SUPPLEMENTAL METHODS AND MATERIALS

Whole genome sequencing

Alignment:

Short insert paired-end reads were aligned to the GRCh37 reference human genome with 1000 genomes decoy contigs using BWA-mem(1).

Somatic mutation calling:

An integrated map of genetic variation from 1092 human genomes

Substitution:

Single base substitutions were called using CaVEMan (Cancer Variants through Expectation Maximisation) (http://cancerit.github.io/CaVEMan/). As described previously(2), the algorithm compares sequence data from each tumor sample to its own matched non-cancerous sample and calculates a mutation probability at each genomic locus. Copy number and cellularity information for CaVEMan were predicted with the Battenberg algorithm (2) using 1000 Genomes (3) loci within the NGS data. To improve specificity, a number of post-processing filters were applied as follows:

1. At least a third of the alleles containing the mutant must have base quality >= 25.

2. If mutant allele coverage >= 10X, there must be a mutant allele of at least base quality 20 in the middle 3rd of a read. If mutant allele coverage is < 10X, a mutant allele of at least base quality 20 in the first 2/3 of a read is acceptable.

3. The mutation position is marked by <3 reads in any sample in the unmatched normal panel.

4. The mutant allele proportion must be >5 times than that in the matched normal sample (or it is zero in the matched normal).

5. If the mean base quality is <20 then less than 96% of mutations carrying reads are in one direction.

6. Mutations within simple repeats, centromeric repeats, regions of excessive depth (https://genome.ucsc.edu/) and low mapping quality were excluded.

Additional unmatched normal filtering was performed using a set of unmatched normal samples. Mutations that were detected in >5% of the unmatched normal normal panel at >=5% mutant allele burden were excluded.

Variant annotation was done in Ensembl v74 using VAGrENT(4).

Small insertions and deletions:

Small somatic insertions and deletions (indels) were identified using a modified version of Pindel (https://github.com/cancerit/cgpPindel)(5). To improve specificity, a number of post-processing filters were applied that required the following:

1) For regions with sequencing depth <200X, mutant variant must be present in at least 8% of total reads.

2) For regions with sequencing depth >=200X, mutant variant must be present in at least 4% of total reads.

3) The region with the variant should have <= 9 small (<4 nucleotides) repeats.

4) The variant is not seen in any reads in the matched normal sample or the unmatched normal panel.

5) The number of Pindel calls in the tumour sample is greater than 4 and either:

a. The number of mutant reads mapped by BWA in the tumour sample is greater than 0 or

b. The number of mutant reads mapped by BWA in the tumour sample is equal to 0 but there are no repeats in the variant region and there are reads mapped by Pindel in the tumour sample on both the positive and negative strand.

6) Pindel 'SUM-MS' score (sum of the mapping scores of the reads used as anchors) >=150

Additional unmatched normal filtering was performed using a set of unmatched normal samples (n=221). Mutations that were detected in >1% of the unmatched normal panel at >=1% mutant allele burden were excluded.

Variant annotation was done in Ensembl v74 using VAGrENT(4).

Structural rearrangements:

Structural rearrangements were detected by an in house algorithm, BRASS (Breakpoints via assembly) [https://github.com/cancerit/BRASS], which first groups discordant read pairs that span the same breakpoint and then using Velvet de novo assembler (6) performs local assembly within the vicinity to reconstruct and determine the exact position of the breakpoint to nucleotide precision.

Copy number changes :

Segmental copy number information was derived for each sample using the Battenberg algorithm as previously described(2). Briefly, the algorithm phases heterozygous SNPs with use of the 1000 genomes genotypes as a reference panel. The resulting haplotypes are corrected for occasional errors in phasing in regions with low linkage disequilibrium. After segmentation of the resulting b-allele frequency (BAF) values, t-tests are performed on the BAFs of each copy number segment to identify whether they correspond to the value resulting from a fully clonal copy number change. If not, the copy number segment is represented as a mixture of 2 different copy number states, with the fraction of cells bearing each copy number state estimated from the average BAF of the heterozygous SNPs

in that segment. The Battenberg algorithm could not be applied to chromosome X since BAFs are uninformative for male subjects. For this chromosome, logR values were segmented and segmented logR values were converted to copy number estimates as described previously(7). Without application of the Battenberg algorithm, the resolution of subclonal copy number states is not possible, so copy number segments are called as single integer values (corresponding to the copy number state of the dominant cancer clone) on chromosome X.

Mutational signature analysis:

Mutational signature analysis of the substitutions was performed using the R package DeconstructSigs(8). Small insertion/deletions were interrogated for the presence of either short tandem repeat or microhomology at the breakpoints as described previously(2). Complex indels were excluded from this analysis.

Clonality analysis:

For each mutation we calculated the cancer cell fraction as previously described(9), using the mutant allele burden, tumor purity and locus specific copy number in the tumor and matched normal. Subclones were identified by clustering the cancer cell fractions with Dirichlet Process-based clustering as described previously(10).

Custom target sequencing

Massively parallel sequencing of postcapture libraries was performed on an Illumina HiSeq2500 system as 2×100 base pair (bp) paired-end reads for 74 primary tumor and 11 metastasis samples as well as for matching normal DNA. All samples were analysed using MSK-IMPACT, as previously described(11). This clinical sequencing platform is a hybridization capture-based next-

generation sequencing assay for targeted deep sequencing of all exons and selected introns of 341 or 410 oncogenes, TSGs, and members of pathways deemed actionable by targeted therapies (Supplemental Table 2).

Sequence reads were aligned to the human reference genome GRCh37 using the Burrows-Wheeler Aligner software (BWA, v0.7.10)(1). The Genome Analysis Toolkit (GATK, v3.1.1) (12)was used for local realignment, duplicate removal and base quality recalibration. Somatic single nucleotide variants (SNVs) were identified using MuTect (v1.0) (13), and small insertions and deletions (indels) were detected using Strelka (v2.0.15) (14)and VarScan 2 (v2.3.7)(15).

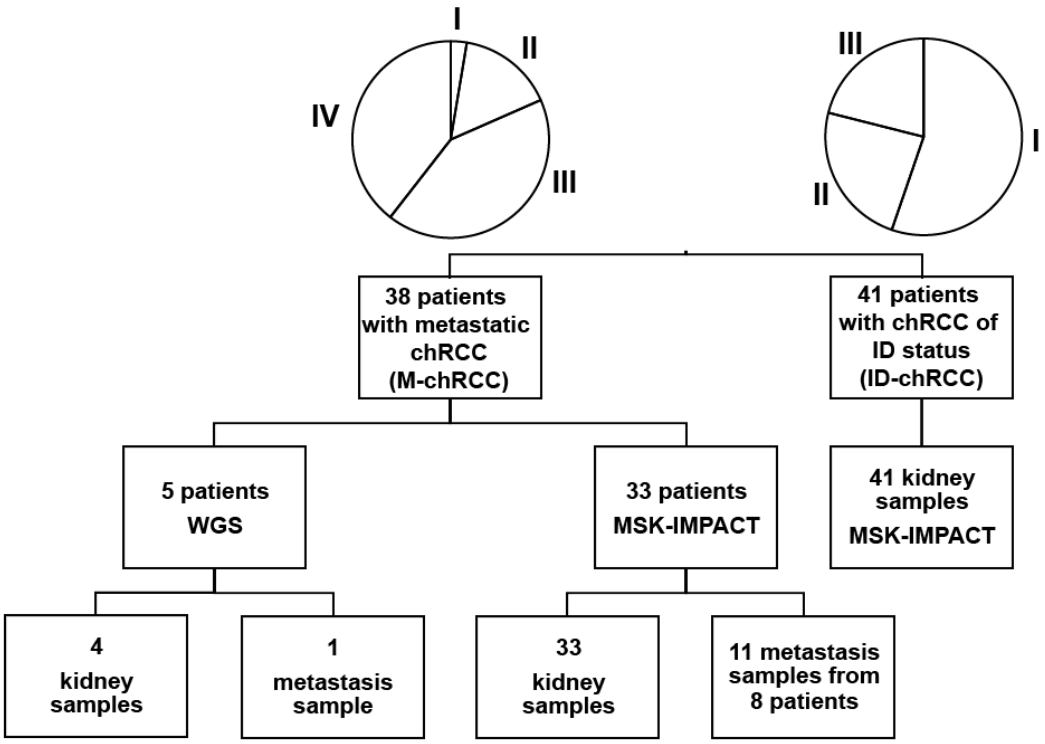
Sequence data were demultiplexed using BCL2FASTQ Conversion Software version 1.8.3 (Illumina, San Diego, CA), aligned in paired-end mode to the hg19 b37 version of the human genome using BWA-MEM software (Burrows-Wheeler Aligner), and used MuTect and SomaticIndelDetector to identify point mutations/single nucleotide variants (SNVs) and small insertions/deletion (indels). In cases where variant calling was performed for a tumor without a normal match, variants with minor allele frequency >1% in the 1,000 genomes cohort were removed.

SNVs and indels outside of the target regions, those with mutant allelic fraction (MAF) of <1% and/or those supported by ≤5 reads (16) were filtered out. We further excluded SNVs and indels for which the tumor MAF was <5 times that of the matched normal MAF, as well as SNVs and indels found at >5% global minor allele frequency of dbSNP. The mean coverage across tested samples was *371X*. Probe sequences and concentrations were optimized to ensure maximally uniform and reproducible coverage across targets. As a result, more than 99% of exons were covered to greater than 20% of the median exon coverage for each sample (71X). To reduce false-positive findings, we included only variants with depth of unique sequencing coverage above 50X and

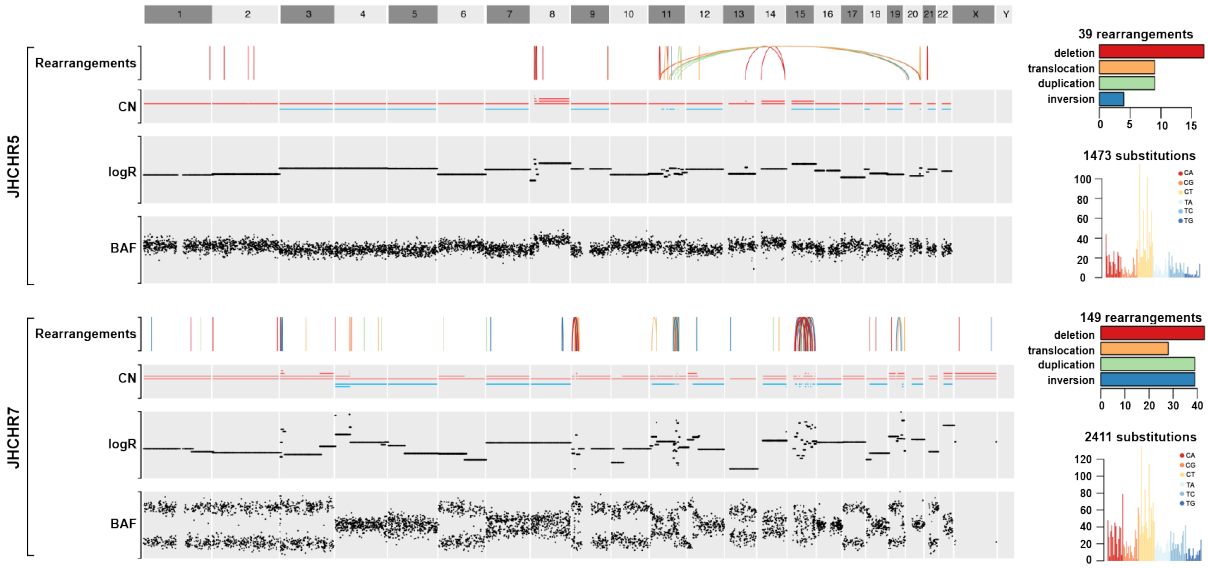
variant allele frequencies above 25%. Variants not annotated as common population polymorphisms (<1% population frequency in the 1000 genomes and NHLBI ESP cohorts http://oncology.jamanetwork.com/article.aspx?articleid=2469517-coi150100r19) but present at frequencies greater than 5% in our cohort were flagged as possible systematic assay artifacts.

All putative mutations were reviewed manually using the Integrative Genomics Viewer (IGV)(17).

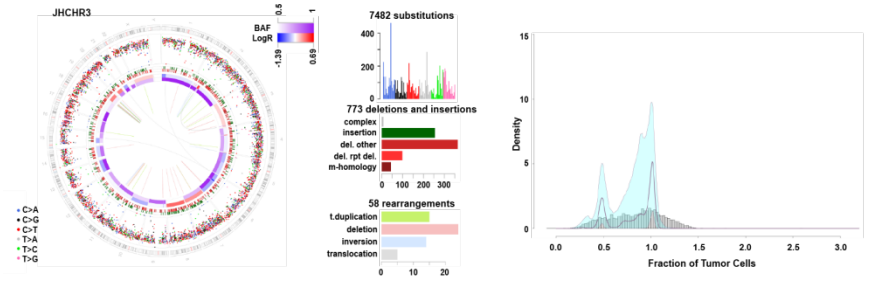
SUPPLEMENTAL FIGURES

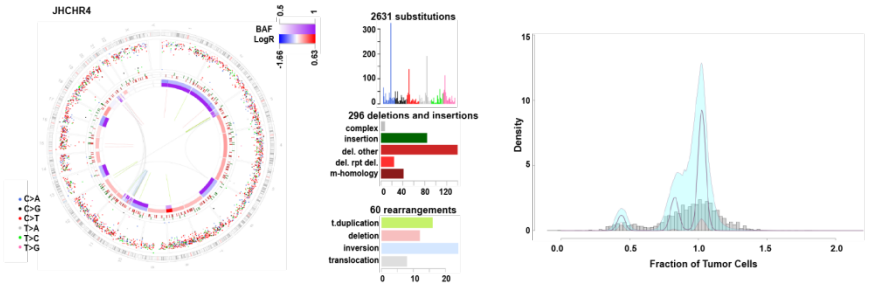


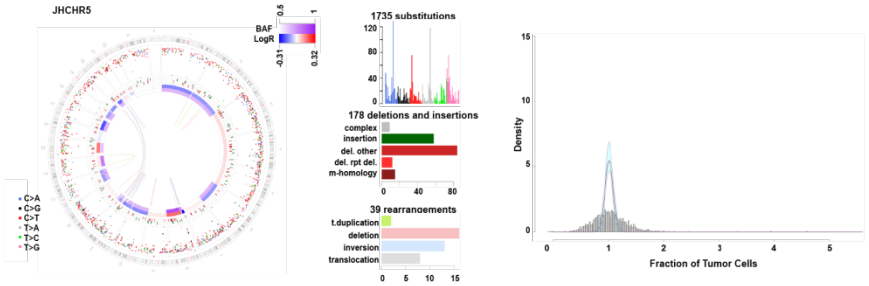
Supplemental Figure 1. Overview of the examined patient's sample cohort of chRCC. Pie charts depict the AJCC stages at the time of diagnosis for 38 metastatic chRCC (M-chRCC) patients and 41 non-metastatic ID-chRCC patients with no documented metastases at final data collection.

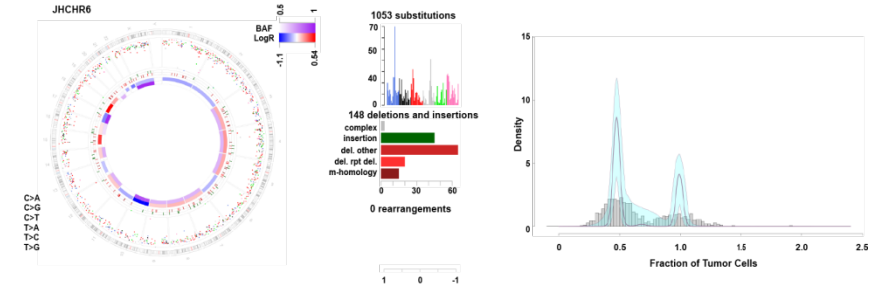


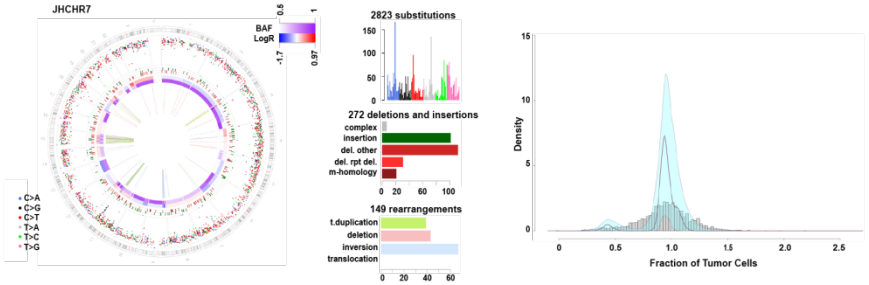
Supplemental Figure 2. Whole Genome Sequencing Analysis of Metastatic chRCC Cases in the exploratory cohort. Somatic alterations identified from WGS data for patients JHCHR5 (lung metastasis) and JHCHR7 (primary tumors). The uppermost track depicts inter- and intra-chromosomal rearrangements as arcs whilst the second and third tracks show the genomic positions of the small insertion/deletions and substitutions, respectively. Inter-mutation distance for substitutions is plotted as a line in the third track. The bottom three tracks depict the total copy number (red, increase; blue, decrease;), logR and B allele frequency (BAF) in the tumor when compared to the normal. Rearrangements, indels and substitutions are coloured according to alteration type which is summarised in the bar plots in the side panel**.** See Supplemental method for details of the classification of insertion/deletions to micro homology or repeat-mediated type.


JHCHR3
BAF
LogR
0.5
1
-1.39
0.69
C>A
C>G
C>T
T>A
T>C
T>G
7482 substitutions
773 deletions and insertions
complex
insertion
del. other
del. rpt del.
m-homology
58 rearrangements
t.duplication
deletion
inversion
translocation
Density
Fraction of Tumor Cells

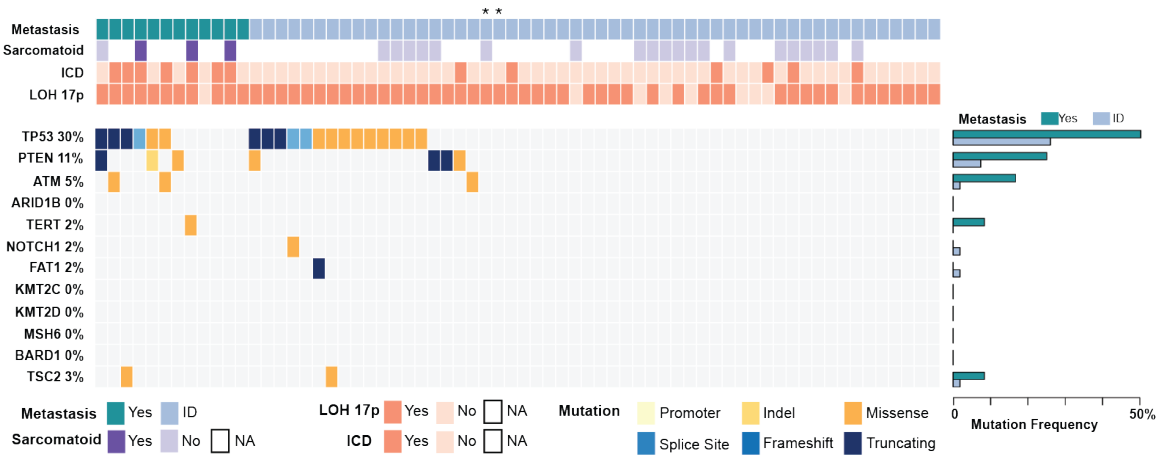

JHCHR4
BAF
LogR
0.5
1
-1.66
0.63
C>A
C>G
C>T
T>A
T>C
T>G
2631 substitutions
296 deletions and insertions
complex
insertion
del. other
del. rpt del.
m-homology
60 rearrangements
t.duplication
deletion
inversion
translocation
Density
Fraction of Tumor Cells


JHCHR5
BAF
LogR
0.5
1
-0.31
0.32
C>A
C>G
C>T
T>A
T>C
T>G
1735 substitutions
178 deletions and insertions
complex
insertion
del. other
del. rpt del.
m-homology
39 rearrangements
t.duplication
deletion
inversion
translocation
Density
Fraction of Tumor Cells

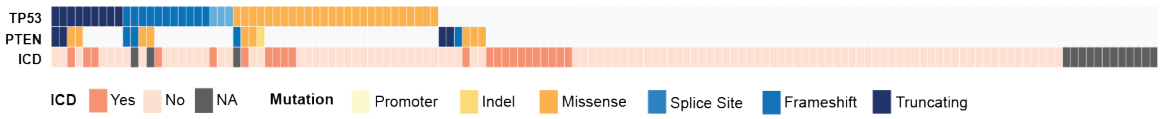

JHCHR6
BAF
LogR
0.5
1
-1.1
0.54
C>A
C>G
C>T
T>A
T>C
T>G
1053 substitutions
148 deletions and insertions
complex
insertion
del. other
del. rpt del.
m-homology
0 rearrangements
Density
Fraction of Tumor Cells


JHCHR7
BAF
LogR
0.5
1
-1.7
0.57
C>A
C>G
C>T
T>A
T>C
T>G
2823 substitutions
272 deletions and insertions
complex
insertion
del. other
del. rpt del.
m-homology
149 rearrangements
t.duplication
deletion
inversion
translocation
Density
Fraction of Tumor Cells

Supplemental Figure 3. Genomic Features revealed by WGS in the exploratory Cohort. Circos plots summarize the different types of alterations identified in the 5 patients that were whole-genome sequenced. Genomic rearrangements are plotted as arcs in the inner most track followed by the two tracks that show copy number changes as LogR and BAF. Genomic positions of the indels are shown in the fourth track. Inter-mutation distance for the substitutions is plotted in the outermost track. All mutations are coloured according to the alteration type summarized in the bar plots in the right panels. The second plot summarizes the results from the statistical analysis of the cancer cell fractions of the clonal and subclonal mutations by a Bayesian Dirichlet process-based clustering. Empiric histogram of substitutions is shown in grey together with the density from clustering in pale green and the fitted distribution in dark pink.



Supplemental Figure 4. Genomic alterations in the TCGA-KICH cohort by WES/WGS. Oncoprint of ploidy status and frequency of nonsynonymous mutations in 66 primary tumors. Asterisks denote samples excluded from the statistical analysis due to insufficient clinical data. The presence of sarcomatoid histology is denoted. Imbalanced chromosome duplication (ICD) status and loss of heterozygosity (LOH) at chromosome 17p were detected with FACETS. Mutation frequencies of individually specified genes are listed on the left. Mutation frequencies of individual genes in M-chRCC and ID-chRCC cases are shown as bar plots on the right. Gene mutation types are color-coded. ICD status and LOH at chromosome 17p were detected with FACETS. Gene mutation types are color-coded.



Supplemental Figure 5. Patterns of 3 high-risk genomic features in 140 chRCC cases (33 M-chRCC, 41 ID-chRCC and 66 TCGA-KICH). *PTEN* mutations and ICD appear mutually exclusive (Fisher, P=0.0033).

SUPPLEMENTAL DATA FILES

Supplemental Table 1. Clinical Data chRCC

Supplemental Table 2. MSK-IMPACT List chRCC

Supplemental Table 3. Mutations chRCC

Supplemental Table 4. Primary Metastasis Pairs chRCC

Supplemental Table 5. Influence of *TP53, PTEN* mutations and ICD on Survival in TCGA-KICH

Supplemental Table 6. Influence of *TP53, PTEN* mutations, sarcomatoid status and ICD on survival in M-chRCC cohort

Supplemental Table 7. TCGA KIRC-KICH Data

REFERENCES

1. Li H, and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England).* 2010;26(5):589-95.
2. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell.* 2012;149(5):994-1007.
3. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56-65.
4. Menzies A, Teague JW, Butler AP, Davies H, Tarpey P, Nik-Zainal S, et al. VAGrENT: Variation Annotation Generator. *Current protocols in bioinformatics.* 2015;52:15.8.1-1.
5. Raine KM, Hinton J, Butler AP, Teague JW, Davies H, Tarpey P, et al. cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events from Paired End Sequencing. *Current protocols in bioinformatics.* 2015;52:15.7.1-2.
6. Zerbino DR, and Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research.* 2008;18(5):821-9.
7. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences of the United States of America.* 2010;107(39):16910-5.
8. Rosenthal R, McGranahan N, Herrero J, Taylor BS, and Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome biology.* 2016;17:31.
9. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun.* 2014;5:2997.
10. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JM, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature.* 2015;520(7547):353-7.
11. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *The Journal of molecular diagnostics : JMD.* 2015;17(3):251-64.
12. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research.* 2010;20(9):1297-303.
13. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology.* 2013;31(3):213-9.
14. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, and Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics (Oxford, England).* 2012;28(14):1811-7.

15. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research.* 2012;22(3):568-76.
16. De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng CK, Nuciforo P, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2014;25(9):1729-35.
17. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nature biotechnology.* 2011;29(1):24-6.