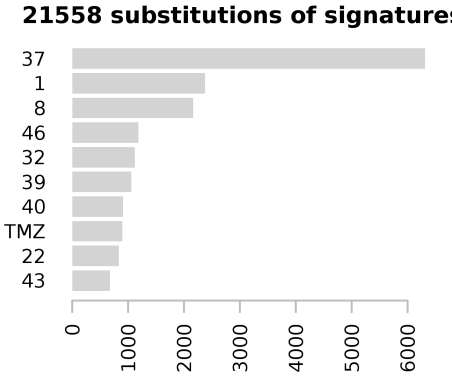


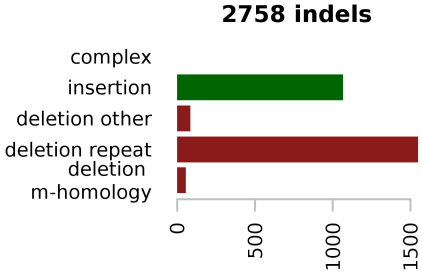
269 rearrangements

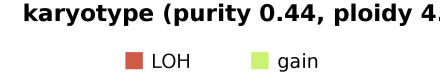


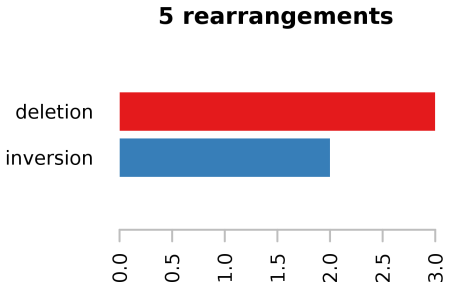
H203519 - FFPE



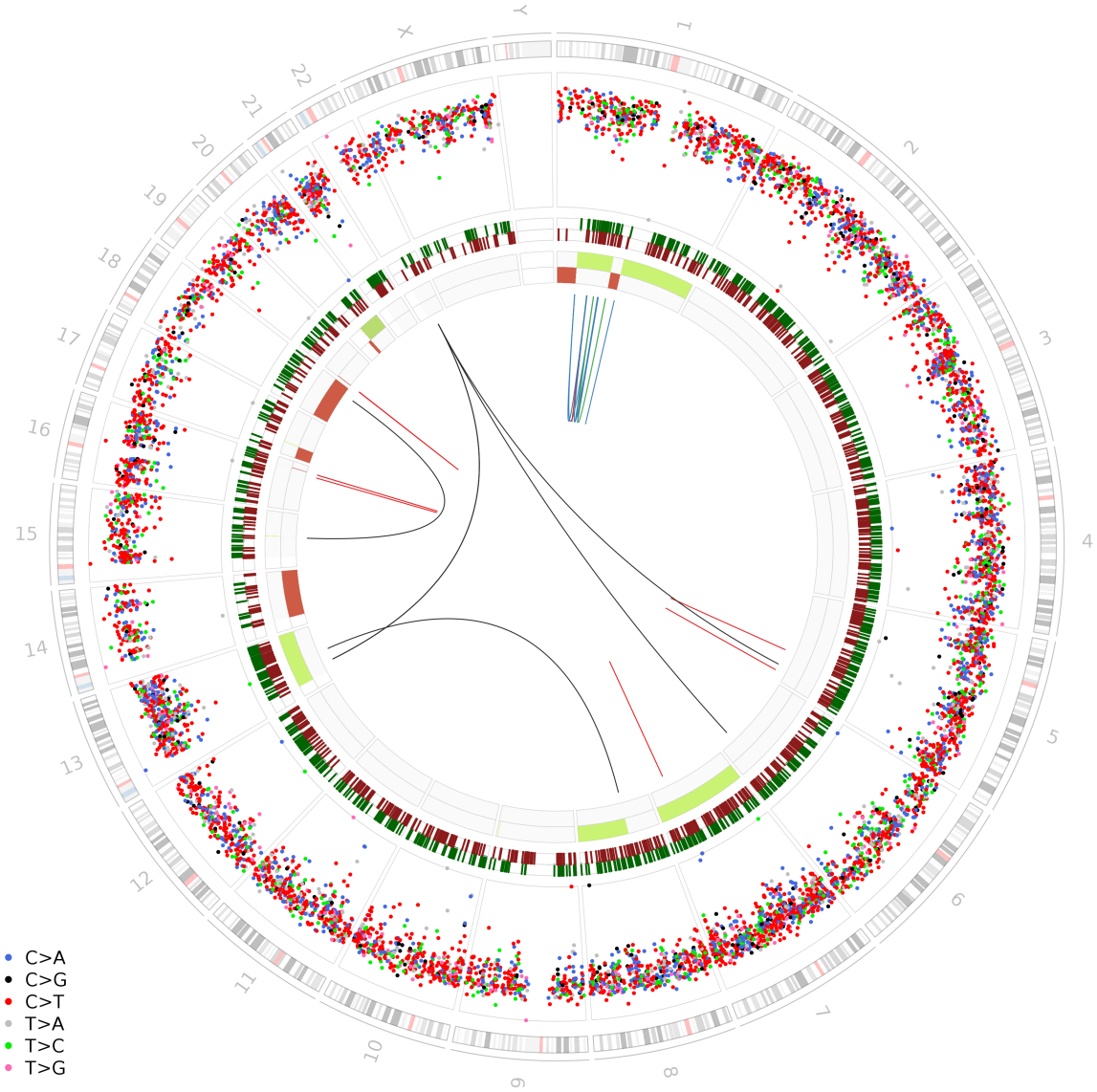




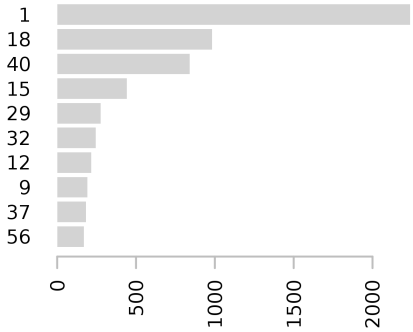




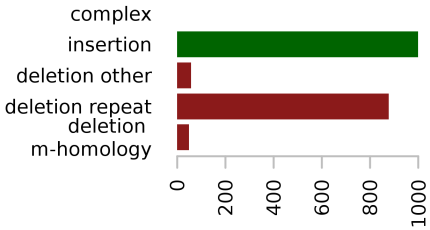
H203519 - Fresh Frozen



7676 substitutions of signatures:



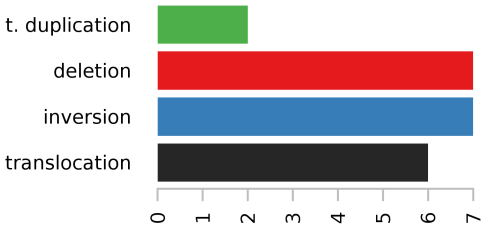
1985 indels



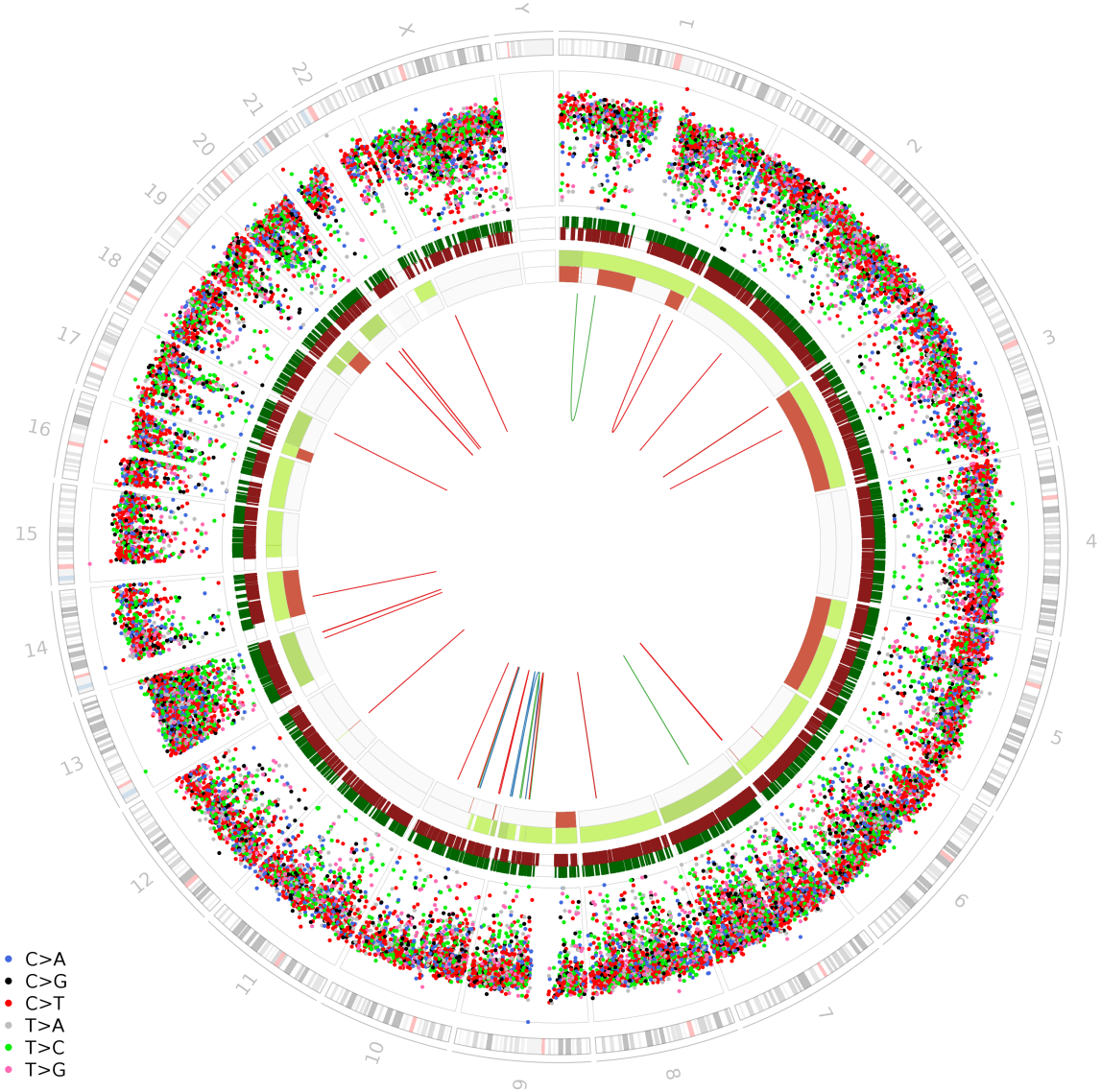
karyotype (purity 0.66, ploidy 2.2)



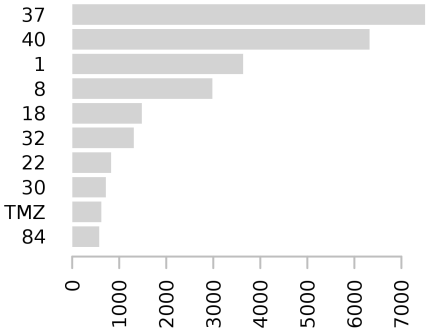
22 rearrangements



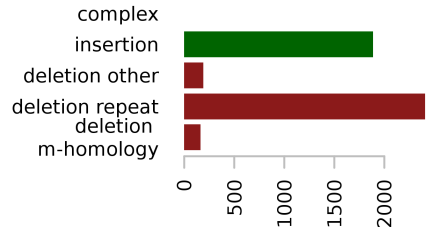
H203520 - FFPE



29242 substitutions of signatures:



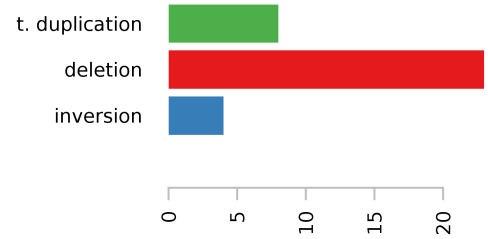
4649 indels



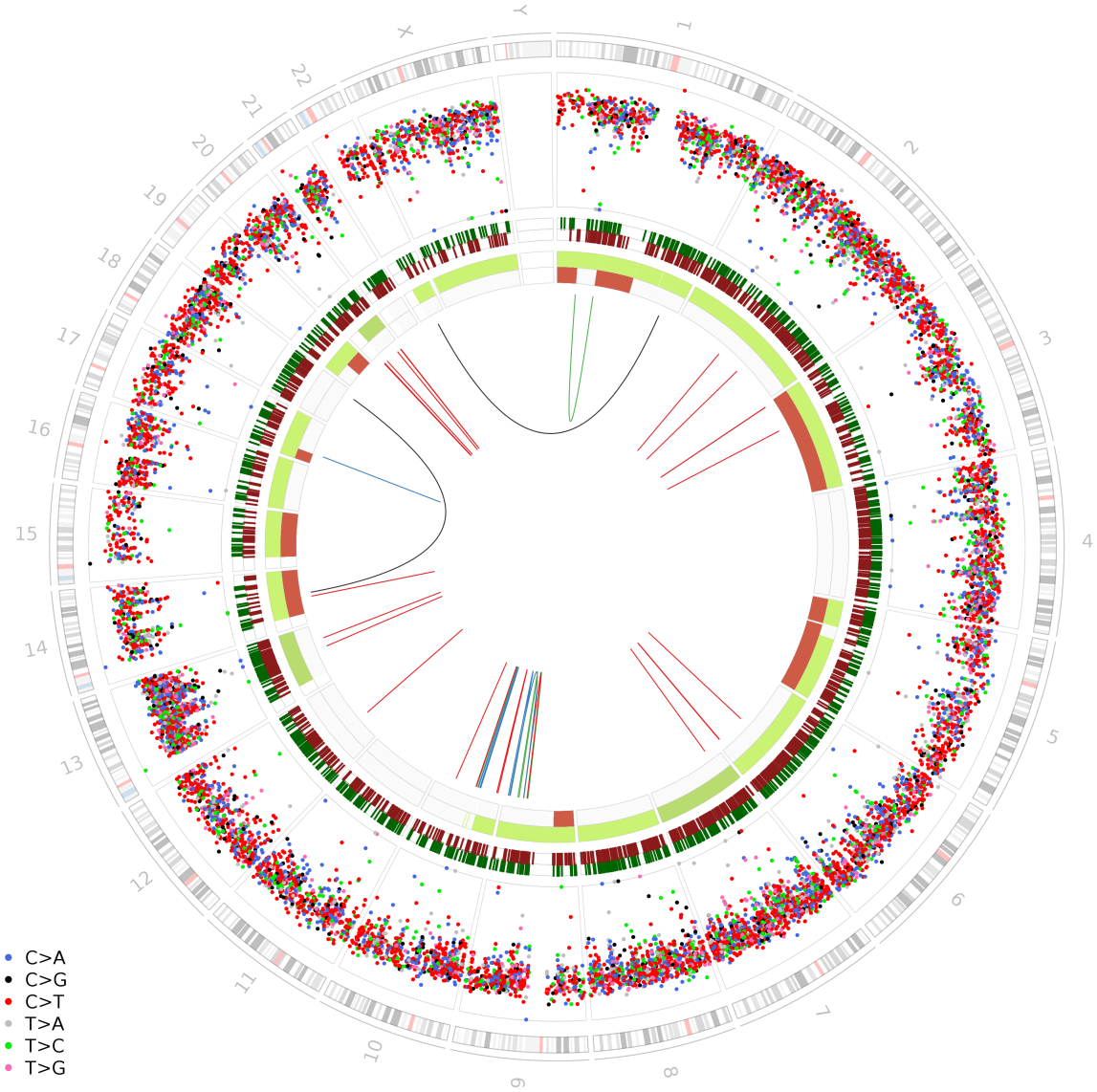
karyotype (purity 0.87, ploidy 4.8)



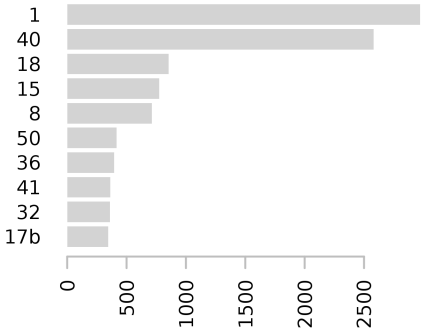
35 rearrangements



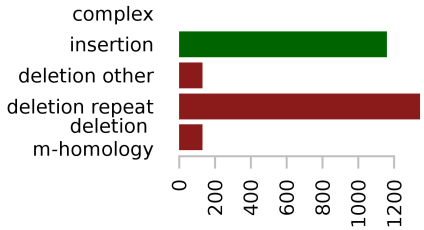
H203520 - Fresh Frozen



12836 substitutions of signatures:



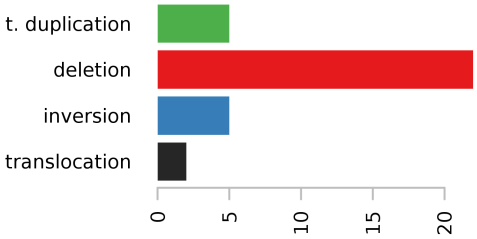
2765 indels

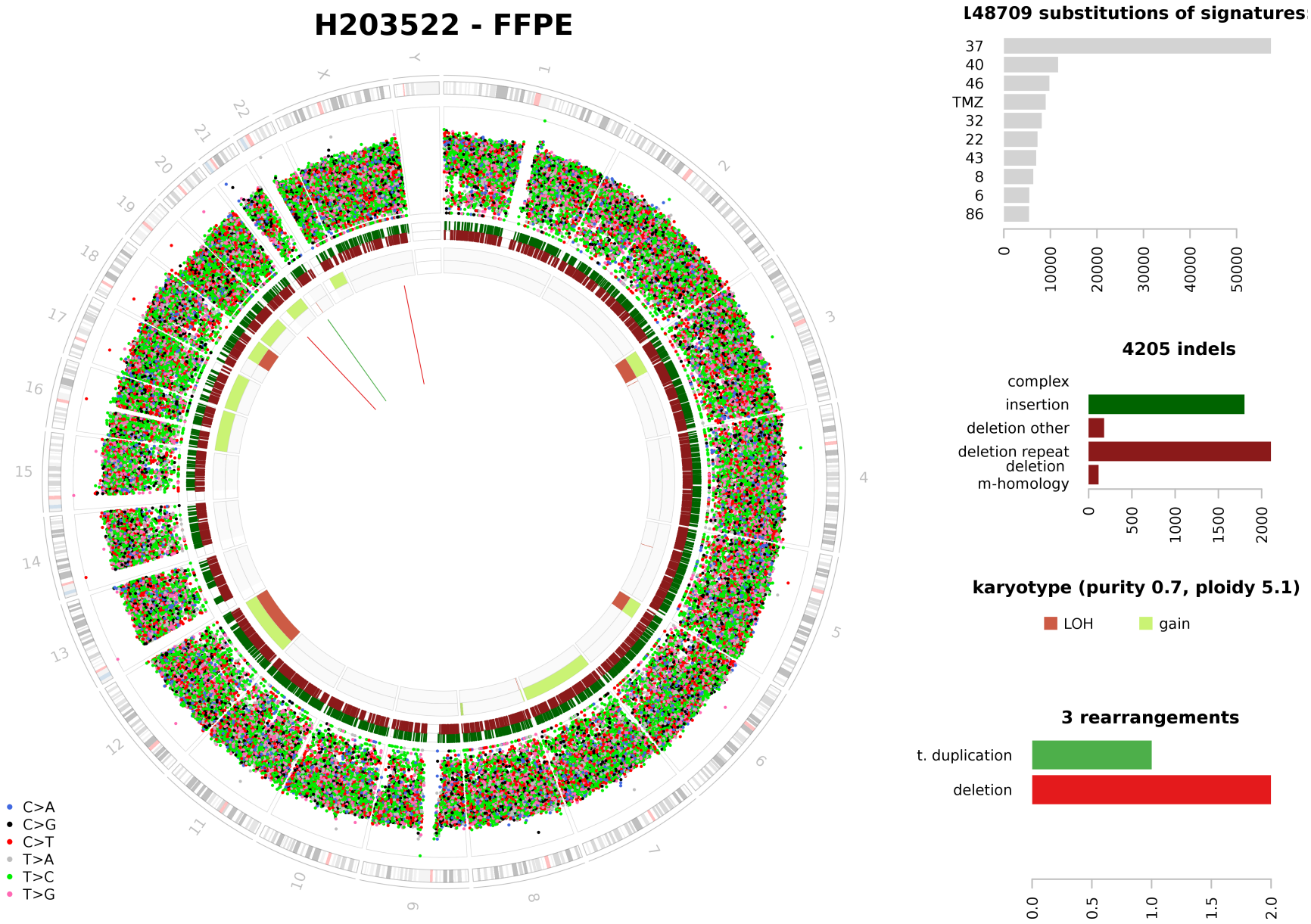


karyotype (purity 0.51, ploidy 3.1)

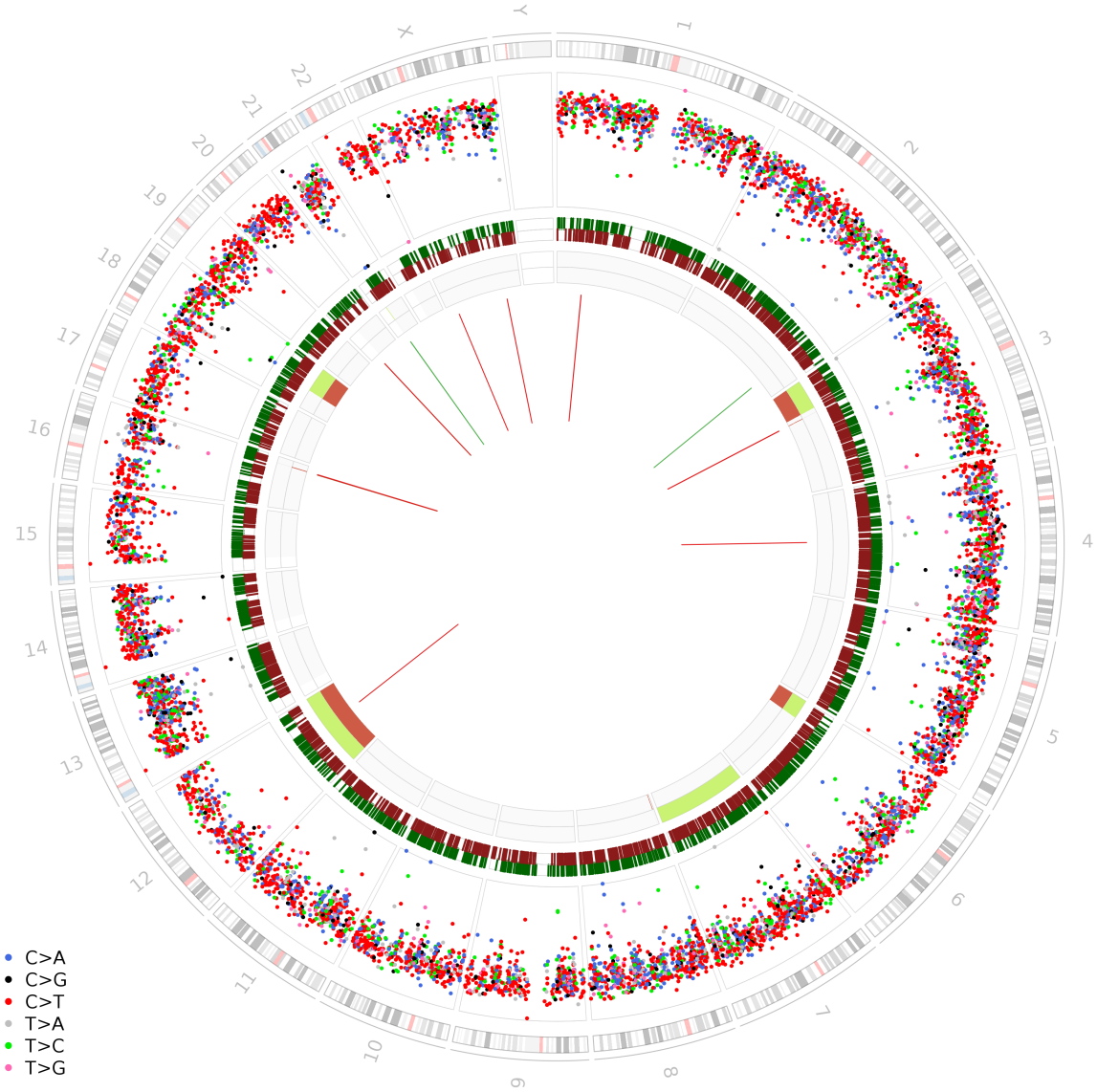


34 rearrangements

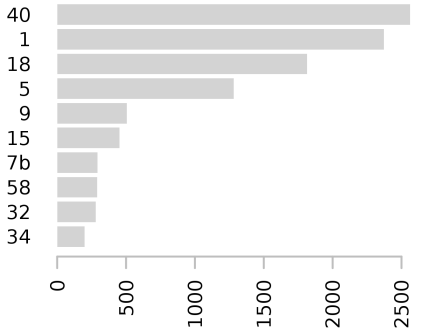



H203522 - FFPE
C>A
C>G
C>T
T>A
T>C
T>G
148709 substitutions of signatures
37
40
46
TMZ
32
22
43
8
6
86
0
10000
20000
30000
40000
50000
4205 indels
complex
insertion
deletion other
deletion repeat
deletion
m-homology
0
500
1000
1500
2000
karyotype (purity 0.7, ploidy 5.1)
LOH
gain
3 rearrangements
t. duplication
deletion
0.0
0.5
1.0
1.5
2.0

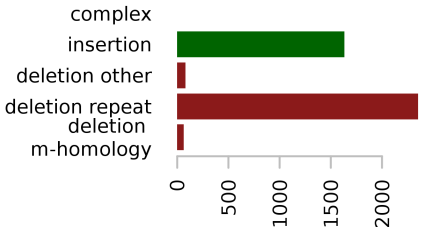
H203522 - Fresh Frozen



9996 substitutions of signatures:



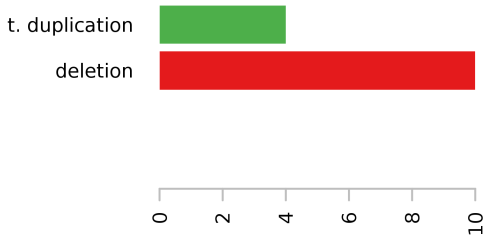
4131 indels



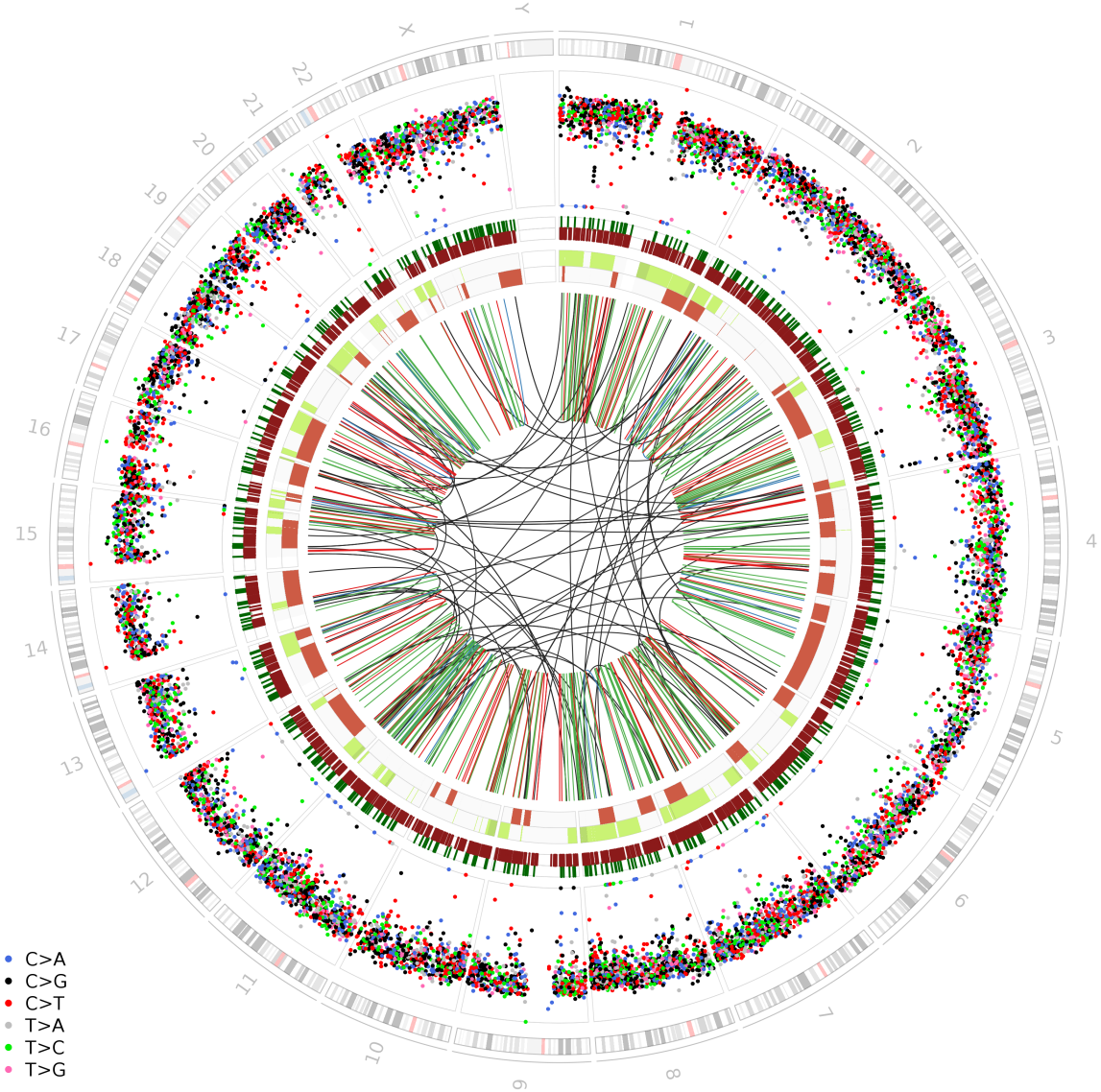
karyotype (purity 0.68, ploidy 2.1)



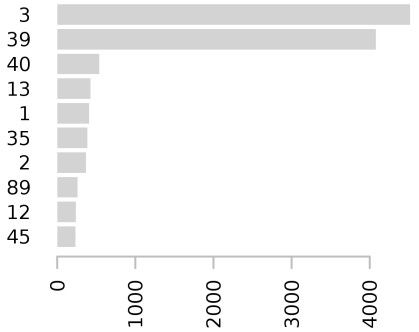
14 rearrangements



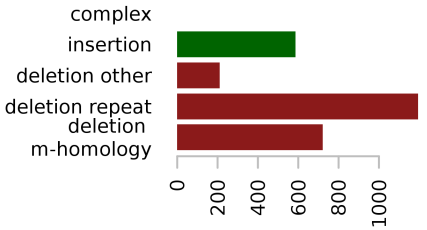
H203523 - FFPE



12825 substitutions of signatures:



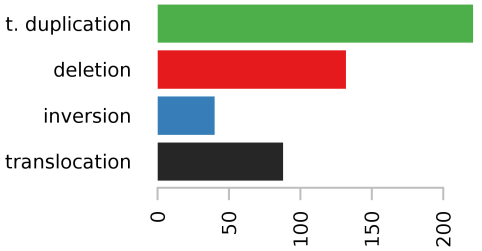
2714 indels



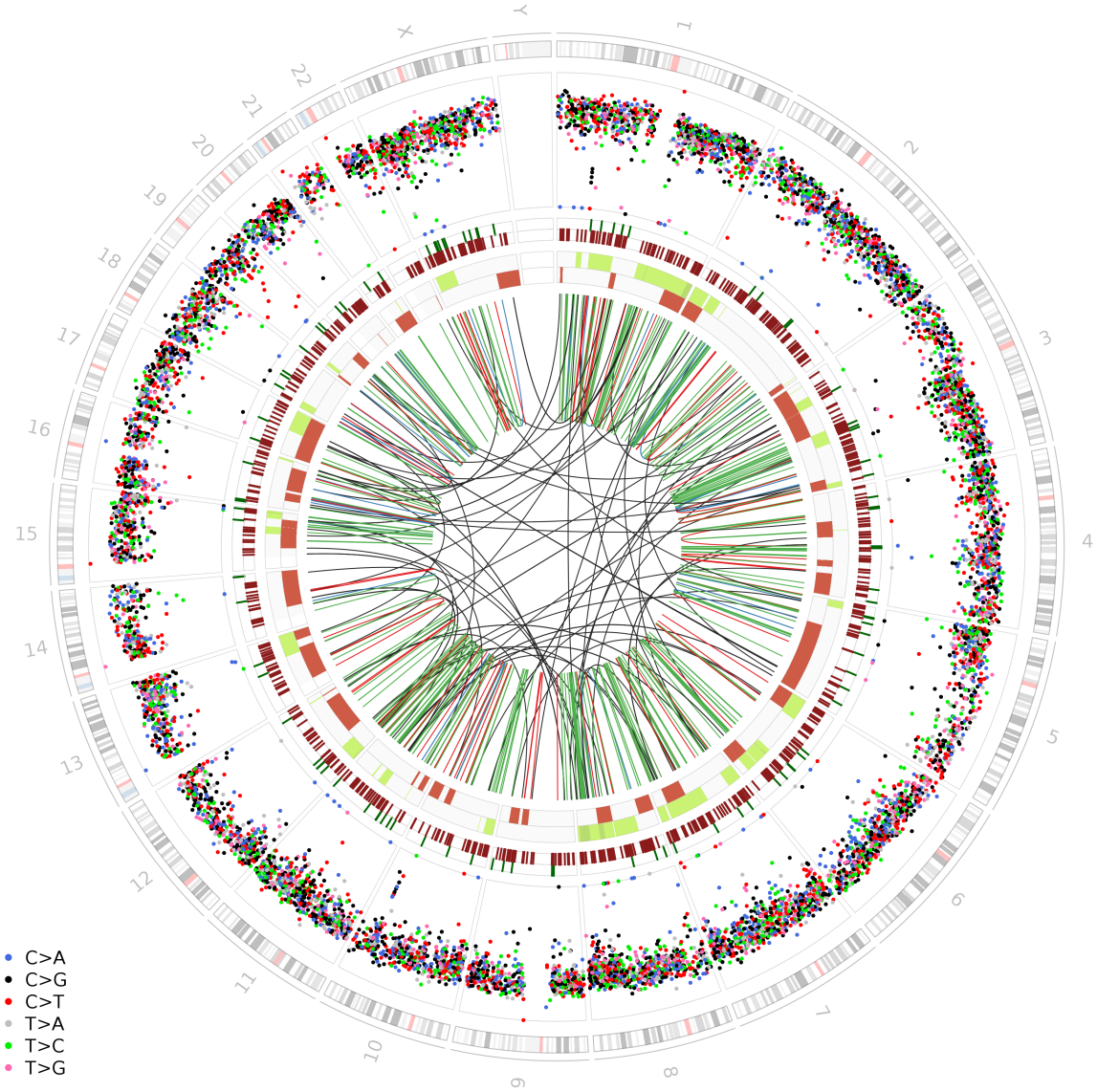
karyotype (purity 0.77, ploidy 2.7)



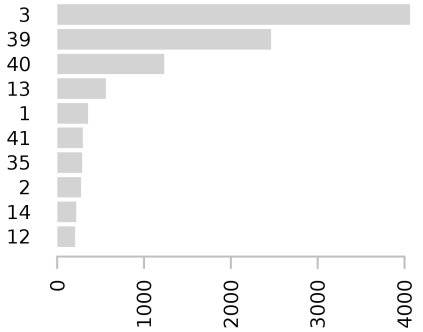
481 rearrangements



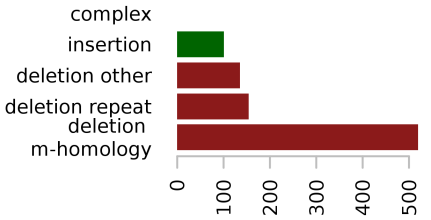
H203523 - Fresh Frozen



10616 substitutions of signatures:



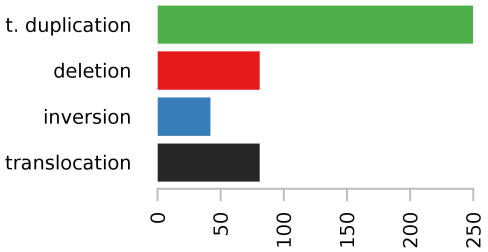
910 indels



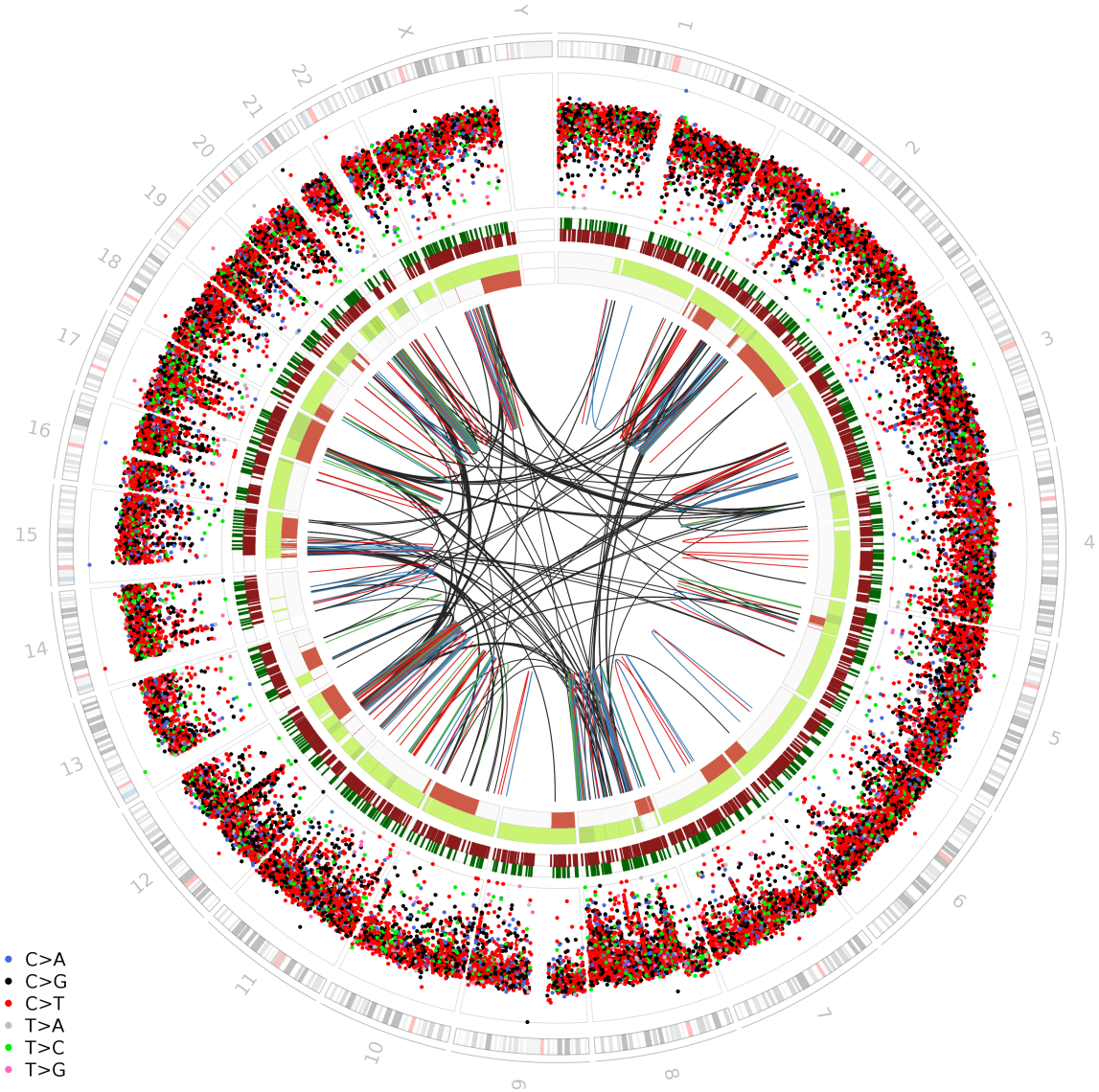
karyotype (purity 0.3, ploidy 2)



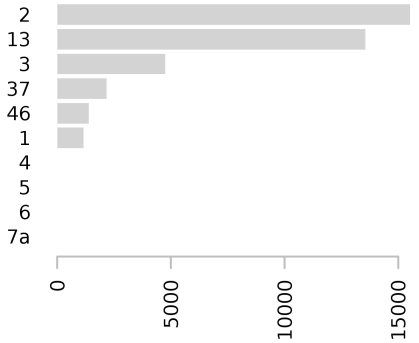
454 rearrangements



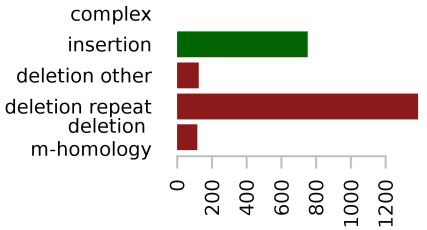
H203524 - FFPE



39235 substitutions of signatures:



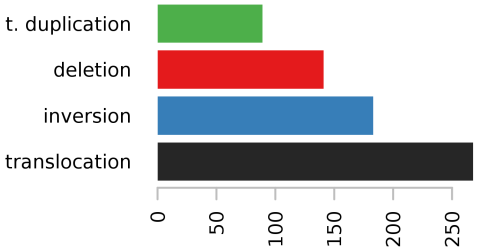
2378 indels



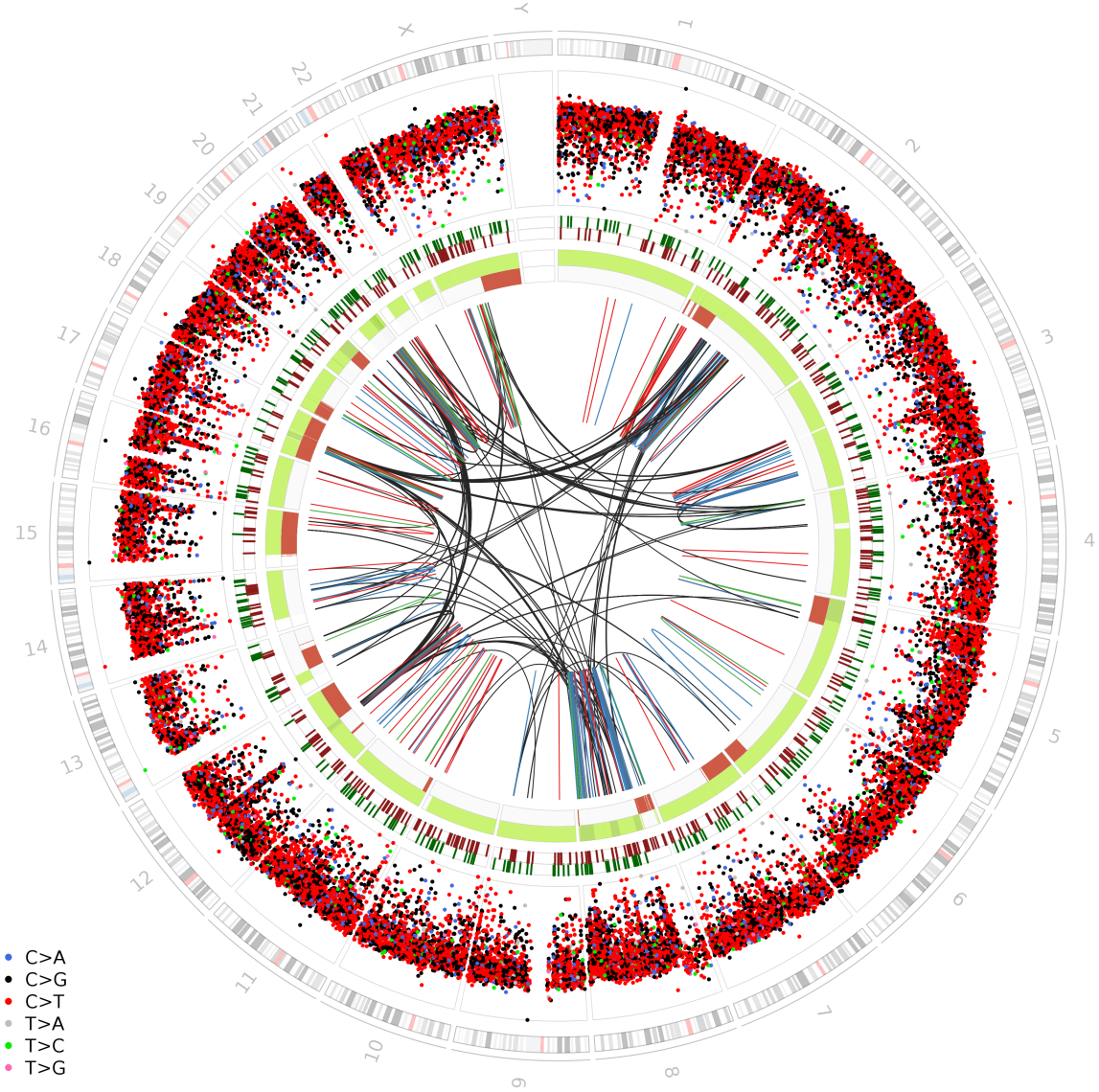
karyotype (purity 0.51, ploidy 4.4)



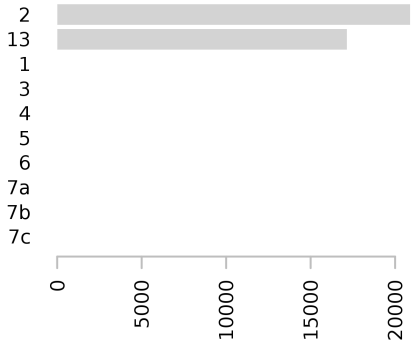
681 rearrangements



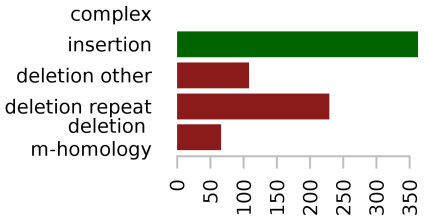
H203524 - Fresh Frozen



42103 substitutions of signatures:



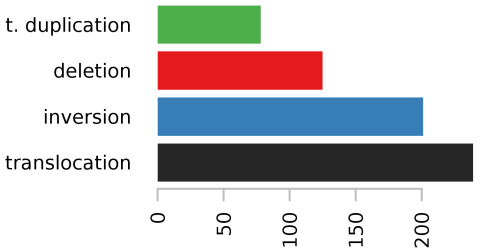
766 indels



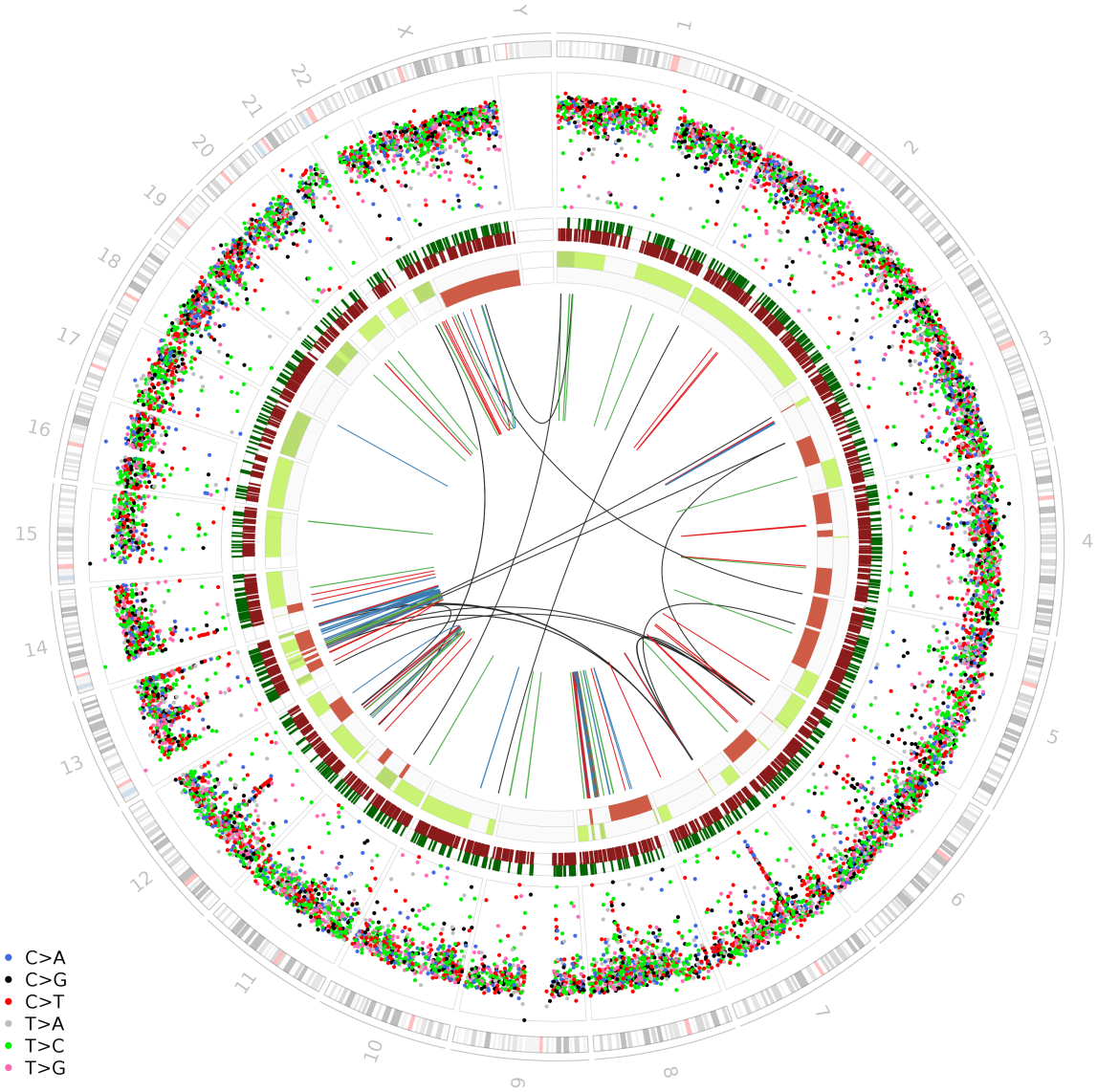
karyotype (purity 0.35, ploidy 3.7)



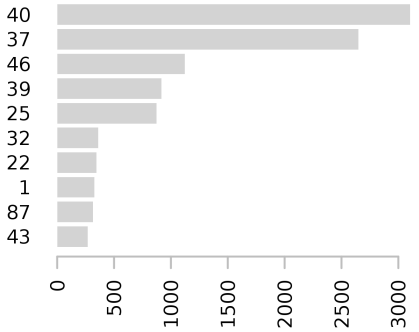
643 rearrangements



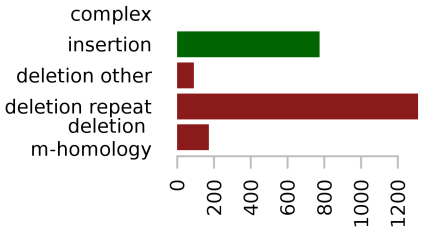
H203525 - FFPE



11704 substitutions of signatures:



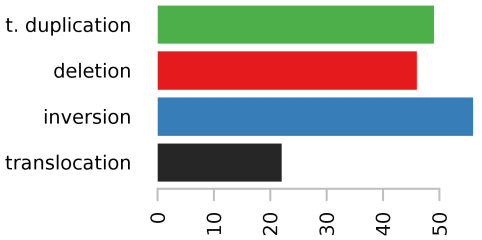
2350 indels



karyotype (purity 0.21, ploidy 4.9)



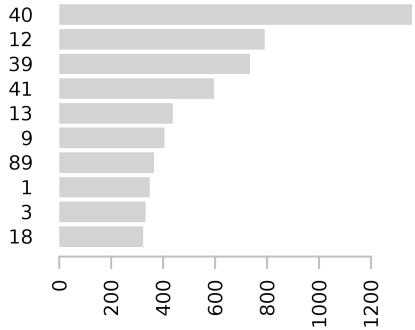
173 rearrangements



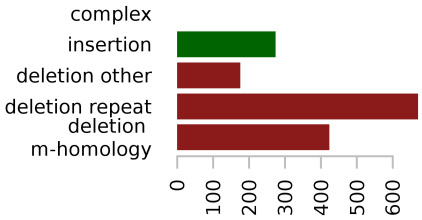
H203525 - Fresh Frozen



9081 substitutions of signatures:



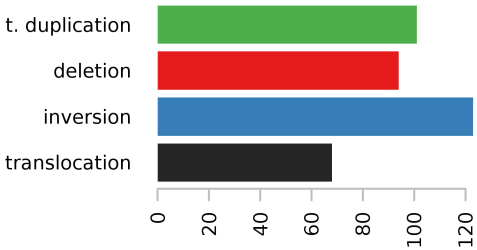
1545 indels



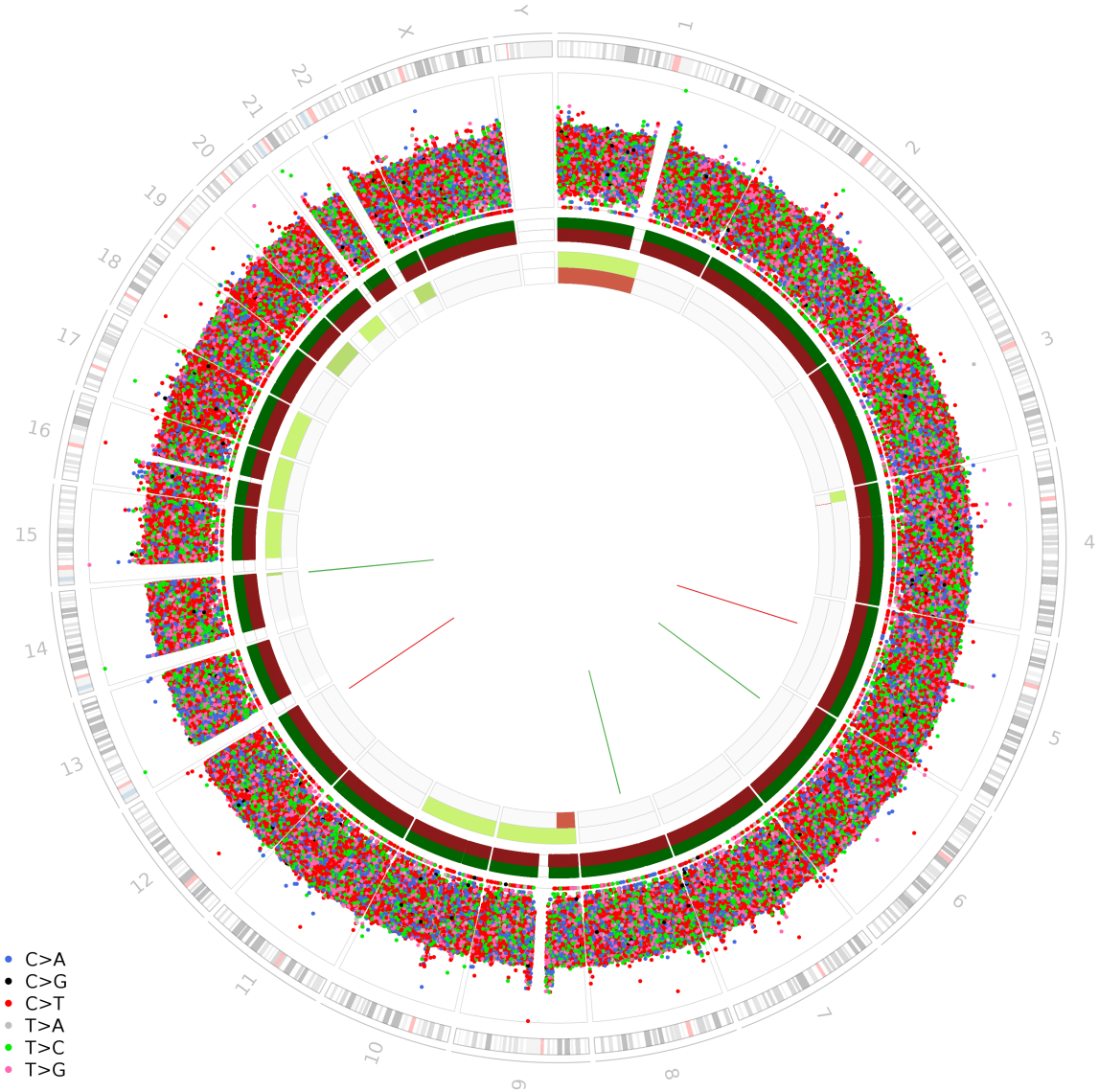
karyotype (purity 0.3, ploidy 3.1)

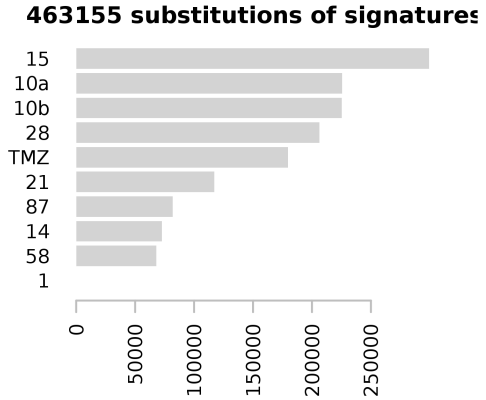


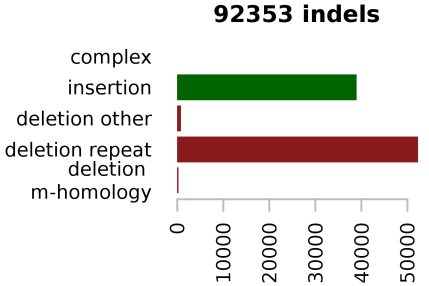
386 rearrangements



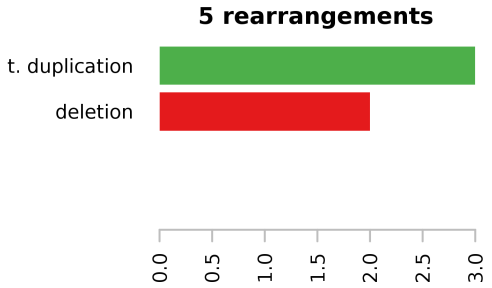
H203526 - FFPE



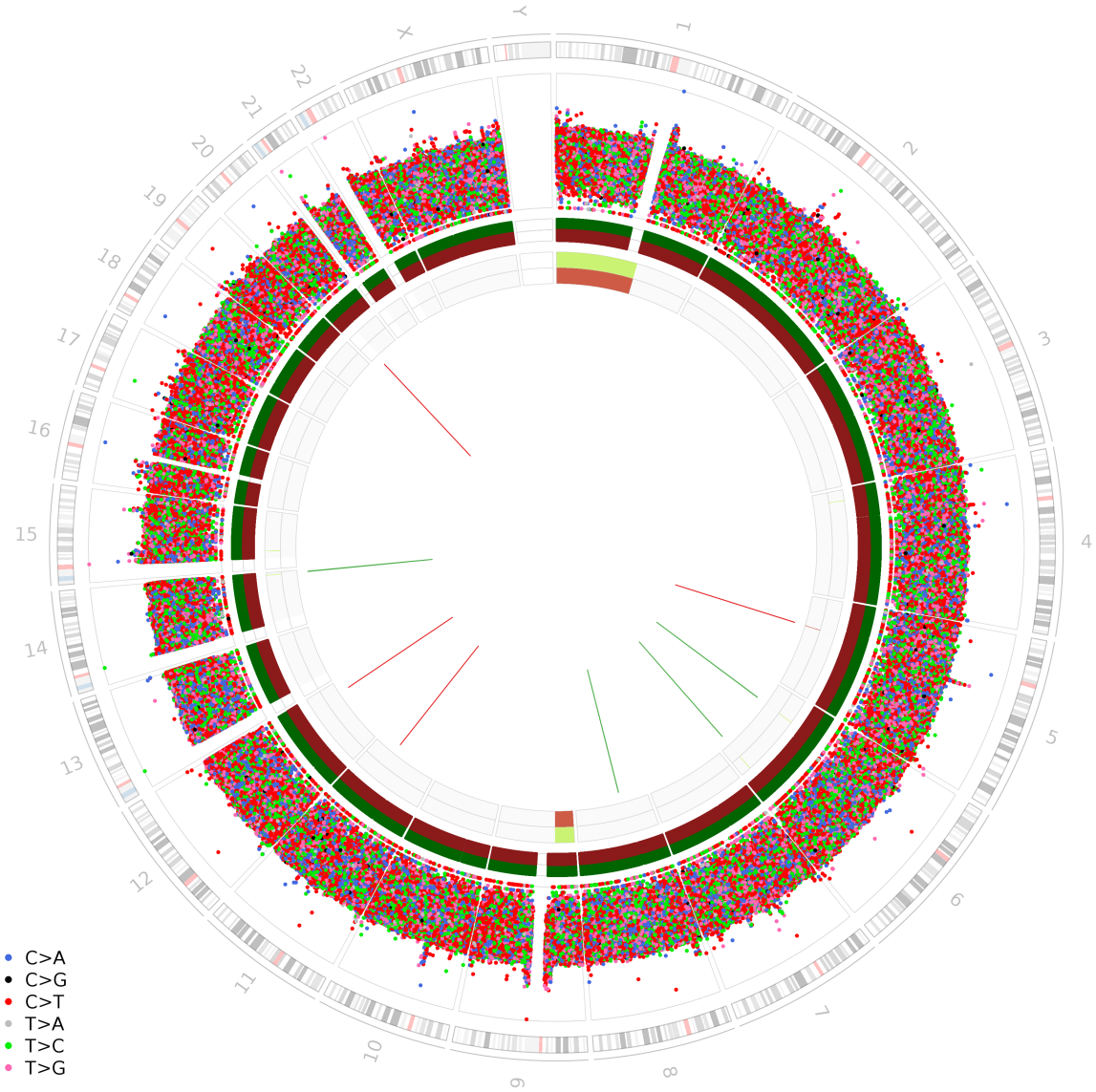


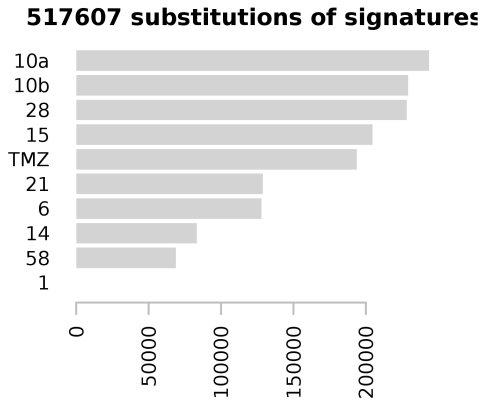


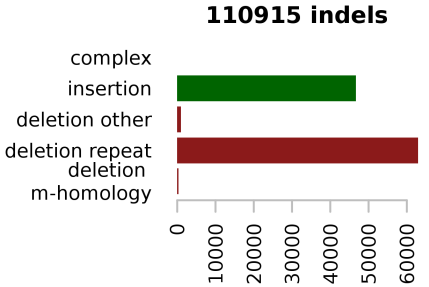




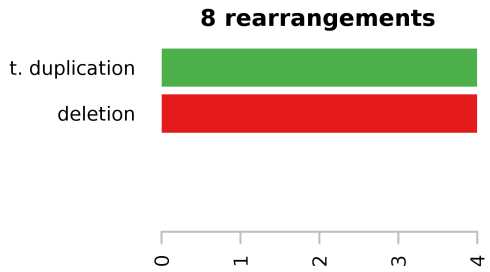
H203526 - Fresh Frozen



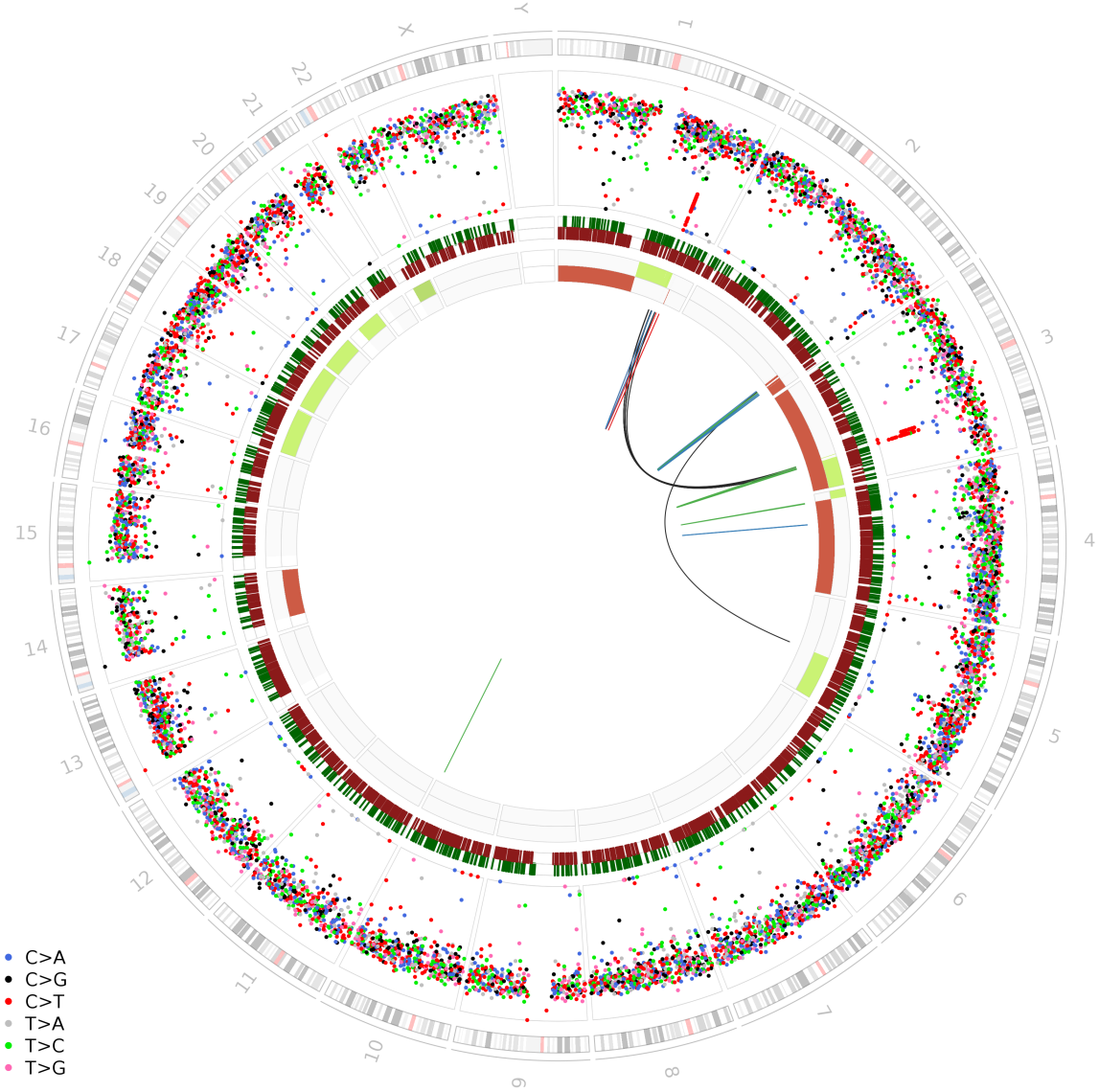


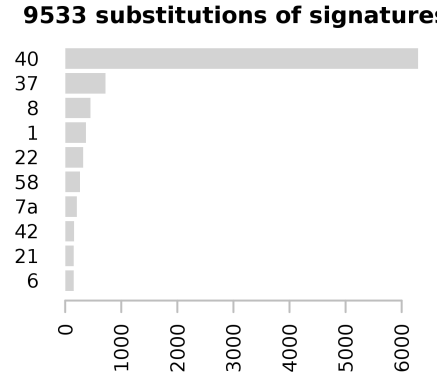


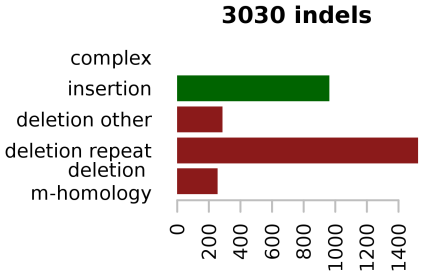




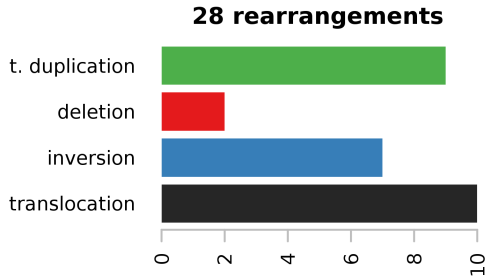
H203527 - FFPE



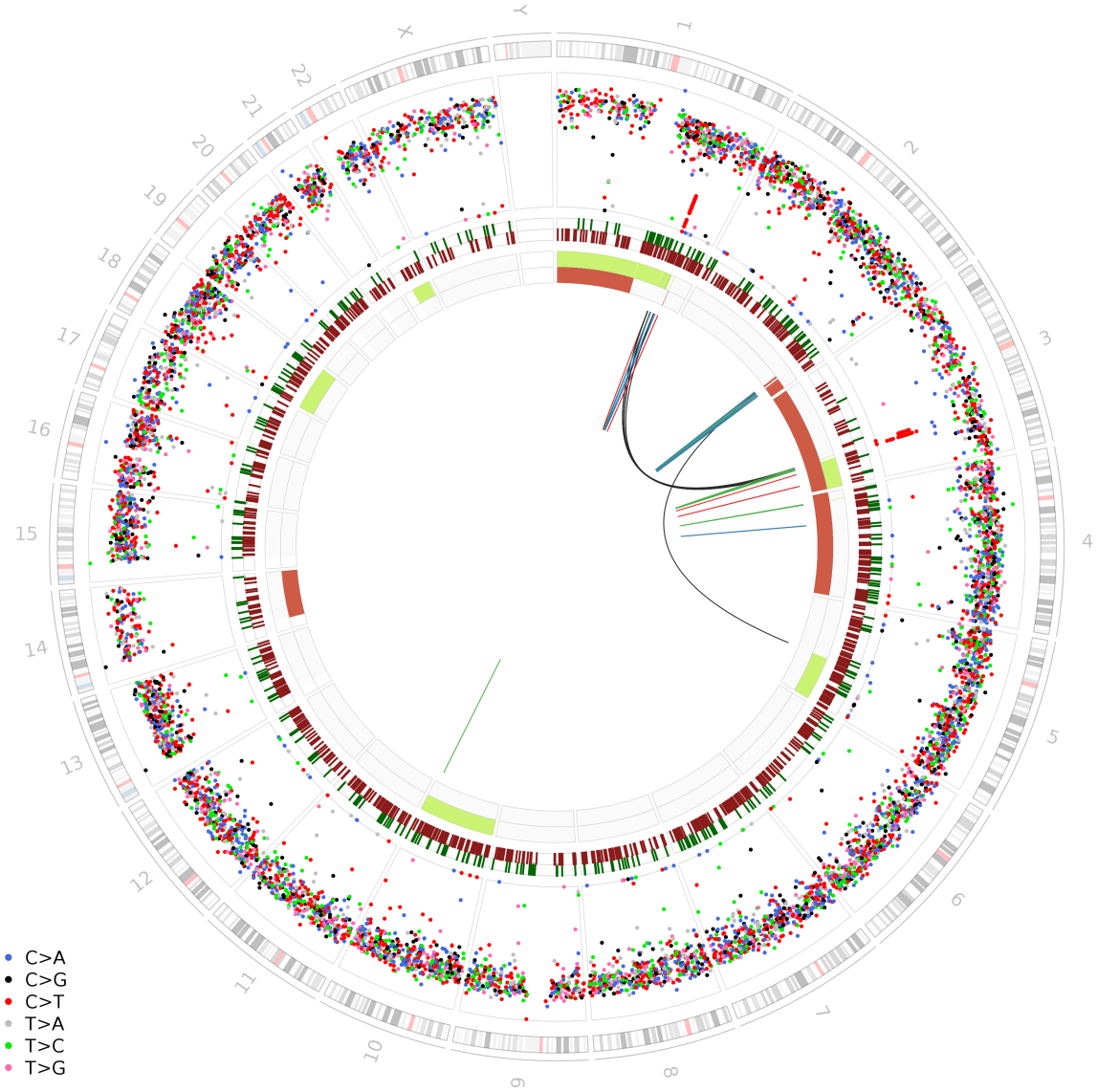




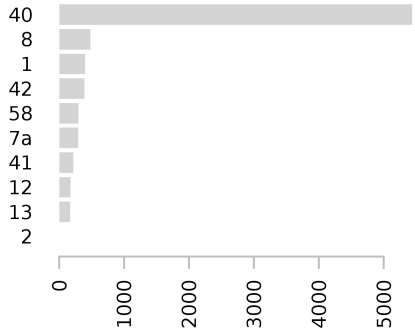




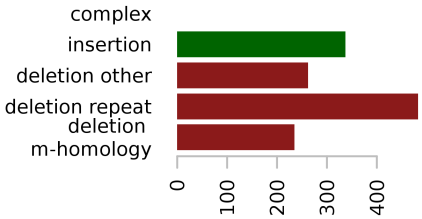
H203527 - Fresh Frozen



8013 substitutions of signatures:



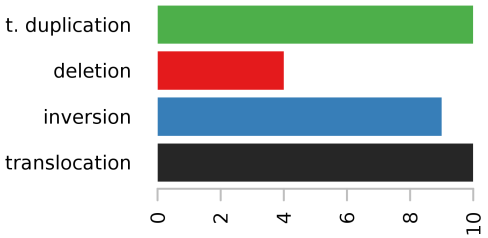
1317 indels



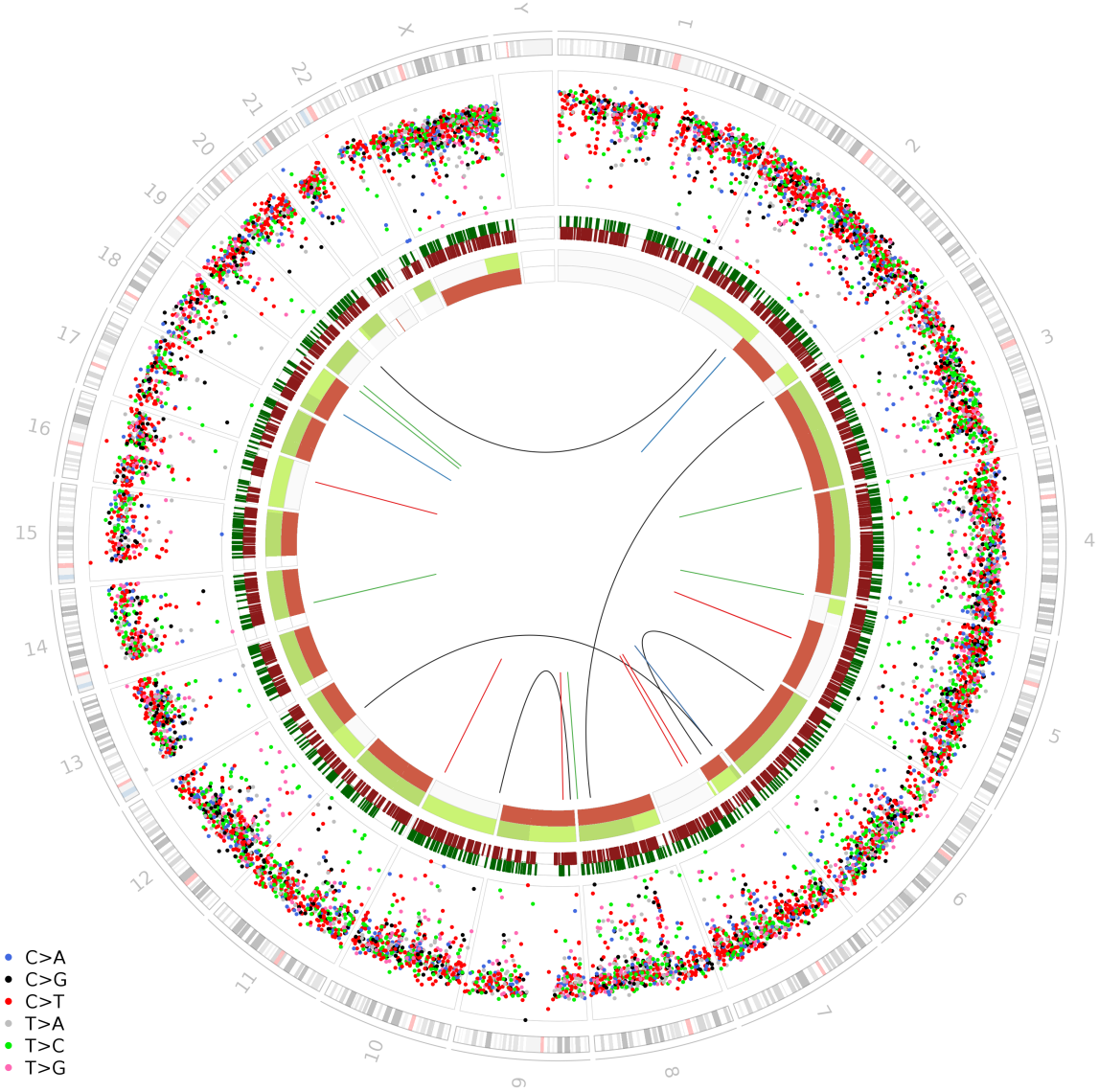
karyotype (purity 0.55, ploidy 2.2)



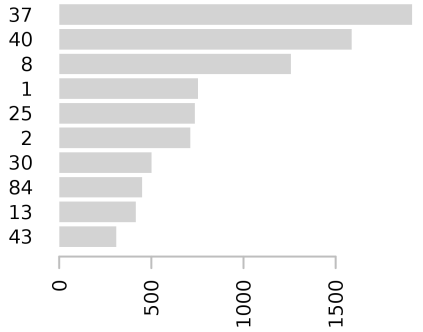
33 rearrangements



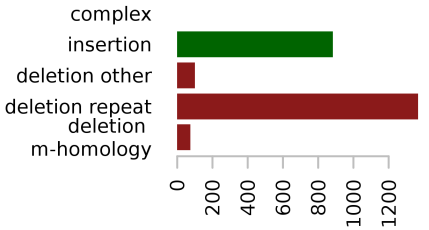
H203530 - FFPE



9753 substitutions of signatures:



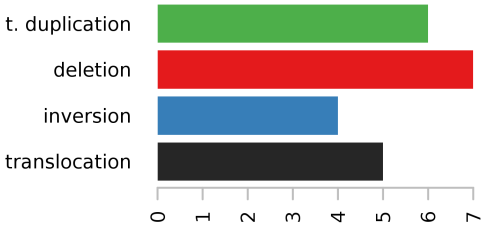
2428 indels



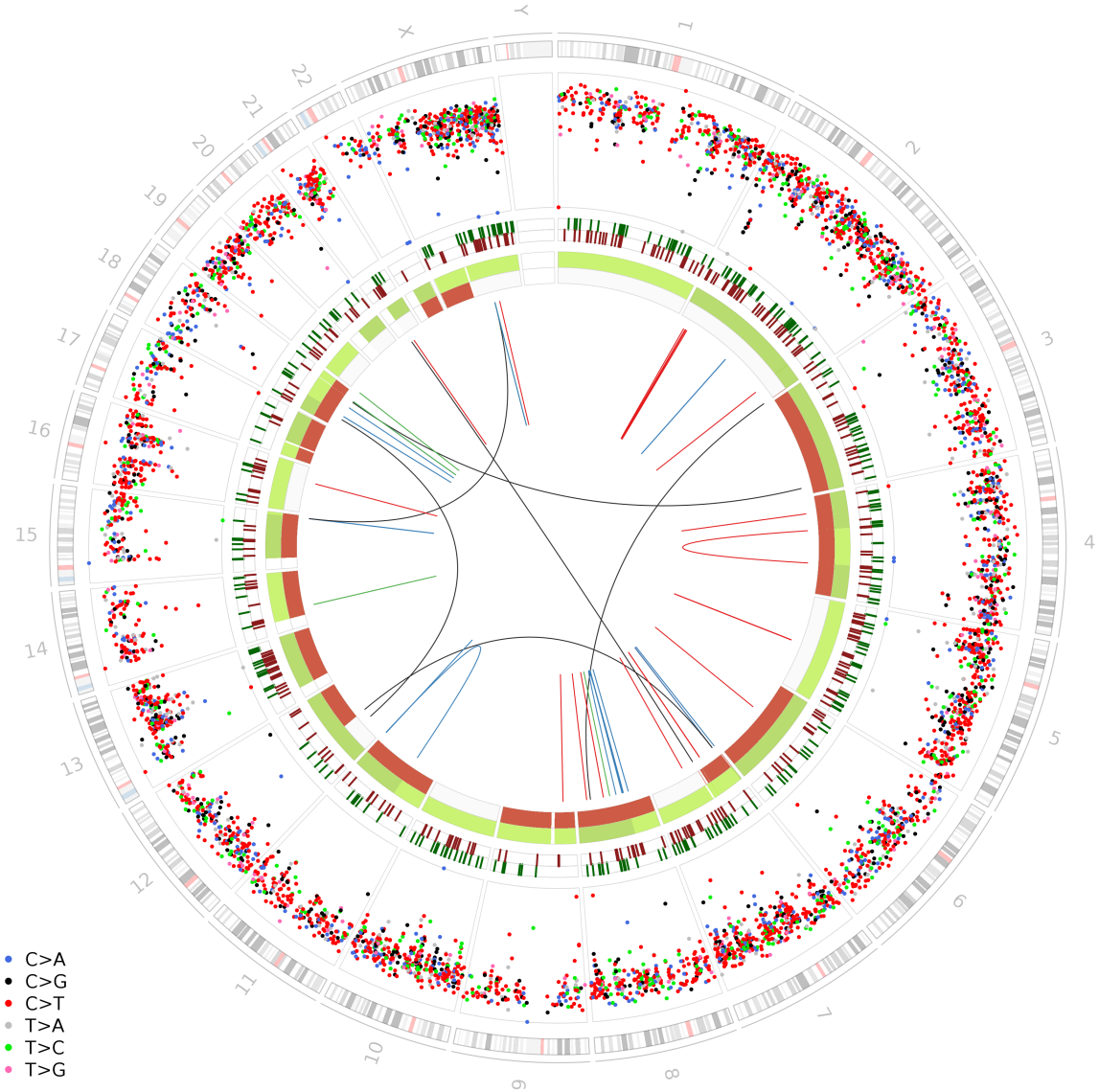
karyotype (purity 0.33, ploidy 5.1)



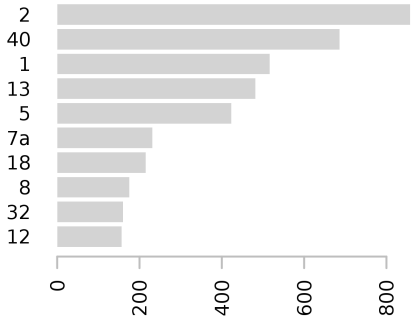
22 rearrangements



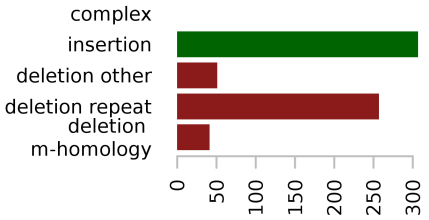
H203530 - Fresh Frozen



5504 substitutions of signatures:



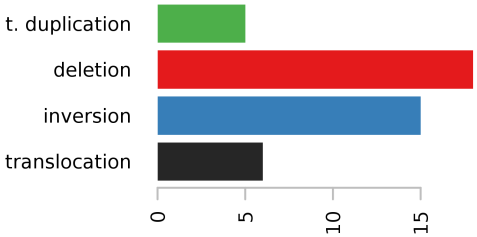
656 indels



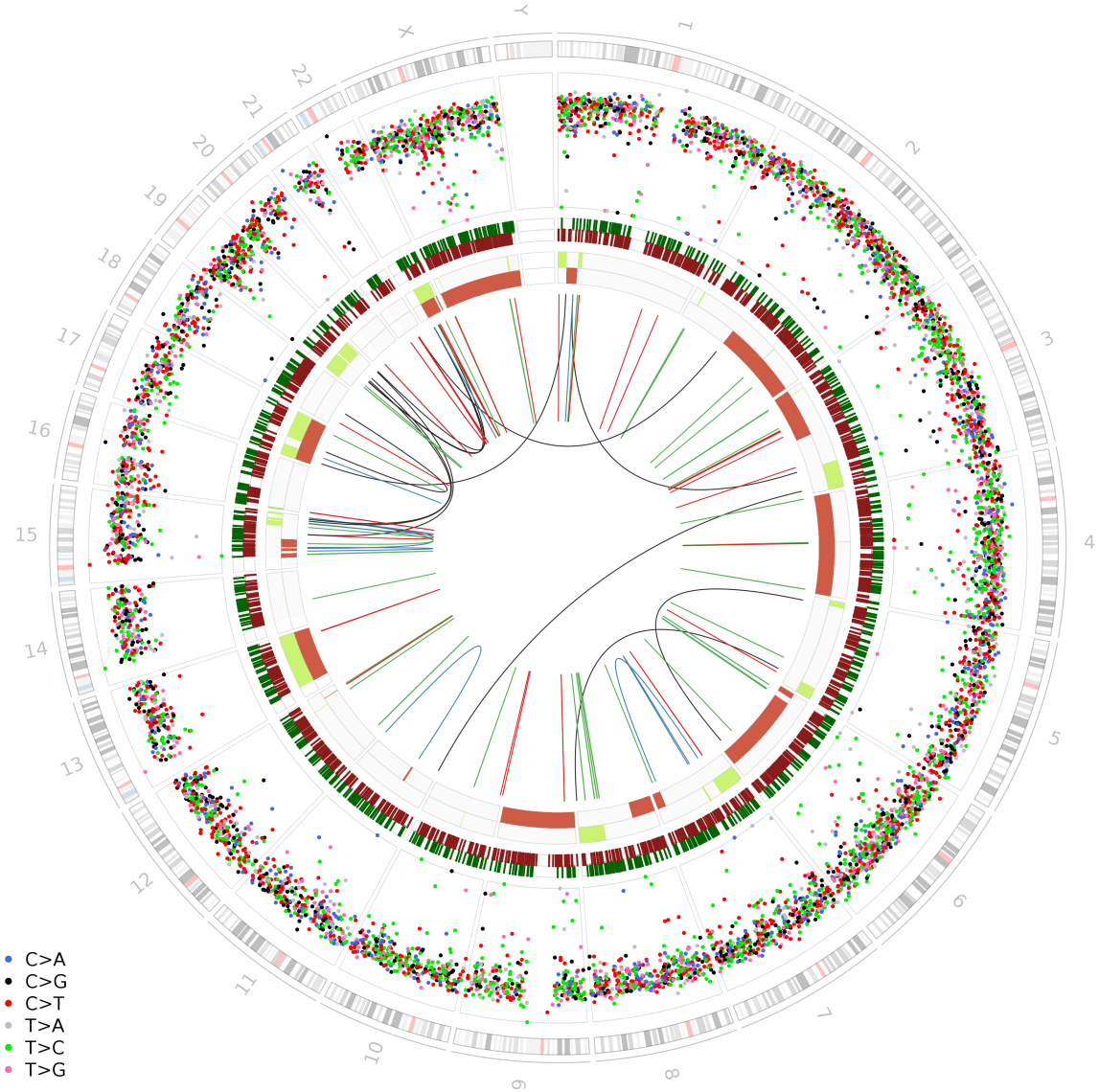
karyotype (purity 0.55, ploidy 4.3)

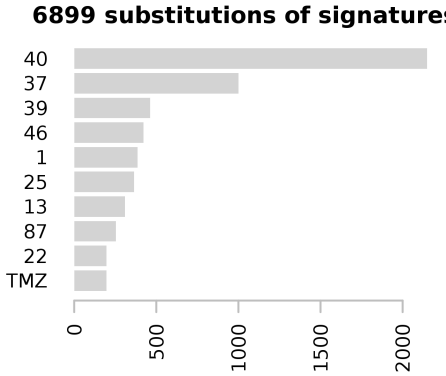


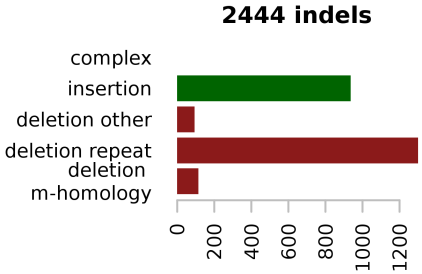
44 rearrangements



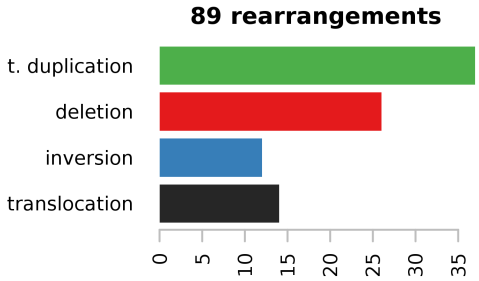
H203531 - FFPE



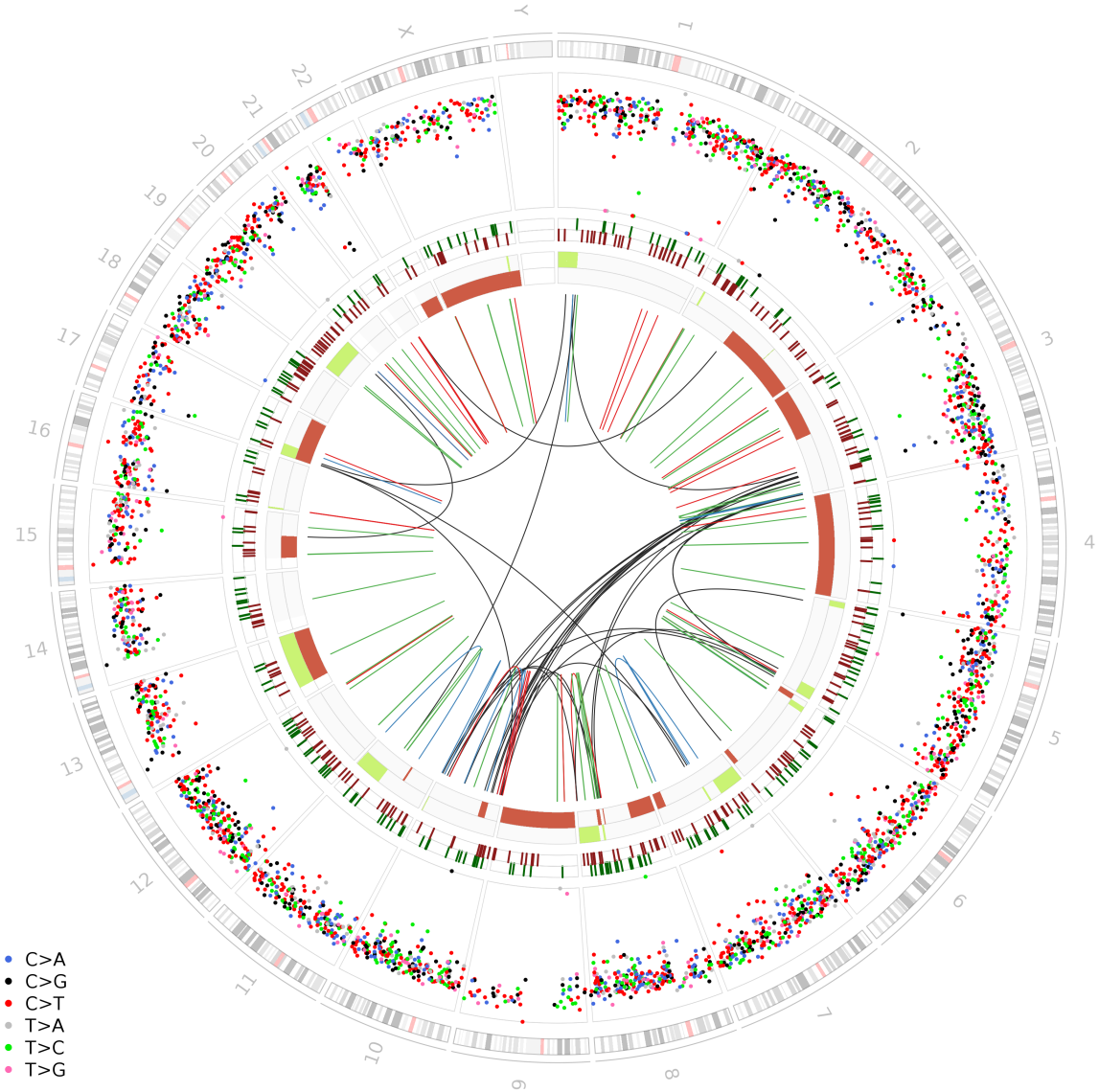




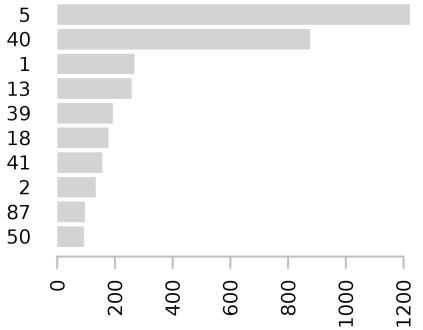




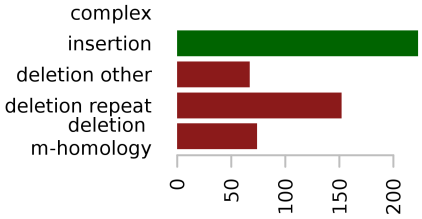
H203531 - Fresh Frozen



3619 substitutions of signatures:



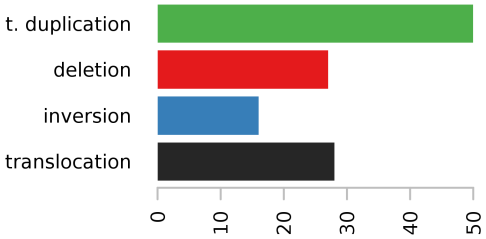
516 indels



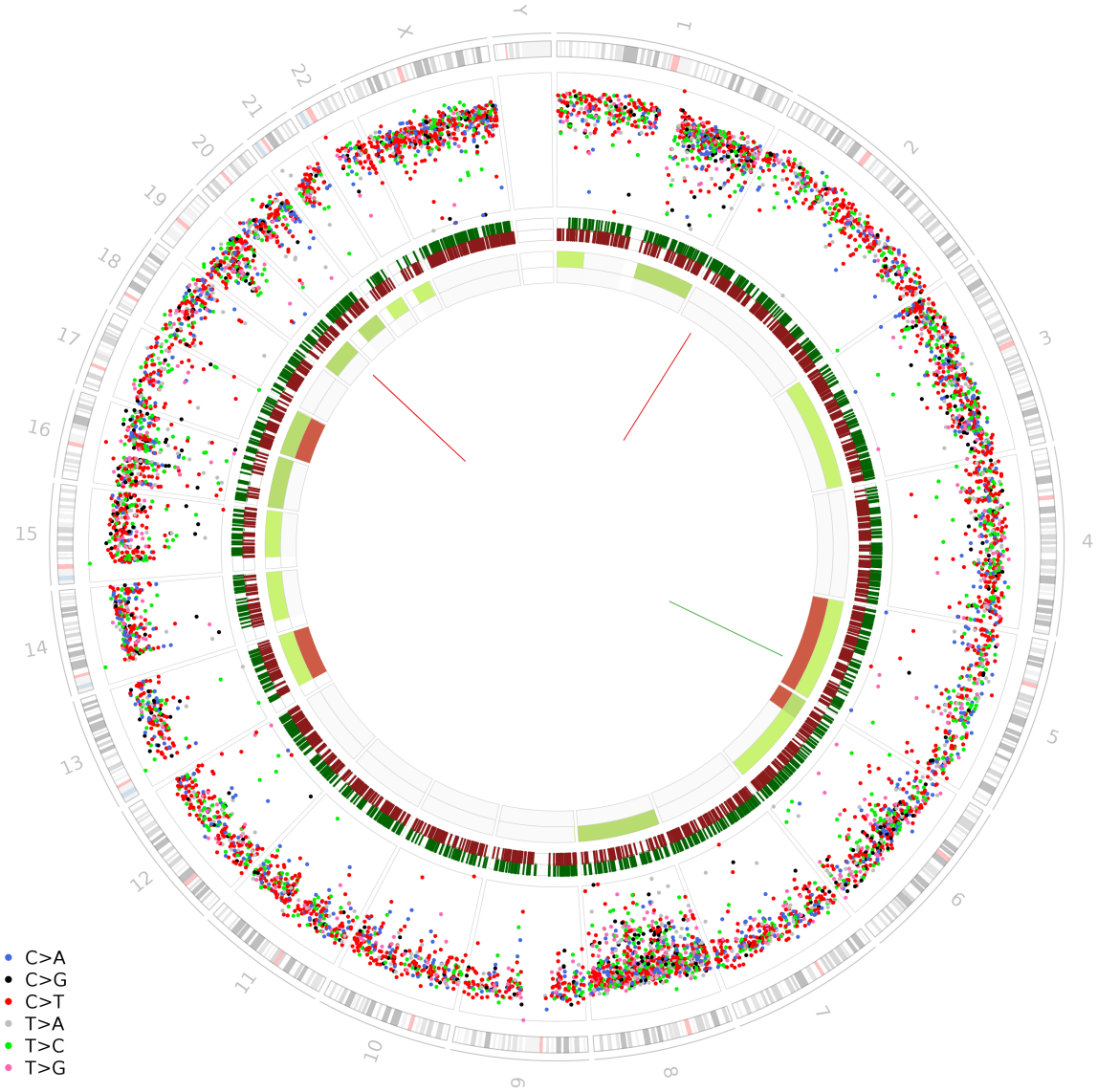
karyotype (purity 0.41, ploidy 1.8)



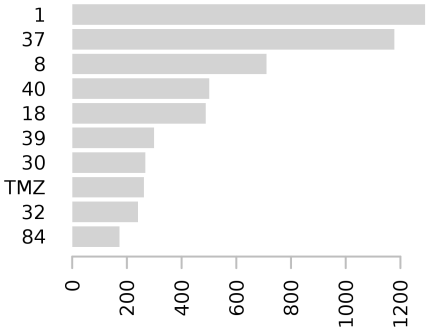
121 rearrangements



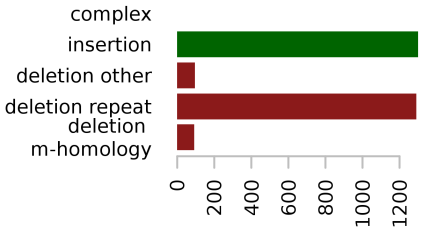
H203532 - FFPE



7240 substitutions of signatures:



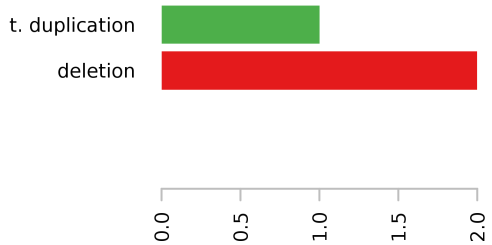
2778 indels



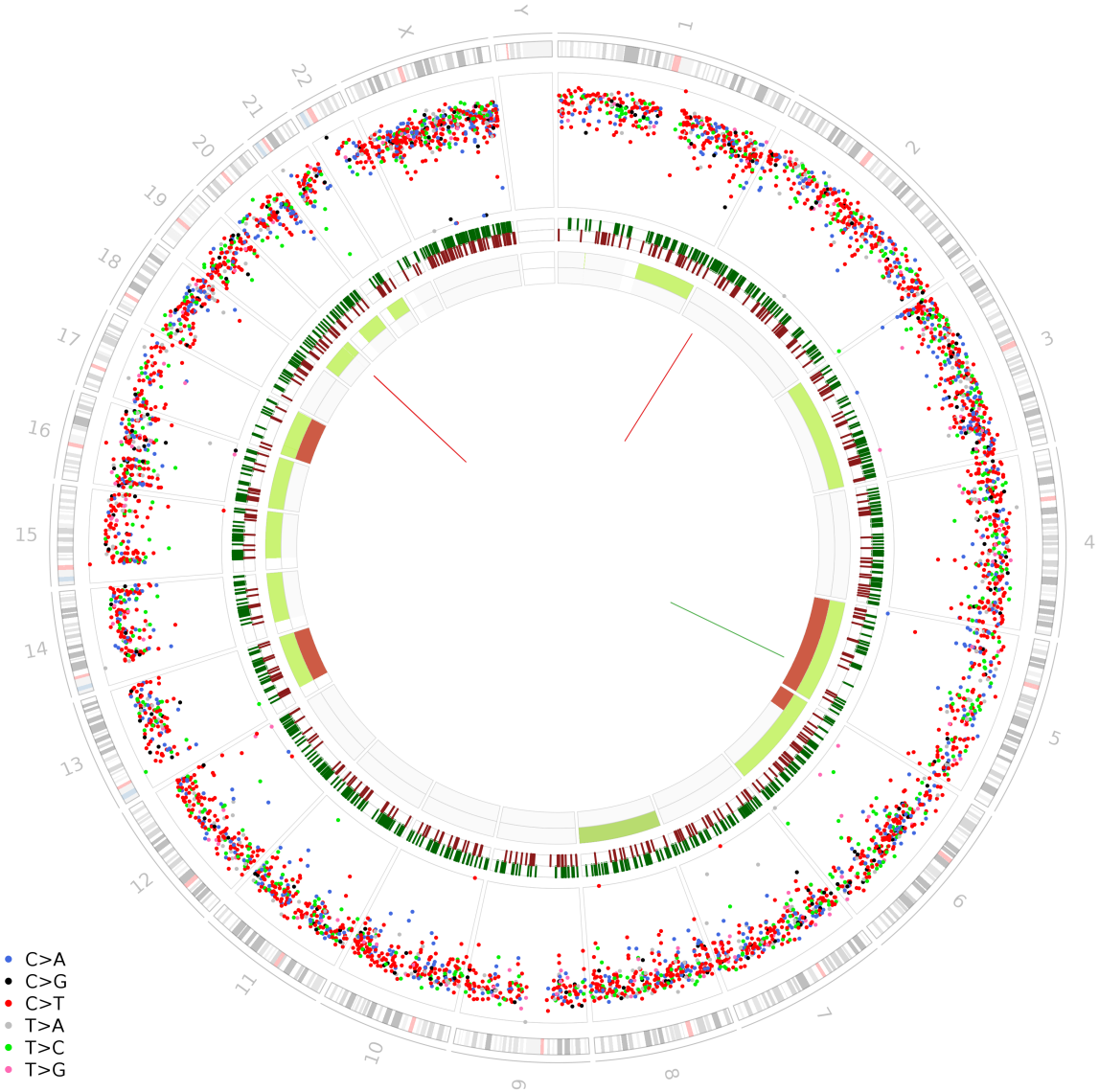
karyotype (purity 0.91, ploidy 4.4)



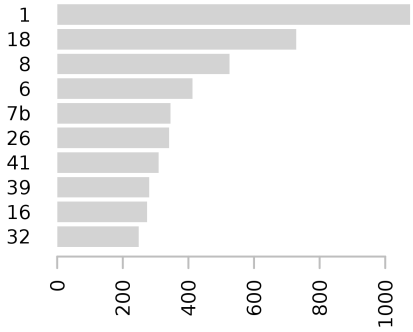
3 rearrangements



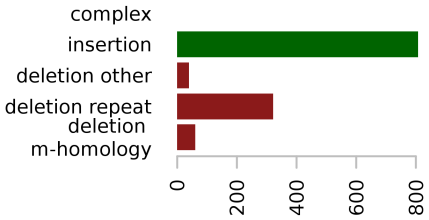
H203532 - Fresh Frozen



4784 substitutions of signatures:



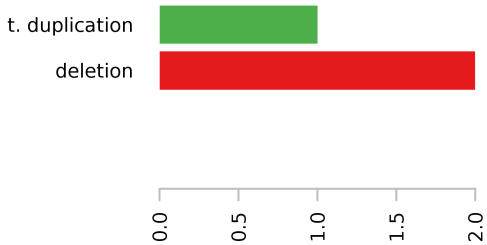
1229 indels



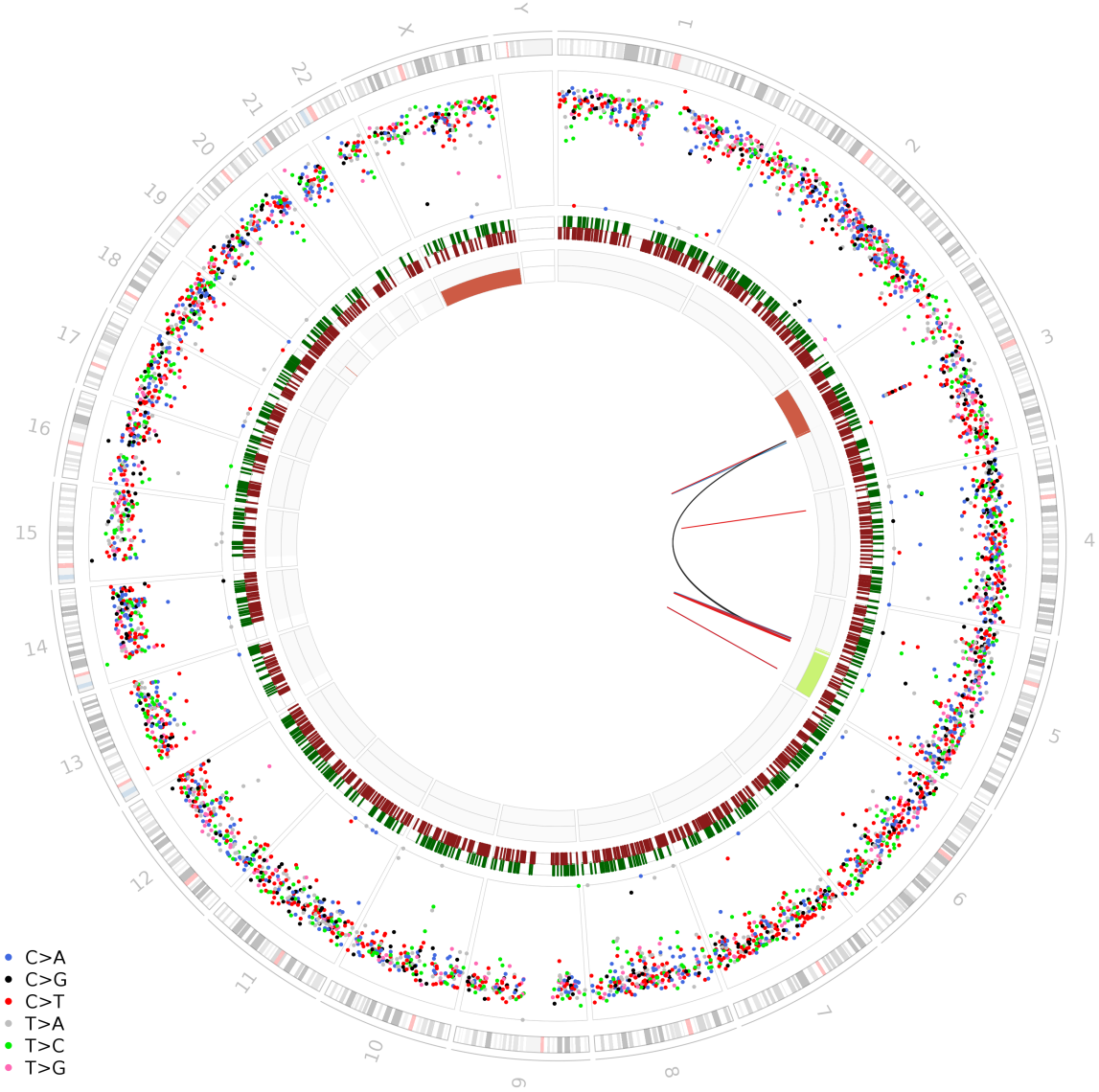
karyotype (purity 0.91, ploidy 2.6)



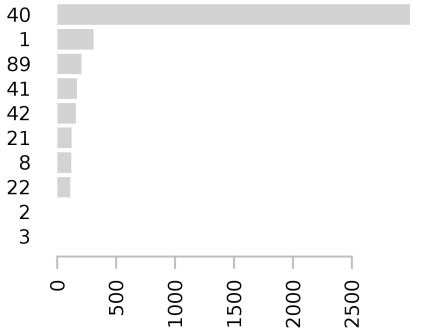
3 rearrangements



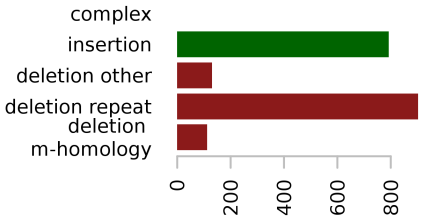
H203533 - FFPE



4179 substitutions of signatures:



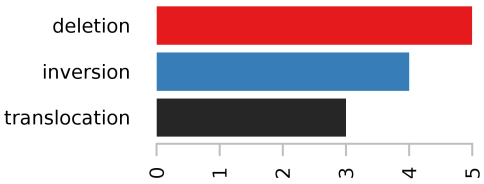
1938 indels



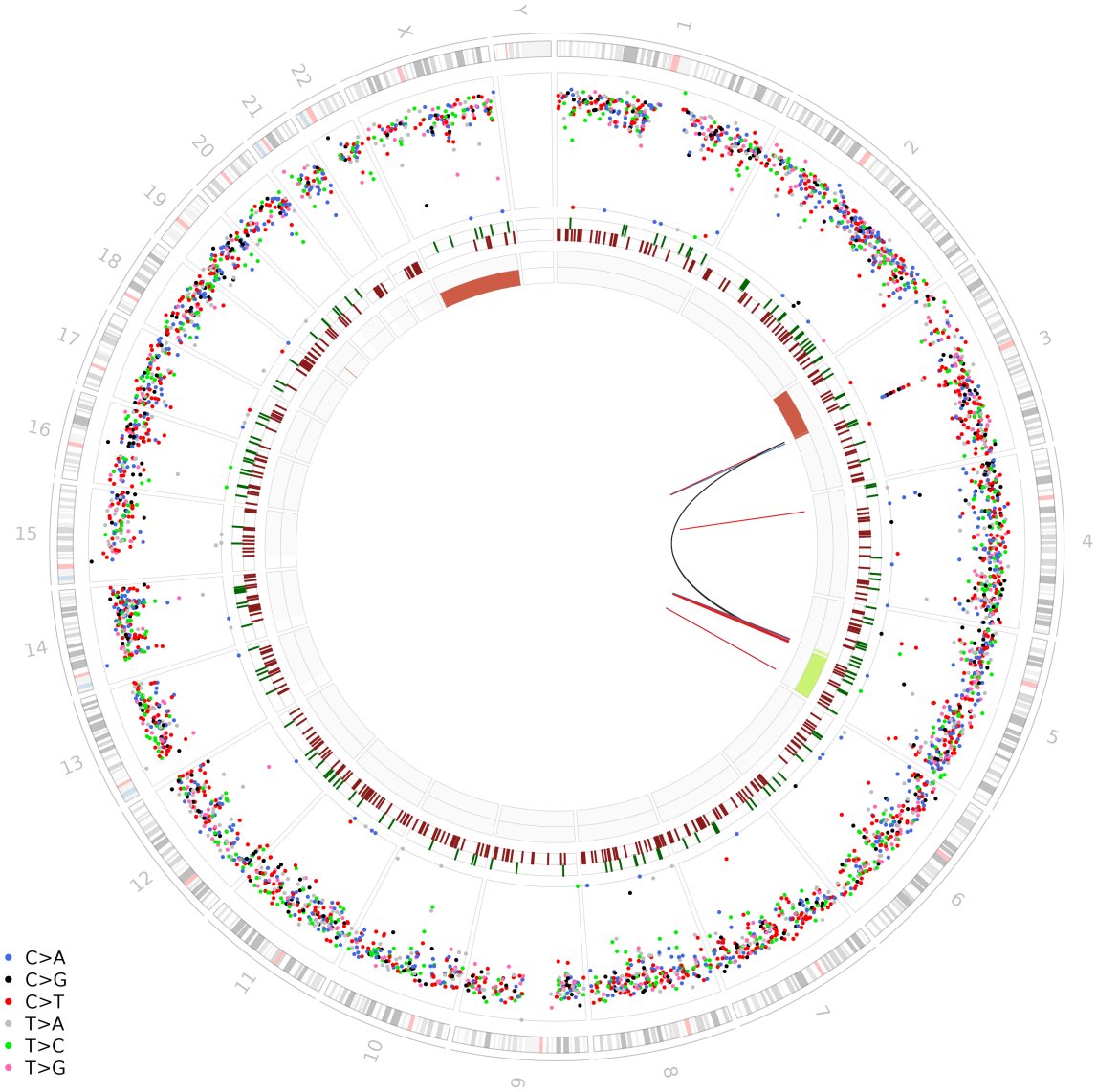
karyotype (purity 0.63, ploidy 2.1)



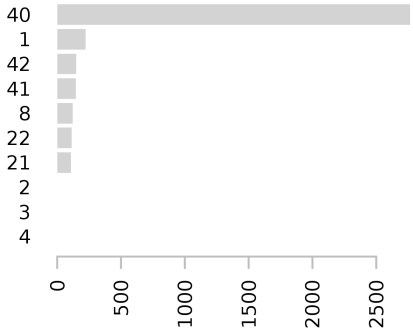
12 rearrangements



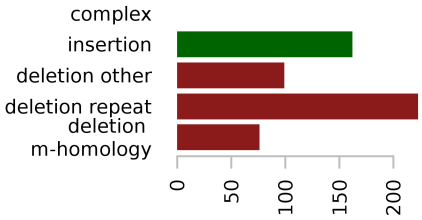
H203533 - Fresh Frozen



3647 substitutions of signatures:



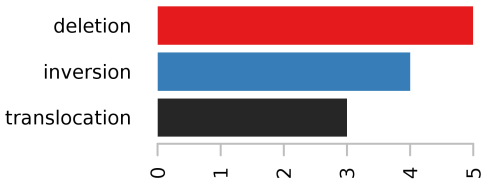
560 indels



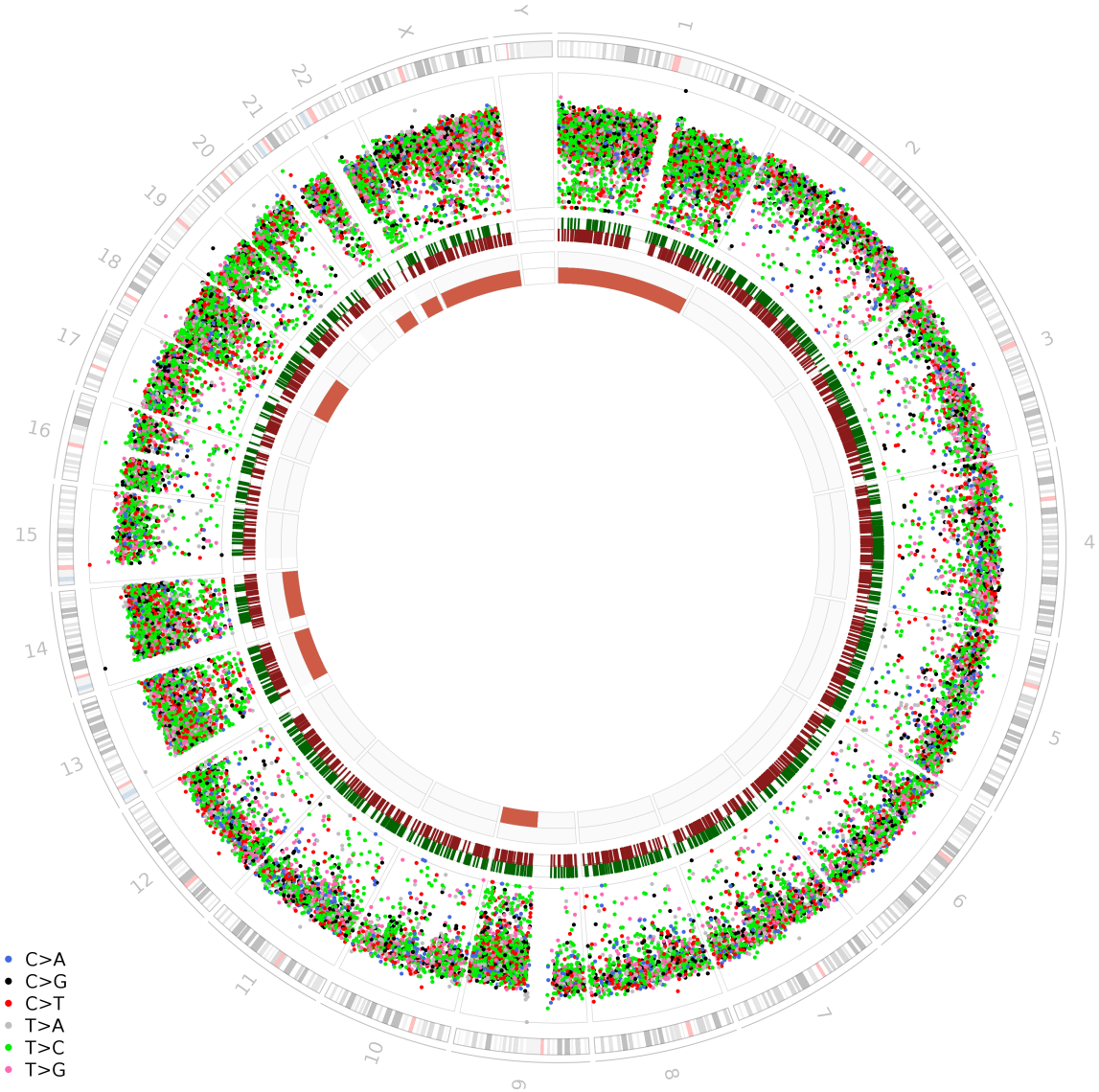
karyotype (purity 0.55, ploidy 2.1)

LOH gain

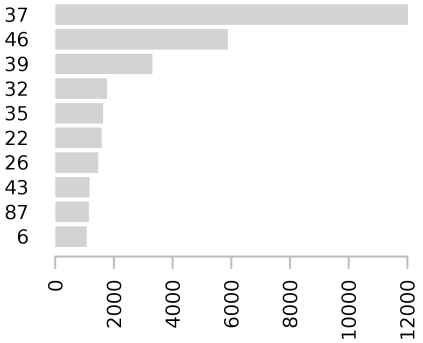
12 rearrangements



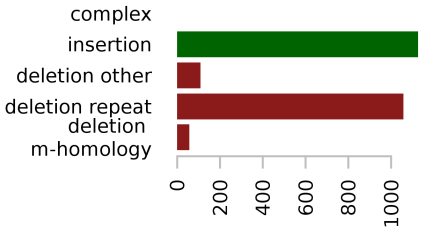
H203534 - FFPE



35288 substitutions of signatures:



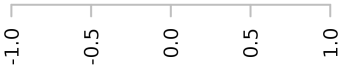
2349 indels



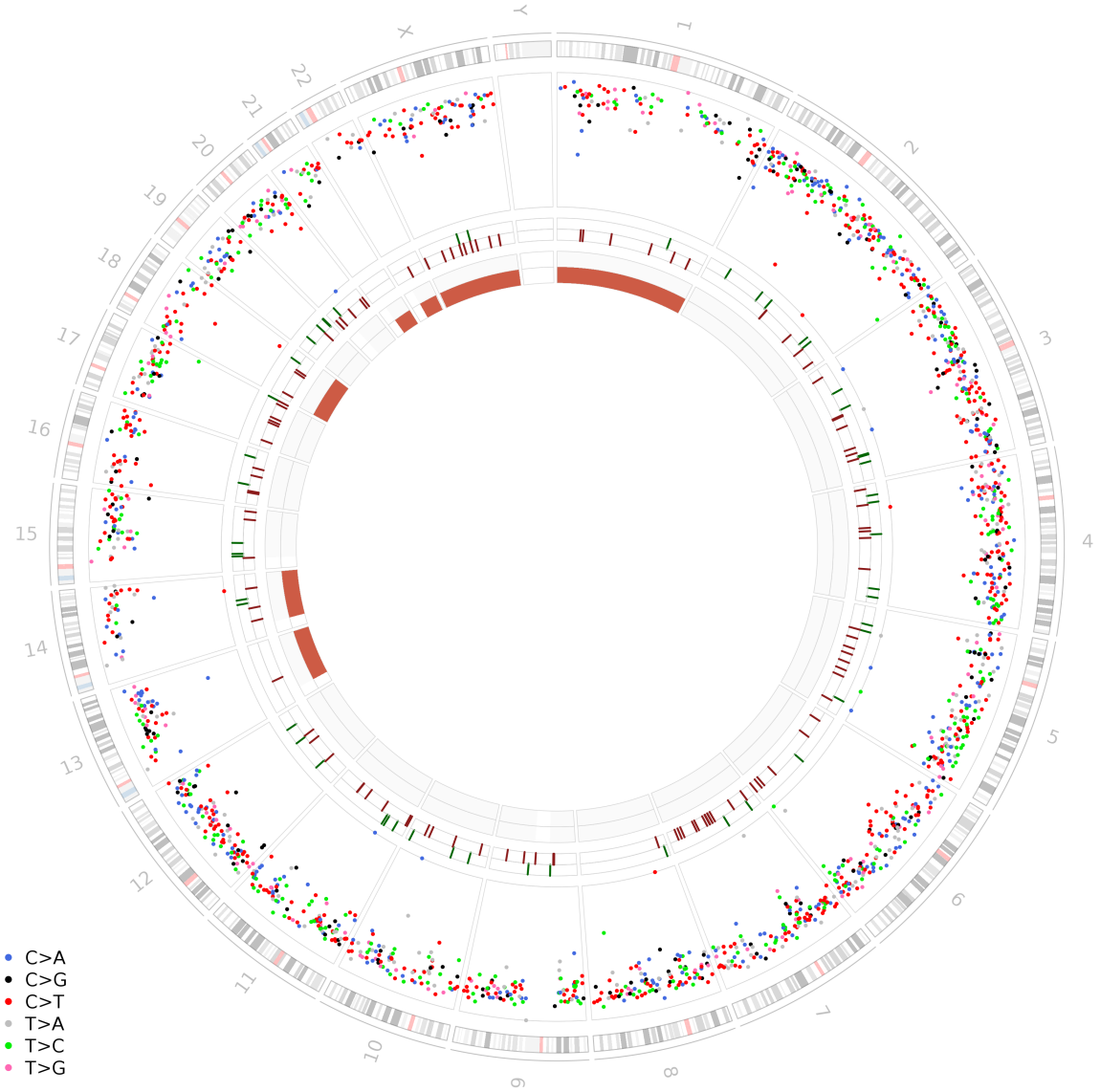
karyotype (purity 0.68, ploidy 2.2)

LOH gain

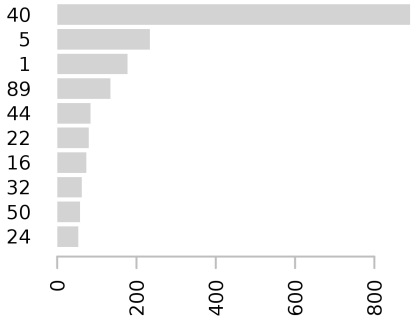
0 rearrangements



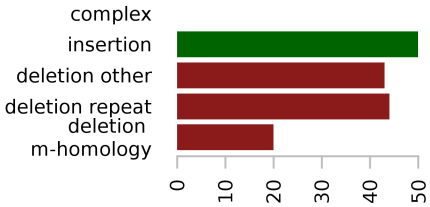
H203534 - Fresh Frozen



1887 substitutions of signatures:



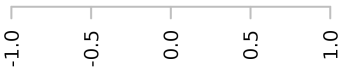
157 indels



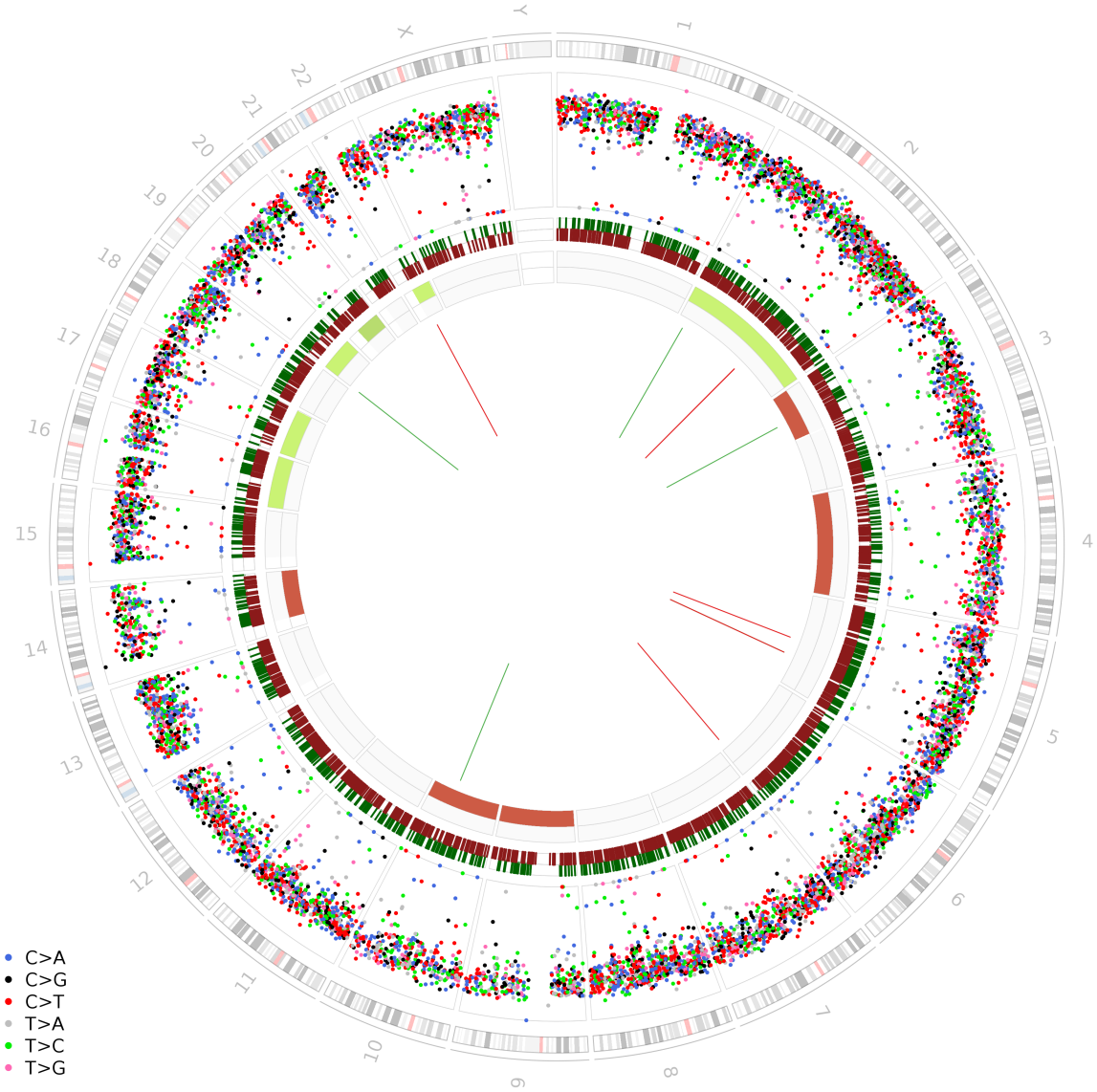
karyotype (purity 0.81, ploidy 1.8)

LOH gain

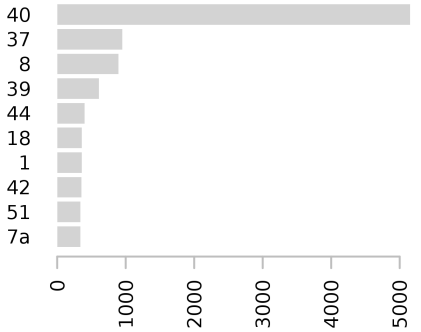
0 rearrangements



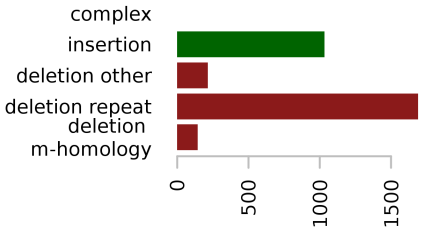
H203535 - FFPE



10971 substitutions of signatures:



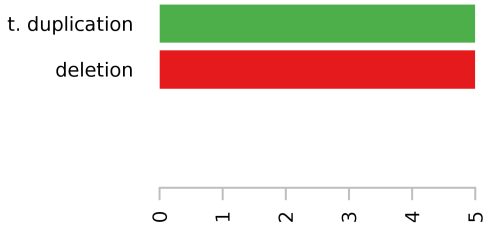
3081 indels



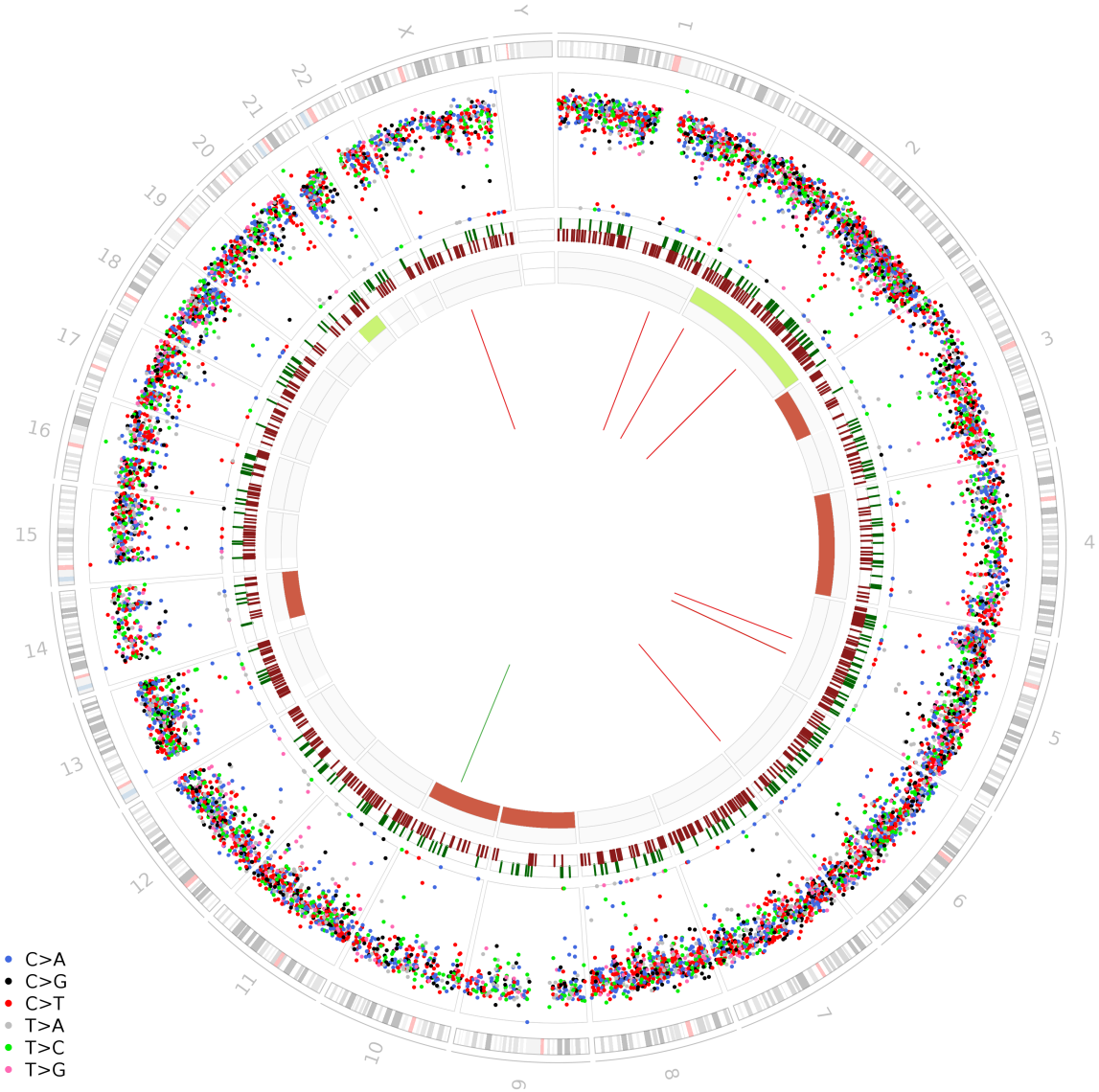
karyotype (purity 0.61, ploidy 3.8)



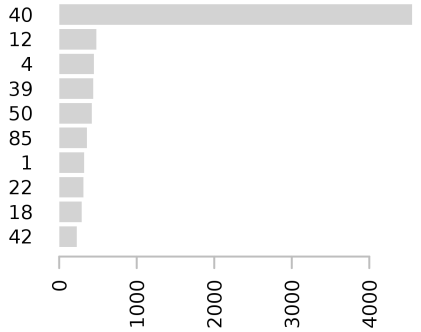
10 rearrangements



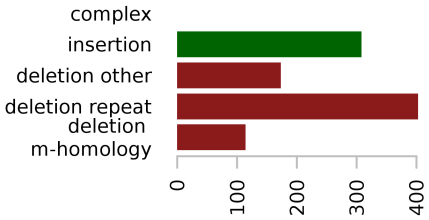
H203535 - Fresh Frozen



9740 substitutions of signatures:



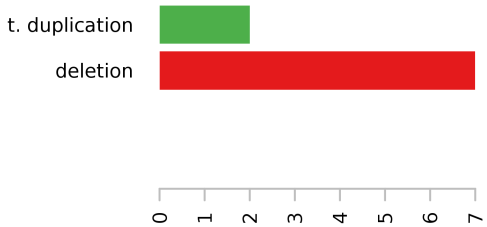
998 indels



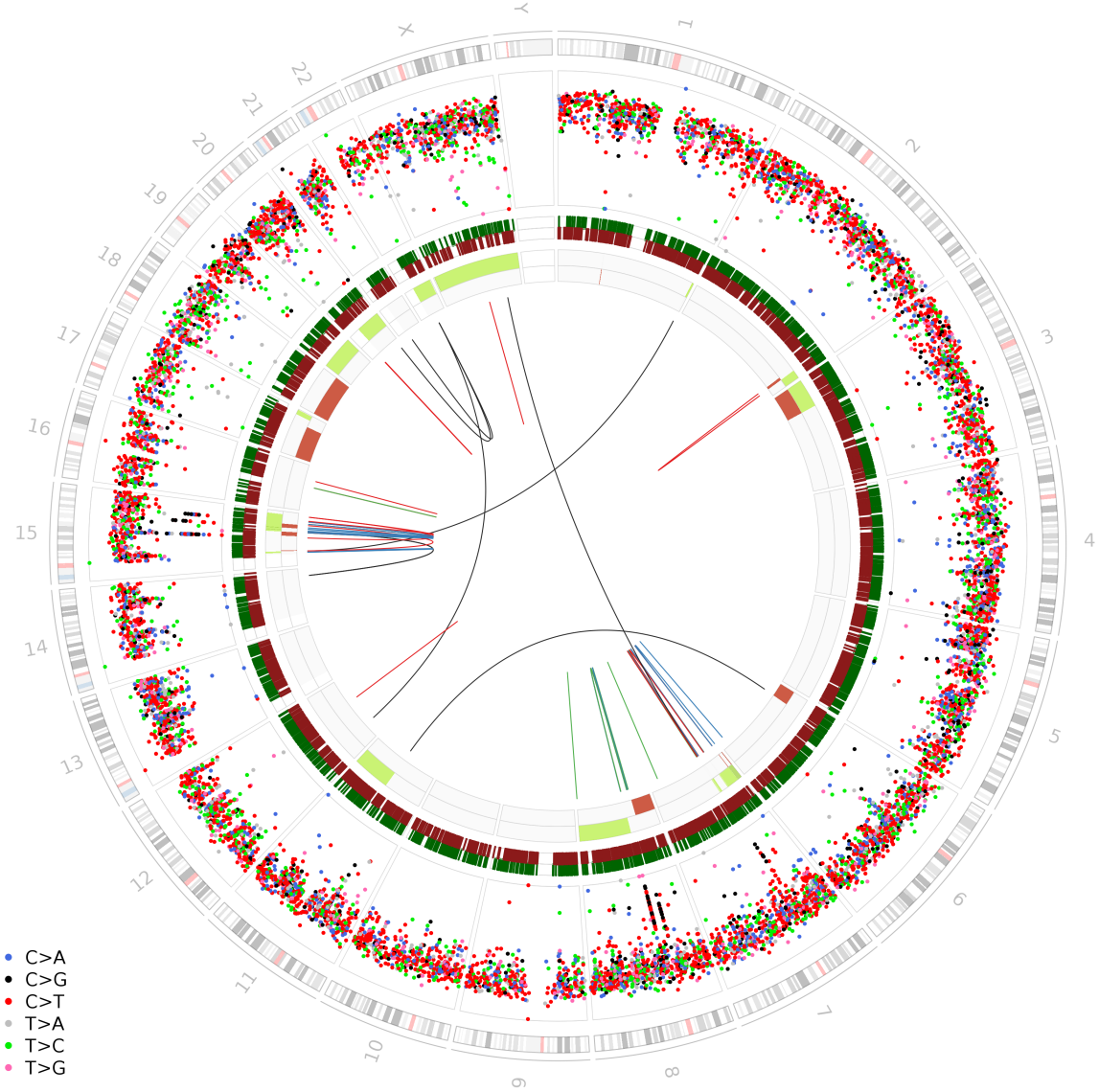
karyotype (purity 0.54, ploidy 2)



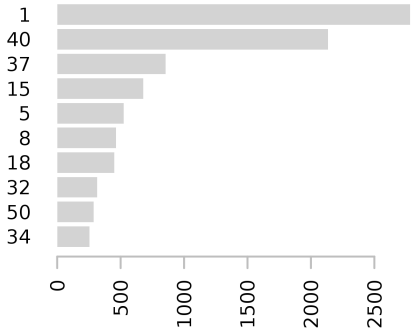
9 rearrangements



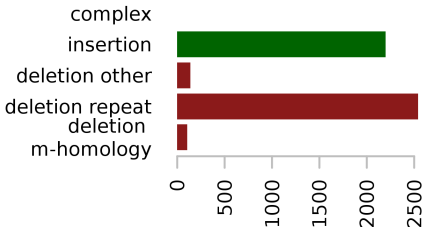
H203537 - FFPE



11727 substitutions of signatures:



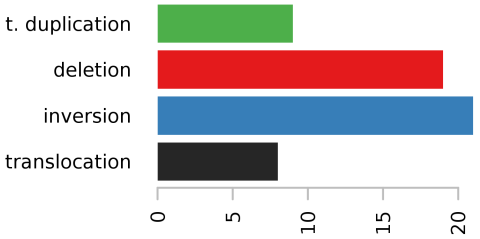
4982 indels



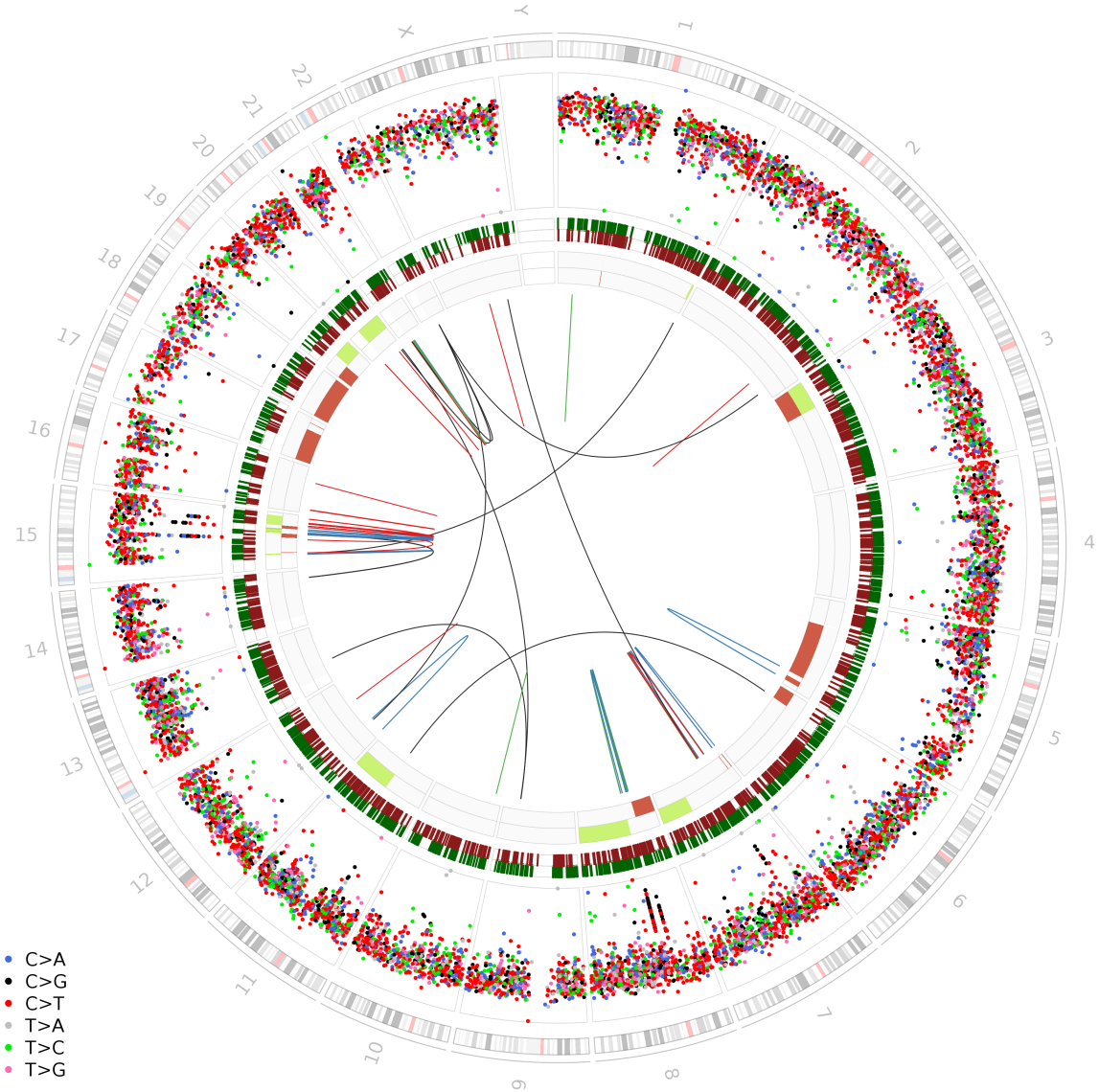
karyotype (purity 0.87, ploidy 3)



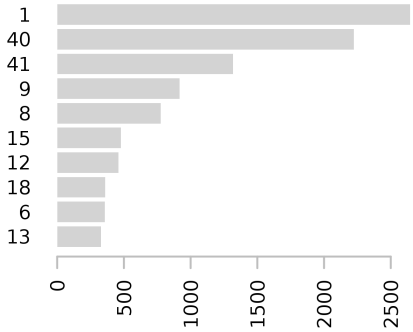
57 rearrangements



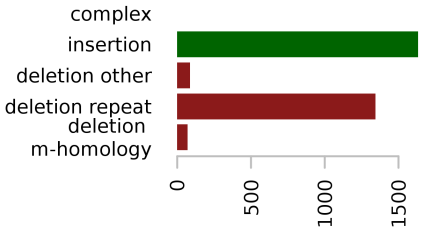
H203537 - Fresh Frozen



12995 substitutions of signatures:



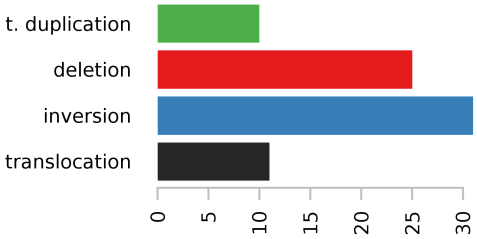
3135 indels



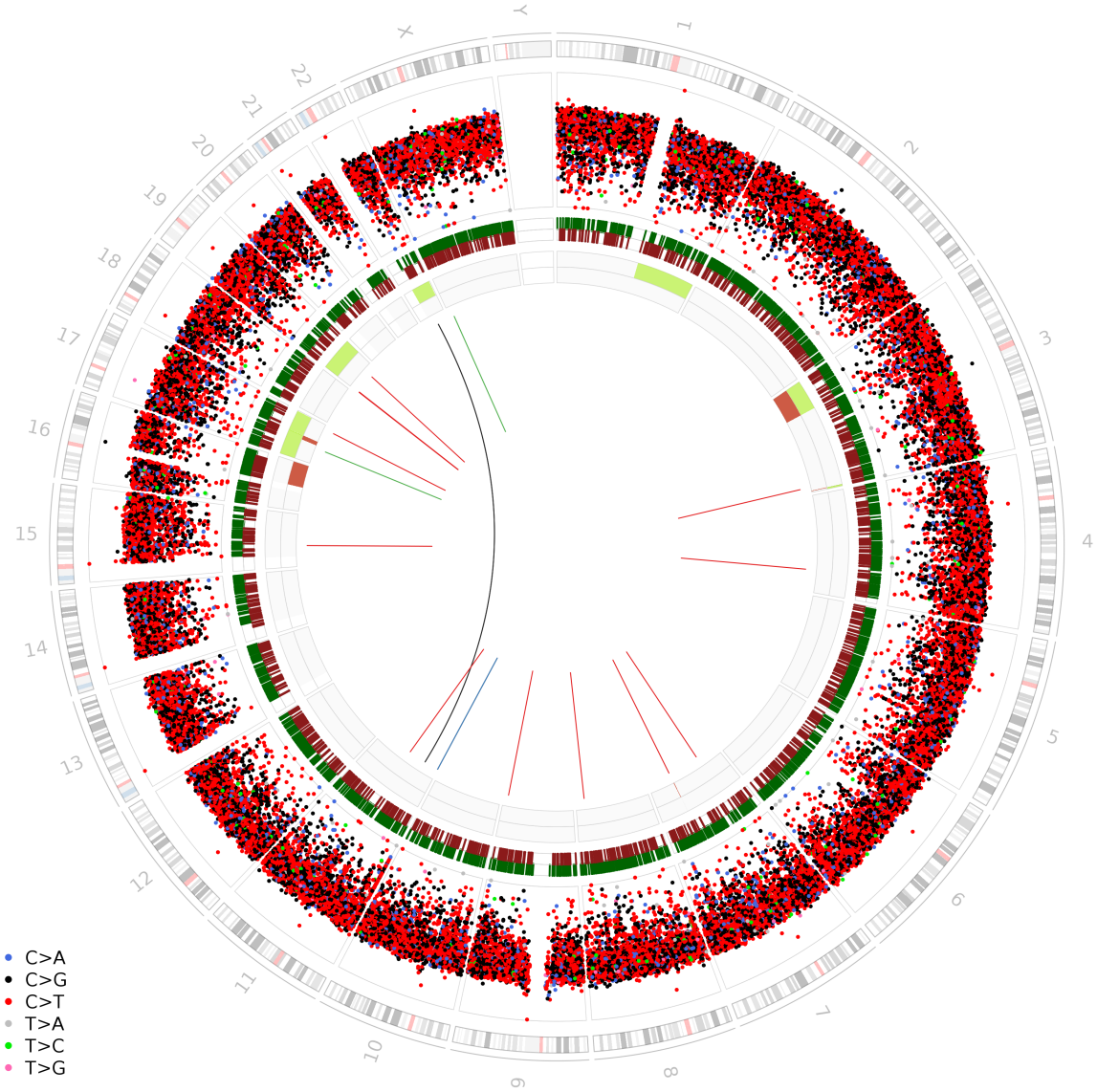
karyotype (purity 0.74, ploidy 2)



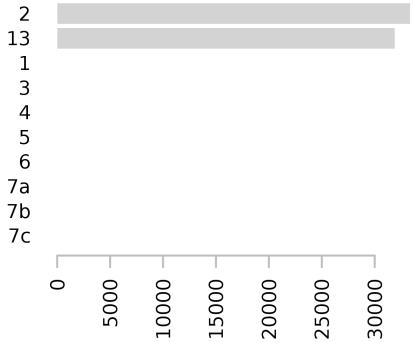
77 rearrangements



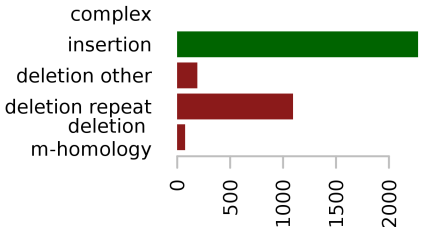
H203539 - FFPE



68277 substitutions of signatures:



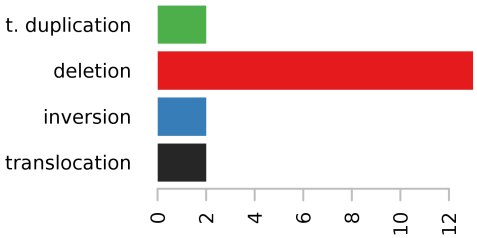
3635 indels



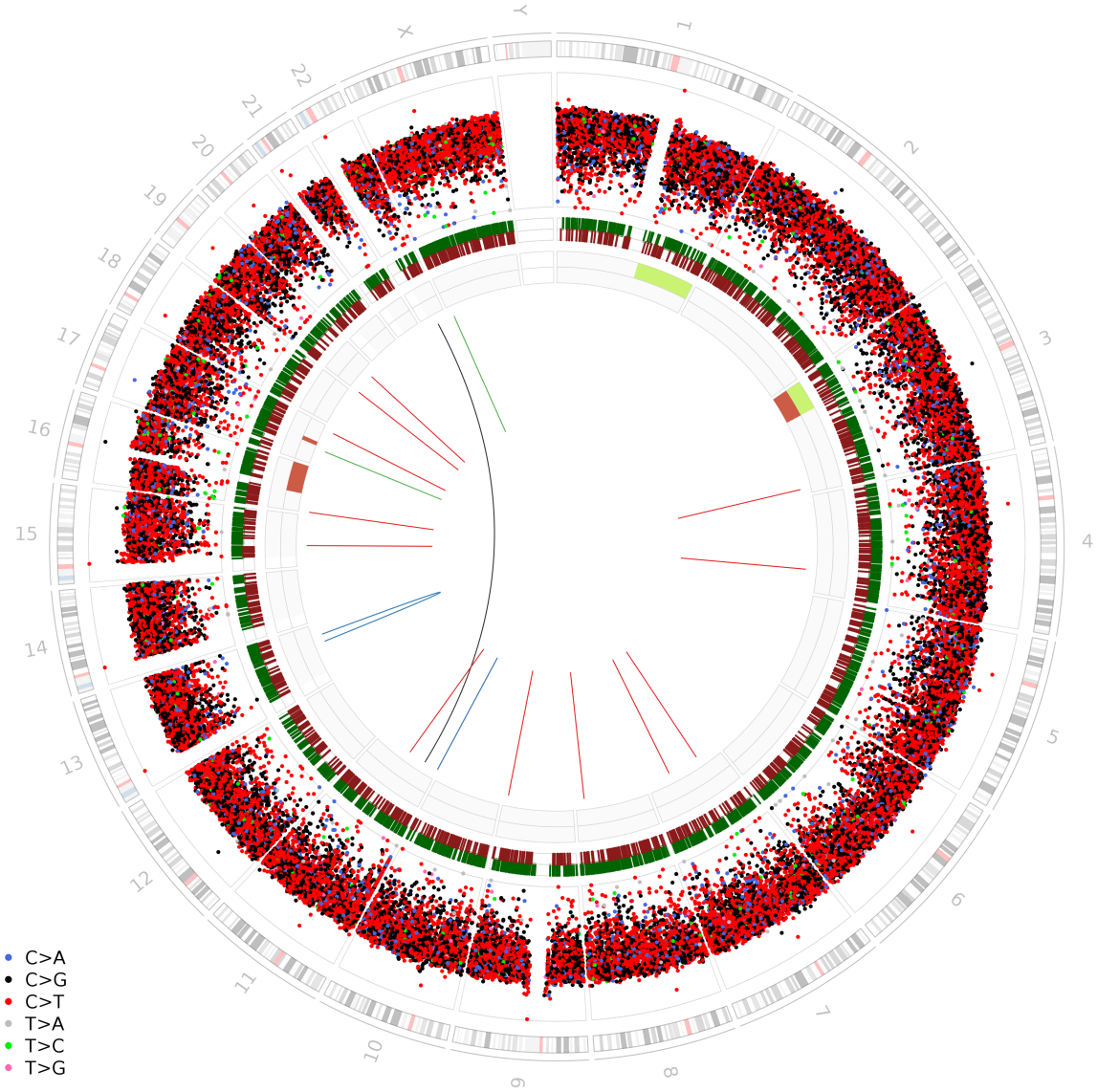
karyotype (purity 0.53, ploidy 2.6)



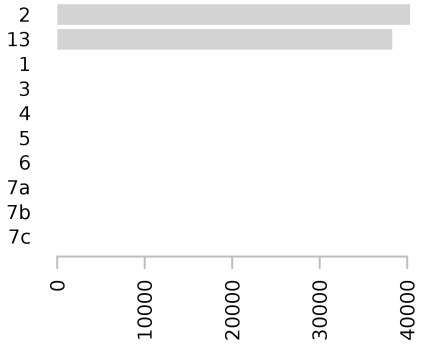
19 rearrangements



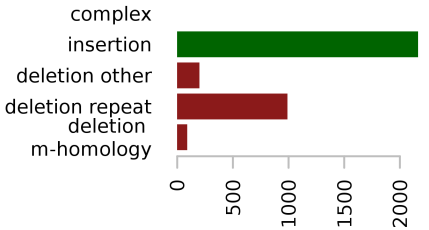
H203539 - Fresh Frozen



82471 substitutions of signatures:



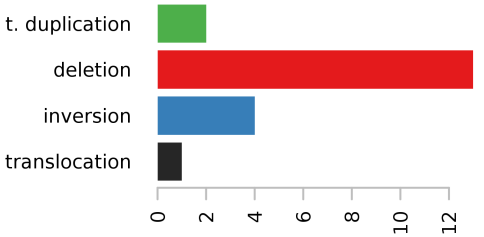
3443 indels



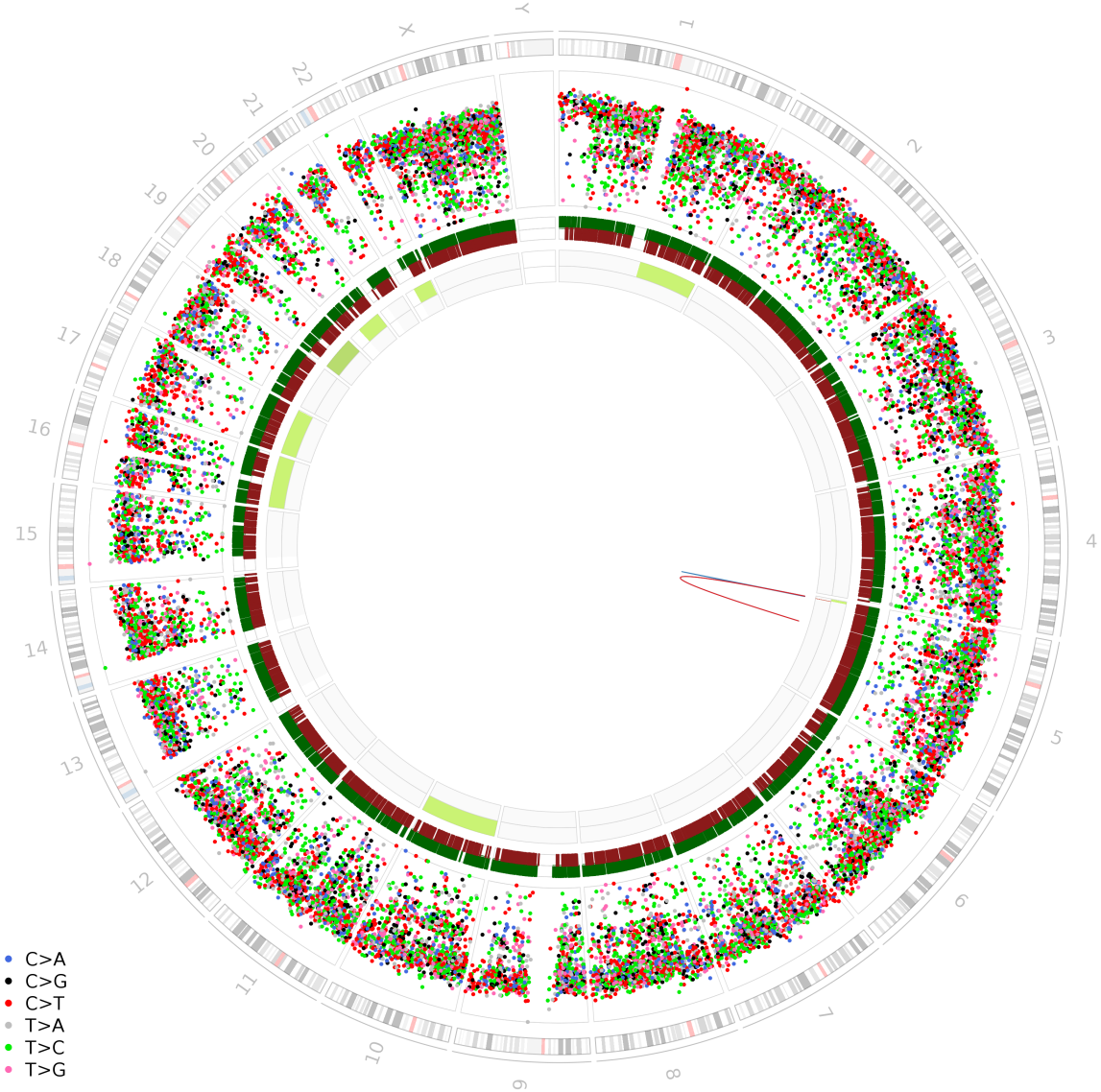
karyotype (purity 0.48, ploidy 2.1)



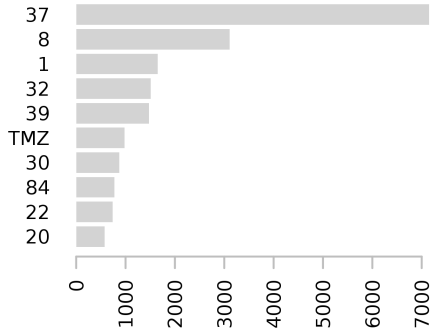
20 rearrangements



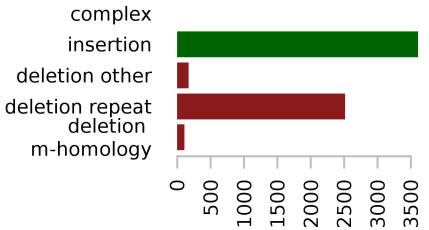
H203540 - FFPE



23428 substitutions of signatures:



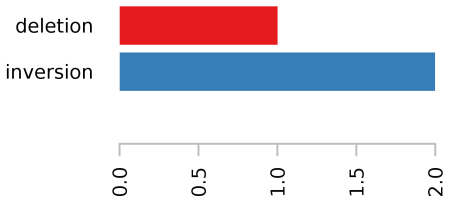
6404 indels



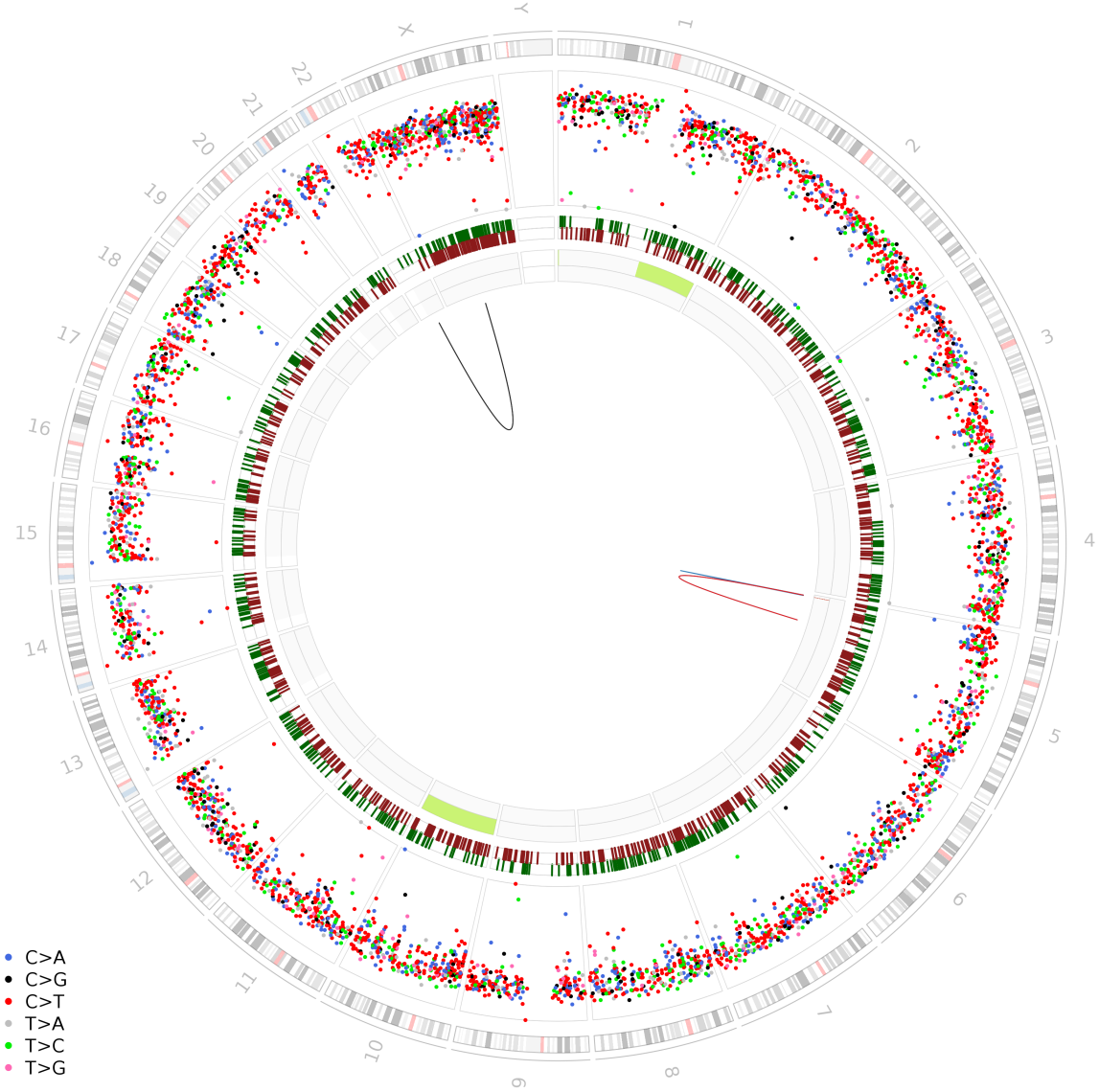
karyotype (purity 0.49, ploidy 4.6)



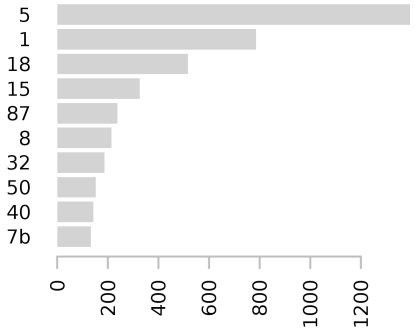
3 rearrangements



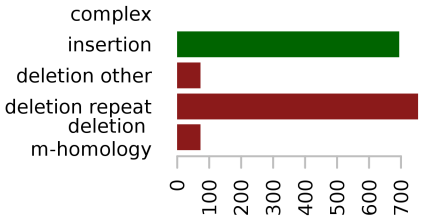
H203540 - Fresh Frozen



5002 substitutions of signatures:



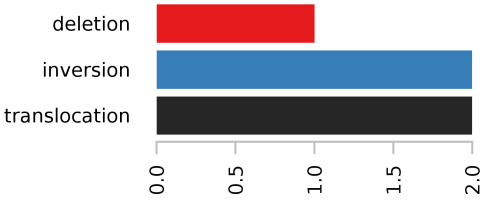
1594 indels



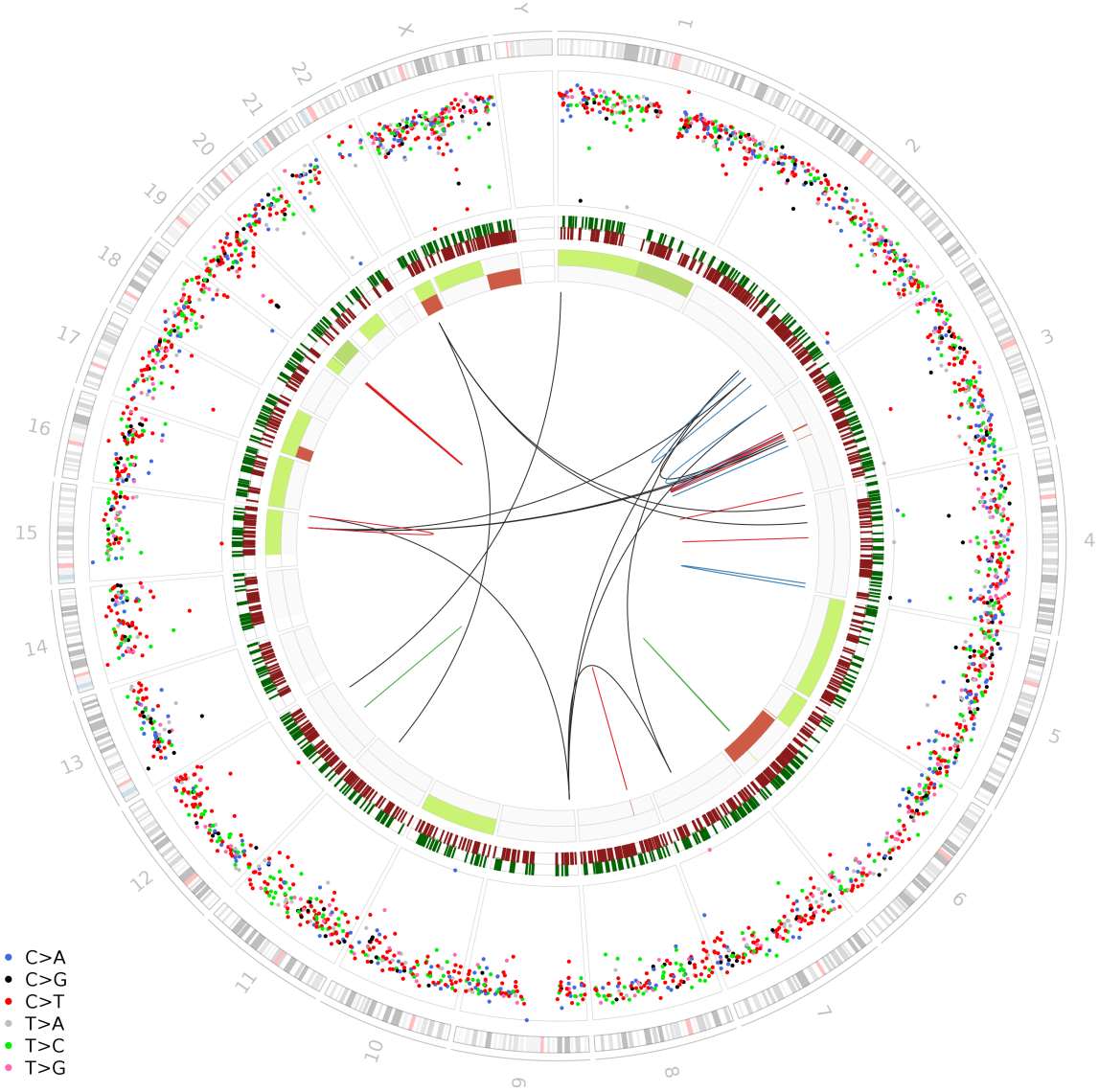
karyotype (purity 0.57, ploidy 2.4)



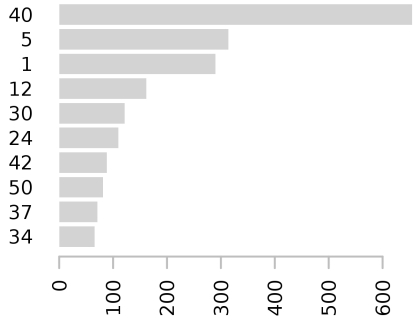
5 rearrangements



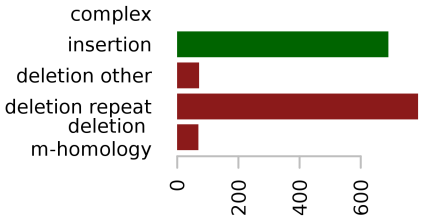
H203541 - FFPE



2506 substitutions of signatures:



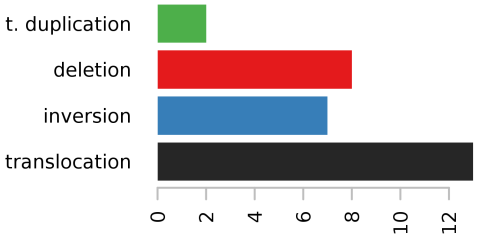
1619 indels



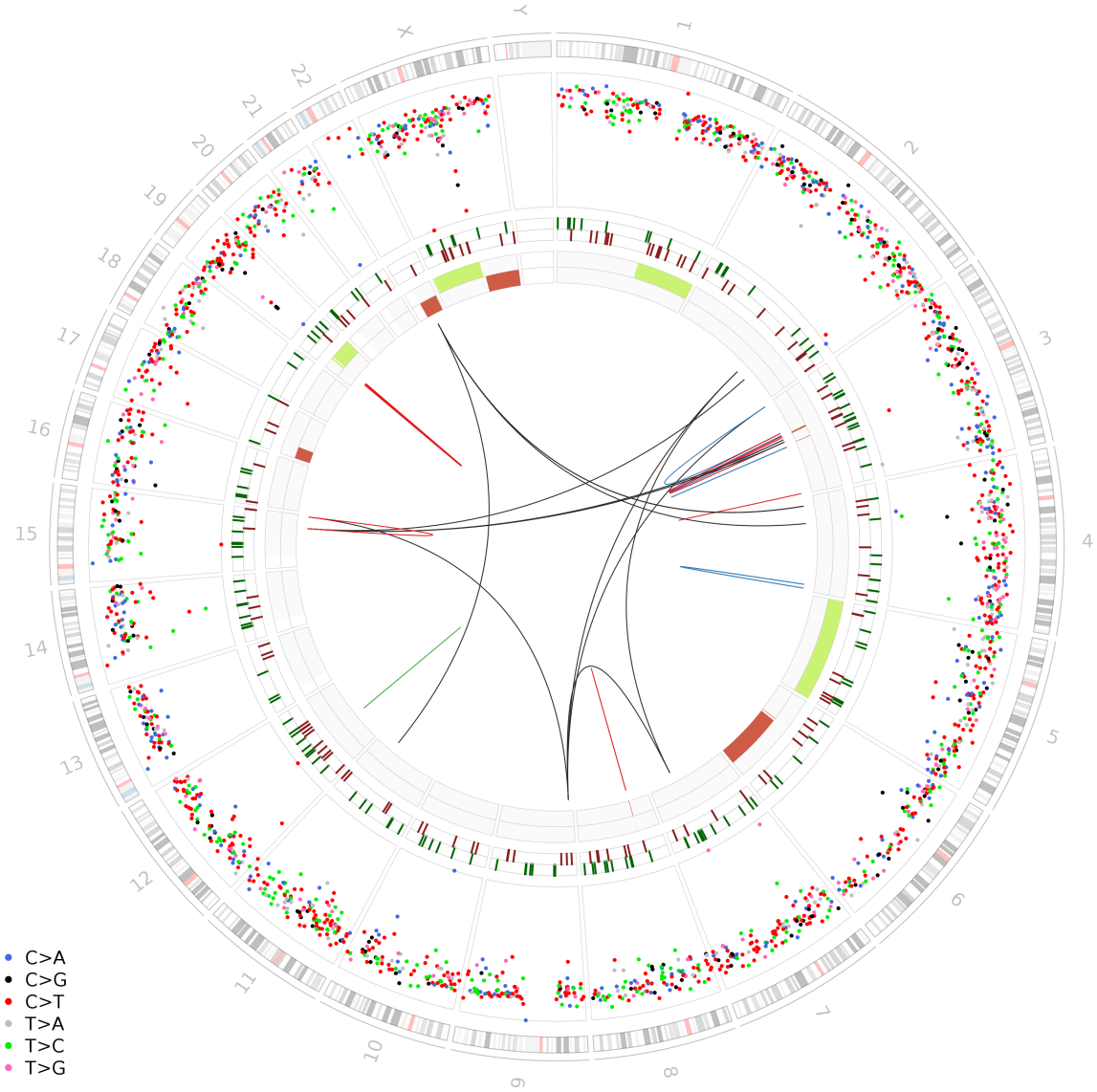
karyotype (purity 0.79, ploidy 3.8)



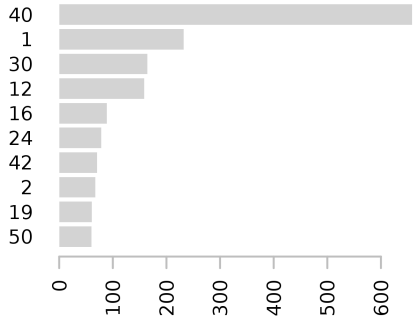
30 rearrangements



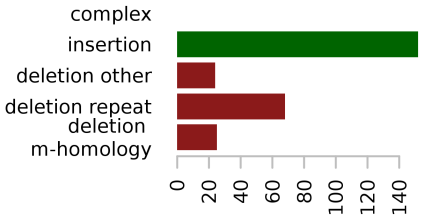
H203541 - Fresh Frozen



2059 substitutions of signatures:



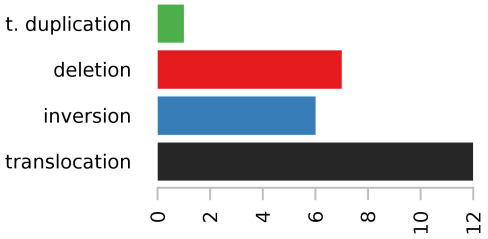
269 indels



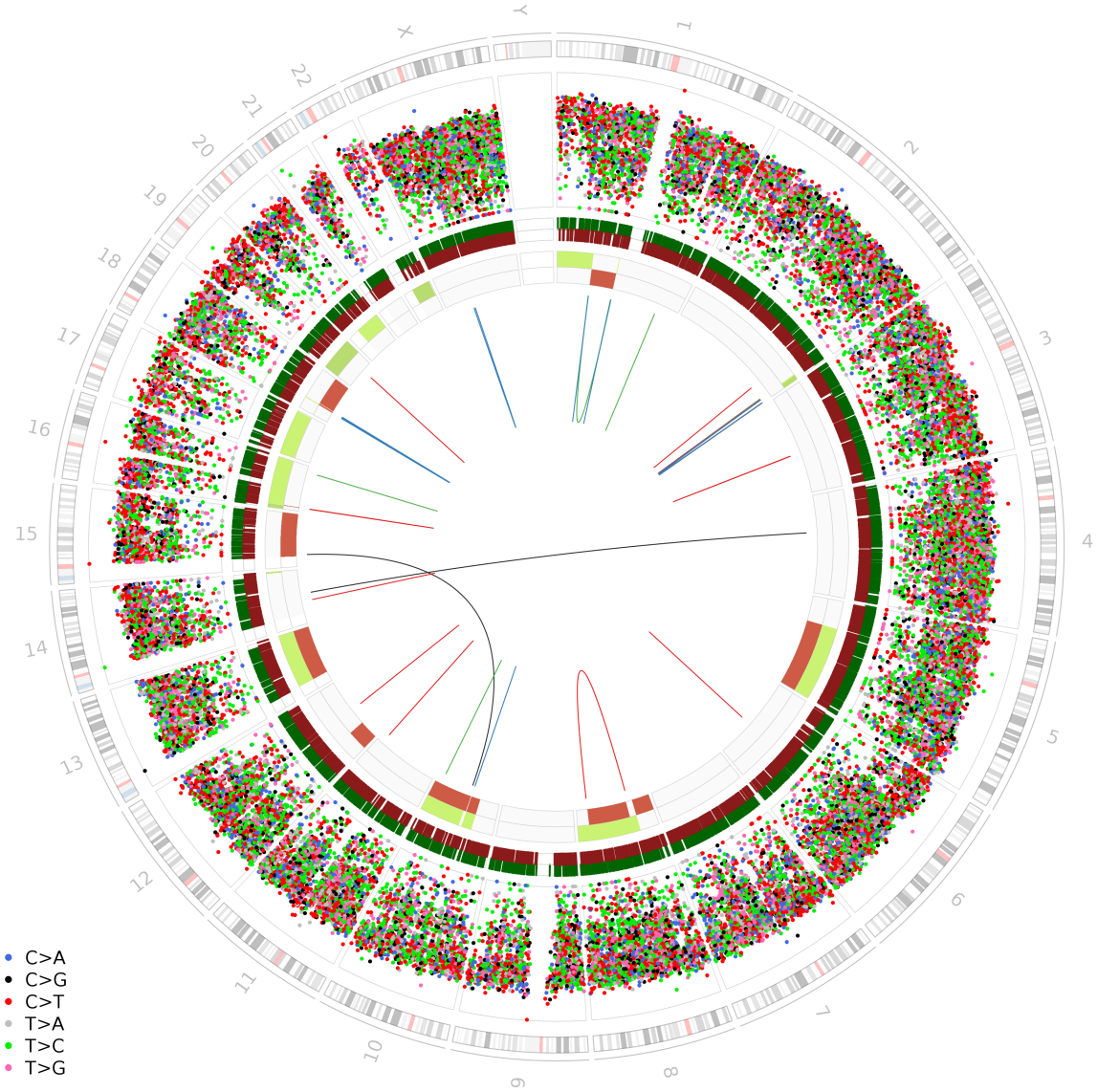
karyotype (purity 0.73, ploidy 2.2)

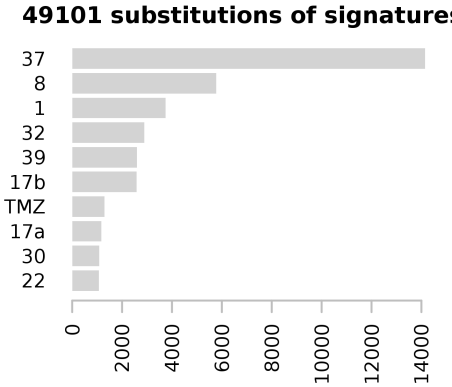


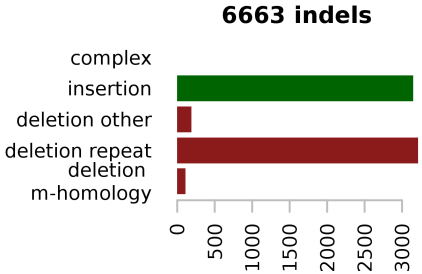
26 rearrangements



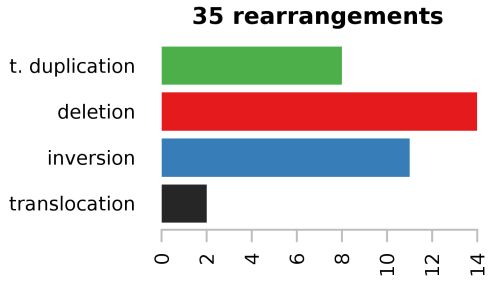
H203542 - FFPE







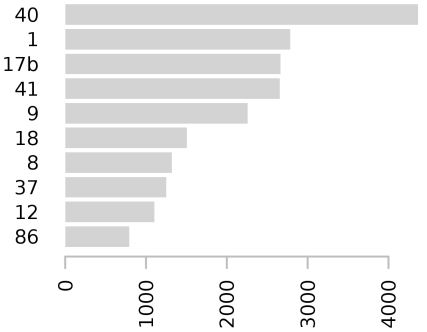




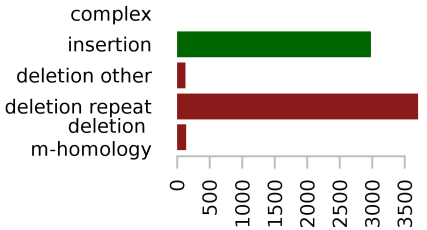
H203542 - Fresh Frozen



23846 substitutions of signatures:



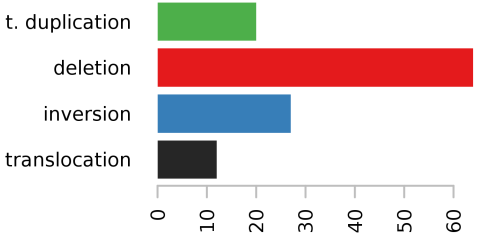
6951 indels



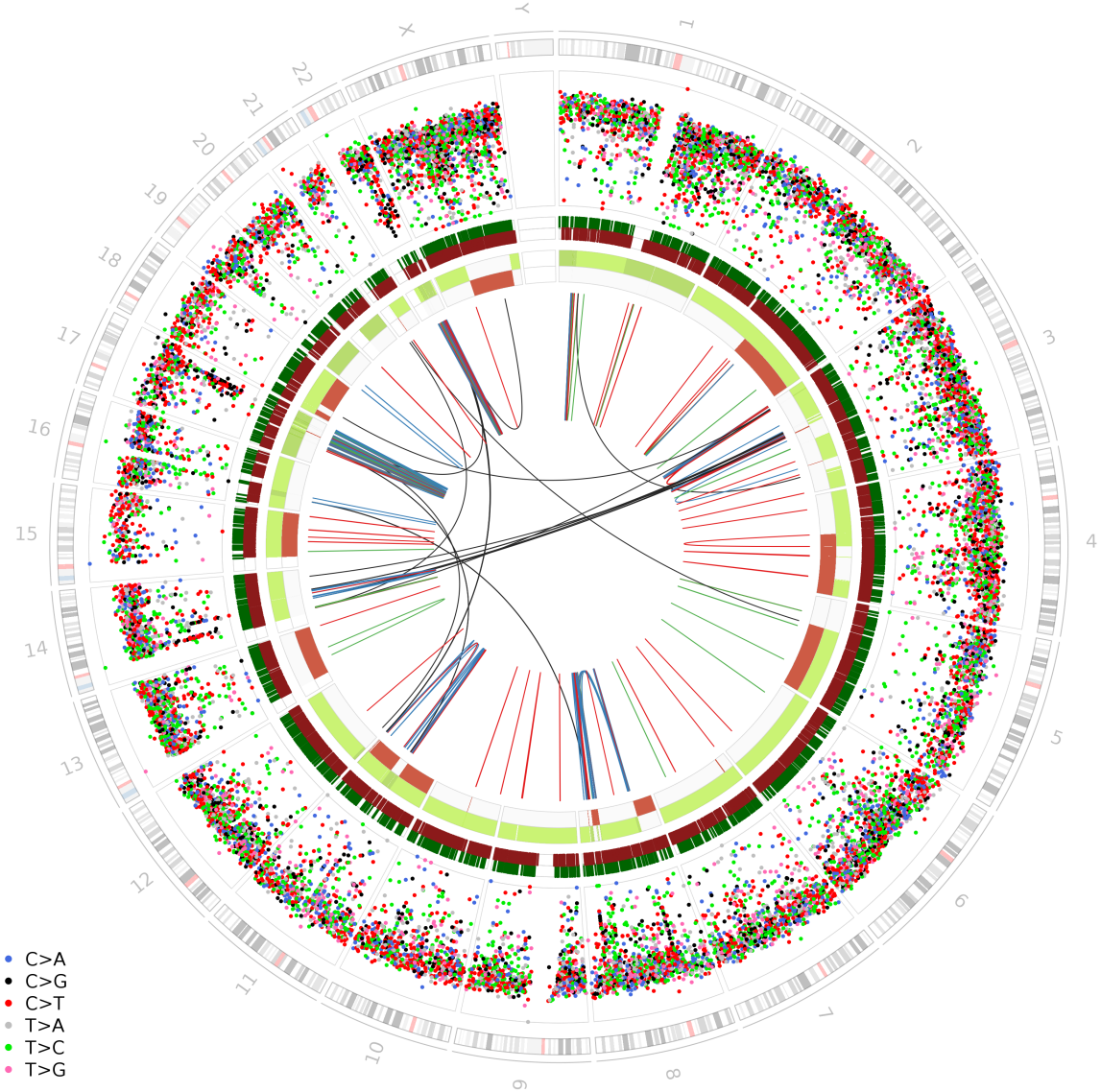
karyotype (purity 0.65, ploidy 2)



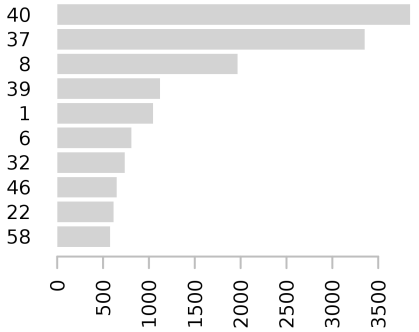
123 rearrangements



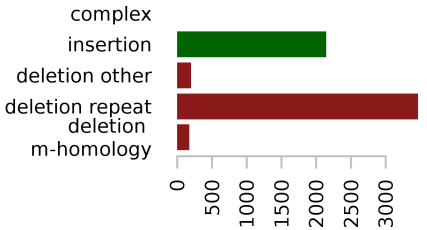
H203544 - FFPE



18080 substitutions of signatures:



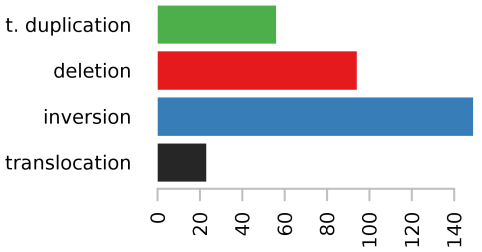
5986 indels



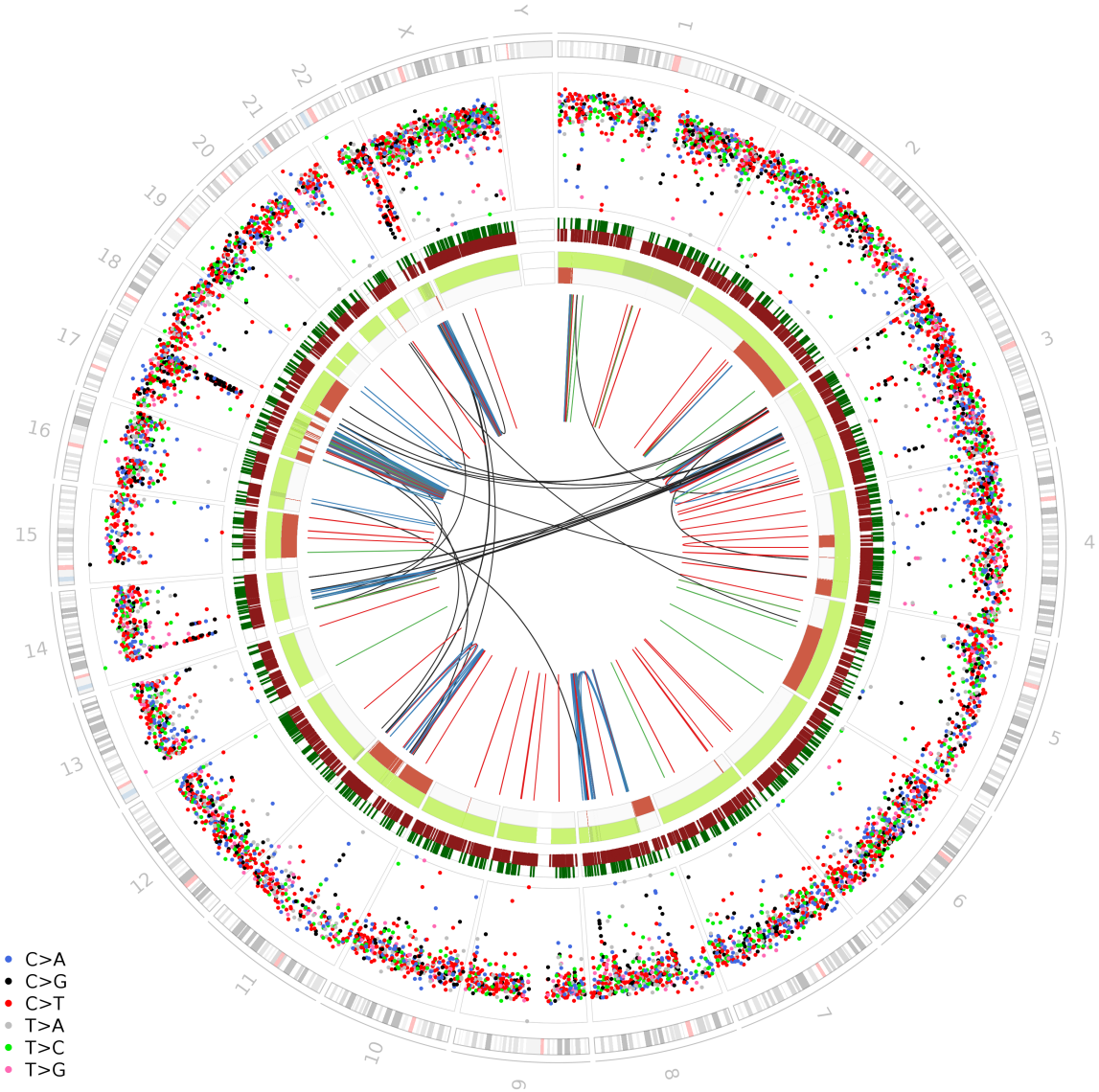
karyotype (purity 0.58, ploidy 5)



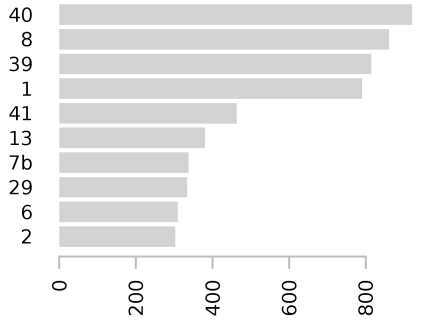
322 rearrangements



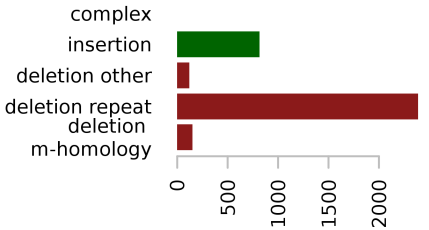
H203544 - Fresh Frozen



8090 substitutions of signatures:



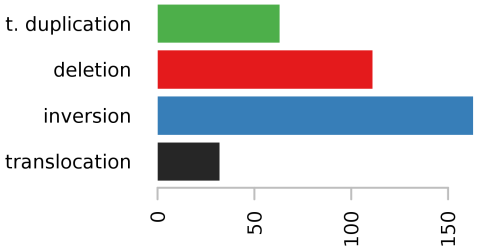
3480 indels



karyotype (purity 0.57, ploidy 3.2)



369 rearrangements



REFERENCES

Alexandrov, Ludmil B., Jaegil Kim, Nicholas J. Haradhvala, Mi Ni Huang, Alvin Wei Tian Ng, Yang Wu, Arnoud Boot, et al. 2020. "The Repertoire of Mutational Signatures in Human Cancer." *Nature* 578 (7793): 94–101.

Basyuni, Shadi, Laura Heskin, Andrea Degasperi, Daniella Black, Gene C. C. Koh, Lucia Chmelova, Giuseppe Rinaldi, et al. 2024. "Large-Scale Analysis of Whole Genome Sequencing Data from Formalin-Fixed Paraffin-Embedded Cancer Specimens Demonstrates Preservation of Clinical Utility." *Nature Communications* 15 (1): 1–12.

Chakravarty, Debyani, Sarah M. Phillips, Jianjiong Gao, Ritika Kundra, Hongxin Zhang, Julia E. Rudolph, Rona D. Yaeger, et al. 2016. "OncoKB: Annotation of the Oncogenic Effect and Treatment Implications of Somatic Mutations in Cancer." *Journal of Clinical Oncology*. https://doi.org/10.1200/jco.2016.34.15_suppl.11583.

Dodani, Dollina D., Matthew H. Nguyen, Ryan D. Morin, Marco A. Marra, and Richard D. Corbett. 2022. "Combinatorial and Machine Learning Approaches for Improved Somatic Variant Calling From Formalin-Fixed Paraffin-Embedded Genome Sequence Data." *Frontiers in Genetics* 13 (April):834764.

Domenico, Dylan, Georgios Asimomitis, Gunes Gundem, Emily Stockfisch, Cassidy Cobbs, Neeman Mohibullah, Leonidas G. Alexopoulos, Neerav Shukla, Andrew Kung, and Elli Papaemmanuil. n.d. "High-Resolution Interpretable Classification of Artifacts versus Real Variants in Whole Genome Sequencing Data from Archived Tissue." Workshop on Computational Biology. https://icml-compbio.github.io/2023/papers/WCBICML2023_paper116.pdf.

Forbes, Simon A., David Beare, Harry Boutselakis, Sally Bamford, Nidhi Bindal, John Tate, Charlotte G. Cole, et al. 2017. "COSMIC: Somatic Cancer Genetics at High-Resolution." *Nucleic Acids Research* 45 (D1): D777–83.

Guo, Qingli, Eszter Lakatos, Ibrahim Al Bakir, Kit Curtius, Trevor A. Graham, and Ville Mustonen. 2022. "The Mutational Signatures of Formalin Fixation on the Human Genome." *Nature Communications* 13 (1): 4487.

Poplin, Ryan, Pi-Chuan Chang, David Alexander, Scott Schwartz, Thomas Colthurst, Alexander Ku, Dan Newburger, et al. 2018. "A Universal SNP and Small-Indel Variant Caller Using Deep Neural Networks." *Nature Biotechnology* 36 (10): 983–87.

Shukla, N., M. F. Levine, G. Gundem, D. Domenico, B. Spitzer, N. Bouvier, J. E. Arango-Ossa, et al. 2022. "Feasibility of Whole Genome and Transcriptome Profiling in Pediatric and Young Adult Cancers." *Nature Communications*. https://doi.org/10.1038/s41467-022-30233-7.