Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Jamal-Hanjani M, Wilson GA, McGranahan N, et al. Tracking the evolution of non–small-cell lung cancer. N Engl J Med 2017;376:2109-21. DOI: 10.1056/NEJMoa1616288

(PDF updated June 1, 2017)

Supplementary Appendix

Supplement to: TRACERx – Tracking Non-Small Cell Lung Cancer Evolution

Corresponding author:
Charles Swanton
Translational Cancer Therapeutics Laboratory
The Francis Crick Institute
3rd Floor South West
1 Midland Road
London
NW1 1AT

Email: Charles.Swanton@crick.ac.uk
Office +44 203 796 2047

Table of Contents

The TRACERx consortium members and collaborators

The TRACERx study (Clinicaltrials.gov no: NCT01888601) is sponsored by University College London (UCL/12/0279) and has been approved by an independent Research Ethics Committee (13/LO/1546). TRACER is funded by Cancer Research UK (grant number C11496/A17786) and coordinated through the Cancer Research UK & UCL Cancer Trials Centre.

Names of consortium members and collaborators

Charles Swanton[1,2,5], Mariam Jamal-Hanjani[1], Selvaraju Veeriah[1], Seema Shafi[1], Justyna Czyzewska-Khan[1], Diana Johnson[1], Joanne Laycock[1], Leticia Bosshard-Carter[1], Gerald Goh[1], Rachel Rosenthal[1], Pat Gorman[1], Nirupa Murugaesu[1], Robert E Hynds[1,3], Gareth Wilson[1,2], Nicolai J Birkbak[1,2], Thomas B K Watkins[2], Nicholas McGranahan[1,2], Stuart Horswell[2], Richard Mitter[2], Mickael Escudero[2], Aengus Stewart[2], Peter Van Loo[2], Andrew Rowan[2], Hang Xu[2], Samra Turajlic[2,4], Crispin Hiley[2], Christopher Abbosh[1], Jacki Goldman[2], Richard Kevin Stone[2], Tamara Denner[2], Nik Matthews[2], Greg Elgar[2], Sophia Ward[2], Jennifer Biggs[2], Marta Costa[2], Sharmin Begum[2], Ben Phillimore[2], Tim Chambers[2], Emma Nye[2], Sofia Graca[2], Maise Al Bakir[2], Kroopa Joshi[1], Andrew Furness[1], Assma Ben Aissa[1], Yien Ning Sophia Wong[1], Andy Georgiou[1], Sergio Quezada[1], John A Hartley[1], Helen L Lowe[1], Javier Herrero[1], David Lawrence[5], Martin Hayward[5], Nikolaos Panagiotopoulos[5], Davide Patrini[5], Shyam Kolvekar[5], Mary Falzon[5], Elaine Borg[5], Teresa Marafioti[5], Celia Simeon[5], Gemma Hector[5], Amy Smith[5], Marie Aranda[5], Marco Novelli[5], Dahmane Oukrif[5], Sam M Janes[5], Ricky Thakrar[5], Martin Forster[5], Tanya Ahmad[5], Siow Ming Lee[5], Dionysis Papadatos-Pastos[5], Dawn Carnell[5], Ruheena Mendes[5], Jeremy George[5], Neal Navani[5], Asia Ahmed[5], Magali Taylor[5], Penny Shaw[5], Junaid Choudhary[5], Yvonne Summers[6], Raffaele Califano[6], Paul Taylor[6], Rajesh Shah[6], Piotr Krysiak[6], Kendadai Rammohan[6], Eustace Fontaine[6], Richard Booton[6], Matthew Evison[6], Phil Crosbie[6], Stuart Moss[6], Faiza Idries[6], Leena Joseph[6], Paul Bishop[6], Anshuman Chaturved[6], Anne Marie Quinn[6], Helen Doran[6], Angela leek[7], Phil Harrison[7], Katrina Moore[7], Rachael Waddington[7], Juliette Novasio[7], Fiona Blackhall[8], Jane Rogan[7], Elaine Smith[6], Caroline Dive[9], Jonathan Tugwood[9], Ged Brady[9], Dominic G Rothwell[9], Francesca Chemi[9], Jackie Pierce[9], Sakshi Gulati[9], Babu Naidu[10], Gerald Langman[10], Simon Trotter[10], Mary Bellamy[10], Hollie Bancroft[10], Amy Kerr[10], Salma Kadiri, [10] Joanne Webb[10], Gary Middleton[10], Madava Djearaman[10], Dean Fennell[11], Jacqui A Shaw[11], John Le Quesne[11], David Moore[11], Apostolos Nakas[12], Sridhar Rathinam[12], William Monteiro[13], Hilary Marshall[13], Louise Nelson[12], Jonathan Bennett[12], Joan Riley[12], Lindsay Primrose[12], Luke Martinson[12], Girija Anand[14], Sajid Khan[15], Anita Amadi[16], Marianne Nicolson[17], Keith Kerr[17], Shirley Palmer[17], Hardy Remmen[17], Joy Miller[17], Keith Buchan[17], Mahendran Chetty[17], Lesley Gomersall[17], Jason Lester[18], Alison Edwards[18], Fiona Morgan[19], Haydn Adams[19], Helen Davies[19], Malgorzata Kornaszewska[20], Richard Attanoos[21], Sara Lock[22], Azmina Verjee[22], Mairead MacKenzie[23], Maggie Wilcox[23], Harriet Bell[24], Natasha Iles[24], Allan Hackshaw[24], Yenting Ngai[24], Sean Smith[24], Nicole Gower[24], Christian Ottensmeier[25], Serena Chee[25], Benjamin Johnson[25], Aiman Alzetani[25], Emily Shaw[25], Eric Lim[26], Paulo De Sousa[26], Monica Tavares Barbosa[26], Alex Bowman[26], Simon Jordan[26], Alexandra Rice[26], Hilgardt Raubenheimer[26], Chiara Proli[26], Maria Elena Cufari[26], John Carlo Ronquillo[26], Angela Kwayie[26], Harshil Bhayani[26], Morag Hamilton[26], Yusura Bakar[26], Natalie Mensah[26], Lyn Ambrose[26], Anand Devaraj[26], Silviu Buderi[26], Jonathan Finch[26], Leire Azcarate[26], Hema Chavan[26], Sophie Green[26], Hillaria Mashinga[26], Andrew G Nicholson[26, 27], Kelvin Lau[28], Michael Sheaff[28], Peter Schmid[28], John Conibear[28], Veni Ezhil[29], Babikir Ismail[29], Melanie Irvin-sellers[29], Vineet Prakash[29], Peter Russell[30], Teresa Light[30], Tracey Horey[30], Sarah Danson[31], Jonathan Bury[31], John Edwards[31], Jennifer Hill[31], Sue Matthews[31], Yota Kitsanta[31], Kim Suvarna[31], Patricia Fisher[31], Allah Dino Keerio[31], Michael Shackcloth[32], John Gosney[32], Pieter Postmus[32], Sarah Feeney[32], Julius Asante-Siaw[32], Tudor Constatin[33], Raheleh Salari[33], Nicole Sponer[33], Ashwini Naik[33], Bernhard Zimmermann[33], Hugo J.W.L. Aerts[34], Stefan Dentro[35], Christophe Dessimoz[36,37,38], Karl Peggs[1,39,40].

Affiliations

1. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, United Kingdom
2. The Francis Crick Institute, United Kingdom
3. Lungs for Living, UCL Respiratory, University College London, United Kingdom
4. The Royal Marsden Hospital, United Kingdom
5. University College London Hospitals NHS Foundation Trust, United Kingdom
6. University Hospital of South Manchester, United Kingdom
7. Manchester Cancer Research Centre Biobank, United Kingdom
8. Christie NHS Foundation Trust, Manchester, United Kingdom
9. Cancer Research UK Manchester Institute, United Kingdom
10. Heart of England NHS Foundation Trust, Birmingham, United Kingdom
11. Cancer Studies and Molecular Medicine, University of Leicester, United kingdom
12. Leicester University Hospitals, United Kingdom
13. National Institute for Health Research Leicester Respiratory Biomedical, Research Unit, United Kingdom
14. North Middlesex Hospital, United Kingdom
15. Royal Free Hospital, United Kingdom
16. Barnet Hospital, United Kingdom
17. Aberdeen Royal Infirmary, United Kingdom
18. Velindre Cancer Centre, Cardiff, Wales, United Kingdom
19. Cardiff & Vale University Health Board, Cardiff, Wales, United Kingdom
20. University Hospital Of Wales Heath Park, Cardiff, Wales, United Kingdom
21. Department of Pathology, University Hospital of Wales and Cardiff University, Heath Park, Cardiff, Wales, United Kingdom
22. The Whittington Hospital NHS Trust, United Kingdom
23. Independent Cancer Patients Voice, United Kingdom
24. Cancer Research UK & UCL Cancer Trials Centre, United Kingdom
25. University Hospital Southampton NHS Foundation Trust, United Kingdom
26. Royal Brompton and Harefield NHS Foundation Trust, United Kingdom
27. National Heart and Lung Institute, Imperial College, United Kingdom
28. Barts Health NHS Trust, United Kingdom
29. Ashford and St. Peter's Hospitals NHS Foundation Trust, United Kingdom
30. The Princess Alexandra Hospital NHS Trust, United Kingdom
31. Sheffield Teaching Hospitals NHS Foundation Trust, United Kingdom
32. Liverpool Heart and Chest Hospital NHS Foundation Trust, United Kingdom
33. Natera Inc., 201 Industrial Road, Suite 410, San Carlos, CA 94070
34. Dana-Farber Cancer Institute, Brigham & Women's Hospital, Harvard Medical School, 450 Brookline Ave, JF518, Boston, MA 02115-5450, USA
35. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, United Kingdom
36. Bioinformatics Group, Department of Computer Science, University College London
37. University of Lausanne
38. Swiss Institute of Bioinformatics
39. Cancer Immunology Unit, University College London Cancer Institute, United Kingdom
40. Research Department of Haematology, University College London Cancer Institute, United Kingdom

TRACERx study inclusion and exclusion criteria

The TRACERx protocol has been included as a supplementary file, but in summary the following eligibility criteria was used to recruit patients into the TRACERx study:

Inclusion criteria

- Written Informed consent
- Patients ≥18 years of age, with early stage I-IIIA disease who are eligible for primary surgery
- Histopathologically confirmed NSCLC, or a strong suspicion of cancer on lung imaging necessitating surgery (e.g. diagnosis determined from frozen section in theatre)
- Primary surgery in keeping with NICE (National Institute for Health and Care Excellence) guidelines planned
- Agreement to be followed up in a specialist centre
- ECOG performance status 0 or 1

Exclusion criteria

- Any other current malignancy or malignancy diagnosed or relapsed within the past 5 years (other than non-melanomatous skin cancer, stage 0 melanoma *in situ*, and *in situ* cervical cancer)
- Psychological condition that would preclude informed consent
- Treatment with neo-adjuvant therapy for current lung malignancy deemed necessary
- Adjuvant therapy other than platinum-based chemotherapy and/or radiotherapy
- Known Human Immunodeficiency Virus (HIV), Hepatitis B Virus (HBV), Hepatitis C Virus (HCV) or syphilis infection.
- Sufficient tissue, i.e. a minimum of two tumor regions, is unlikely to be obtained for the study based on pre-operative imaging

The TRACERx 100 cohort

The 100 patients presented in this paper are those who met the eligibility criteria and from whom collected tumor samples could be sequenced prospectively depending on different filtering steps as outlined in the CONSORT diagram (CONSORT flow chart, Fig.S1). It therefore follows that these patients do not have consecutive study ID numbers from 0001 to 0100.

Data collection relating to environmental exposure

In a study questionnaire, all patients were asked to provide details relating to exposure to the following substances:

1. Asbestos
2. Diesel fumes
3. Natural fibres, such as silica, wood dust
4. Metals, such as aluminum, arsenic, beryllium, cadmium, chromium, nickel
5. Radon
6. Reactive chemicals, such as bis(chloromethyl) ether, mustard gas, vinyl chloride
7. Solvents, such as benzene, toluene
8. Arsenic
9. Benzene
10. BisphenolA (BPA)
11. Chromium Hexavalent compounds
12. Dioxins
13. Formaldehyde
14. Polybrominated diphenylethers (PBDEs)

15. Polycyclicaromatic hydrocarbons (PAHs)
16. Vinyl Chloride

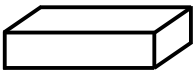
Experimental Procedures

Sample collection

All surgically resected tumor samples were macroscopically reviewed by a pathologist. Spatially separated tumor regions, documented by photography, were collected and snap frozen in liquid nitrogen for subsequent DNA extraction using the following Standard Operating Procedure.

All specimens were cut fresh in a class 2 cabinet (or possibly class 1 depending on local Standard Operating Procedures (SOP)/ or as per local practice. The tumor specimen was dissected by the pathologist with new and sterile scalpel blades on a clean surface to avoid contamination with other DNA. Large dissection knives were autoclaved, or new blades were used. When required sterile forceps were used and changed between samples. The pathologist examined the specimen macroscopically to assess whether there was enough tissue surplus to diagnostic requirements without compromising the bronchial or pleural resection margins. A cut (or more than one cut for tumors larger than 3-4 cm) fully bisecting the tumor, without compromising integrity of the whole specimen, was made. Where possible incisions toward the pleural surface, rather than through the pleura, were made to aid later microscopic assessment of the pleural surface.

Research tissues were excised from the newly exposed tumor surface using a scalpel or a 5-10mm disposable skin biopsy punch. A fresh blade or punch was used for every region sampled to avoid DNA carryover. At least 2 regions from each tumor, separated by at least 3mm, were collected for research purposes. Areas that were obviously necrotic, fibrotic or hemorrhagic were avoided to maximize viable tumor cellularity. Where possible, regions samples reflected the macroscopic morphological heterogeneity of the tumor.

Tumour diameter	**Number of regions**	**Volume of each sample**
16-29 mm	2-3 regions	Small (≈4 mm cube) 
30 – 59 mm	3 - 5 regions	Medium (≈7 mm cube) 
>59 mm	5 - 8 regions	Large (≈7x7x20 mm) 

Intrapulmonary involved lymph nodes or other malignant deposits (e.g. in chest wall resections) present in the specimen were sampled if possible. Hilar and station lymph nodes were also examined by the pathologist in the 'fresh' state. If there was clear malignant involvement, a portion, surplus to diagnostic requirements, was sampled. Individual fresh tissue samples were wrapped in thick sterile foil, labeled with the region number corresponding to the annotated photographs, and snap frozen by immediate immersion into liquid nitrogen or rapid coolant spray (Frostbite Rapid Coolant, Leica Biosystems) and immediately put on dry ice until placed in a -80°C freezer for storage.

Sample processing
Approximately 3x3x3mm of tumor tissue from each region was used for genomic DNA extraction using a modification of the DNA/RNA AllPrep kit (Qiagen). Peripheral blood was collected at the time of surgery, of which 2ml was used for germline DNA extraction using the QiaAmp blood midi kit (Qiagen). DNA was quantified by the Qubit (Invitrogen) and DNA integrity was assessed by TapeStation (Agilent Technologies).

Central histopathological review
Tumor sections from all cases were converted into digital images and reviewed centrally by two subspecialty pulmonary pathologists. All cases were reviewed in detail by at least one pathologist, and in cases of uncertainty, images were examined by both. Histological patterns were recorded as per IASLC categories, with additional incorporation of assessment of cribriform pattern.

All diagnostic tumor sections from lung adenocarcinoma (LUAD) and lung squamous cell cancer (LUSC) cases were subject to structured histopathological grading for the relevant tumor types, using pathological scoring systems previously described in large patient cohorts as being of prognostic value[1-3].

Tissue microarray creation
We designed tissue microarrays (TMAs) comprising of 100 NSCLC cases for ALK testing. Representative tumor areas were defined by examination of H&E stained sections from all 100 tissues blocks by a pathologist. From each NSCLC case two 2mm cores were selected from different regions within each biopsy resulting in a TMA of 200 cores. The selected area was then excised and re-embedded in recipient blocks. Additionally, four normal lung cores were placed as a negative control. The H&E staining confirmed that each core contained tumor tissue.

TMAs were created using the following protocol:

Donor blocks were melted and re-embedded so wax in donor and recipient blocks were of approximate depths. 2mm arraymold kits (Arraymold Kit A, USA) were selected and appropriate sized dermal biopsy punches and stylets were used. The silicone arraymold was warmed for 15 minutes at 55°C then filled with molten wax. A cassette lid was placed on top and more wax added until the cassette was completely covered, prior to transfer to a cold plate to set. Once set, the arraymold was peeled off to produce the recipient TMA block. Relevant areas were marked on H/E slides of the donor block. The dermal needle was inserted on regions of interest to make the biopsy punches. This was performed at room temperature to prevent cracks. The stylet was injected inside the dermal biopsy punch, extracting cores into the recipient TMA block. The TMA block was placed with uncharged glass slide on top faced down and put into the oven at 40°C overnight. An uncharged glass slide was heated to 70°C for 10 minutes and placed under the block. Air bubbles were removed to smooth the surface of the TMA block. The glass slide and TMA block were placed onto a cold plate for 20 minutes prior to cutting the microtome.

ALK and ROS1 status testing by IHC
The Ventana D5F3 Rabbit monoclonal antibody was used on the BenchMark XT platform according to the manufacturers guidelines and ALK expression was reported as negative or positive (no grading again as recommended by Ventana guidelines). The signaling antibody

ROS1 (D4D6) was used at a 1 in 100 dilution on the Ventana ultra using the Optiview kit with amplification and was reported as negative or positive.

Ki67 staining and scoring by IHC

Tumor samples were fixed in 10% buffered formalin and embedded in paraffin according to conventional histological protocols. 2-5µm sections from TMAs were cut and transferred onto poly-l-lysine–coated slides, dewaxed in two changes of xylene and rehydrated in a series of graded alcohols. Single immunohistochemistry for the monoclonal antibody anti-Ki67 (Dilution 1:100; clone MIB-1; DAKO Agilent Technologies LDA, UK Limited, Stockport, Cheshire SK8 3GR, UK) was carried out using the automated platform BenchMark Ultra (Ventana/Roche) according to a protocol described elsewhere[4] . The primary antibody was incubated for 32 minutes using CC1 antigen retrieval standard protocol. Detection was performed using the peroxidase-based detection reagent conjugate (OptiView DAB IHC Detection Kit; Ventana Medical Systems, Inc). Optimal working dilutions were first optimised on paraffin-embedded sections of human reactive tonsil; used as positive control. Specificity of staining reactivity pattern (i.e. nuclear labeling) was assessed by TM and the Ki-67 proliferation fraction in each core was evaluated by approximate counting of the positive nuclear staining over the total number of neoplastic cells at 40X magnification (NIKON Eclipse E400).

Whole-exome sequencing

For each tumor region (n=327) and matched germline (n=100), exome capture was performed on 1-2 µg DNA isolated from genomic libraries with median insert size of 190bp, using a customised version of the Agilent Human All Exome V5 kit according to the manufacturer's protocol (Agilent). The capture kit was extended to cover intronic regions in genomic areas linked with fusion events in NSCLC (ALK, EML4, FGFR1, FGFR2, FGFR3, KIF5B, NTRK1, RET, ROS1, TCF3). Following cluster generation, samples were 100bp paired-end multiplex sequenced on the Illumina HiSeq 2500 and HiSeq 4000 at the Advanced Sequencing Facility at The Francis Crick Institute, Lincoln's Inn Fields Laboratories, as described previously[5,6]. The generated data was aligned to the reference human genome (hg19) achieving a median sequencing depth of 431 (range 83-986) for the tumor regions and 415 (range 107-765) for the matched germline. To assess the stability in coverage across the target regions, the correlation of coverage between regions from each tumor was calculated. The median of these correlation coefficients across the cohort was 0.96, with 99% of the tumors > 0.87. The single outlier, CRUK0053, was manually corrected to ensure no mutations were erroneously classified as subclonal.

Further sequencing was performed for the purpose of validation of exome calls. Multiplex PCR and subsequent next generation sequencing (mPCR-NGS) was performed by Natera on 48 patients (CRUK0001, CRUK0002, CRUK0004, CRUK0005, CRUK0006, CRUK0008, CRUK0009, CRUK0013, CRUK0015, CRUK0017, CRUK0018, CRUK0021, CRUK0023, CRUK0024, CRUK0027, CRUK0029, CRUK0030, CRUK0031, CRUK0034, CRUK0035, CRUK0036, CRUK0039, CRUK0045, CRUK0048, CRUK0049, CRUK0051, CRUK0052, CRUK0057, CRUK0065, CRUK0066, CRUK0067, CRUK0068, CRUK0069, CRUK0070, CRUK0071, CRUK0072, CRUK0076, CRUK0077, CRUK0078, CRUK0079, CRUK0080, CRUK0081, CRUK0083, CRUK0084, CRUK0085, CRUK0088, CRUK0095, CRUK0098). This involved 175 tumor regions comprising of 792 predicted non-silent mutations with a median coverage of 3958 (range 32-53,030). Multiplex PCR assays specific to each patient were designed, representing both clonal and subclonal mutations. Following primer design, corresponding primers were ordered and primer pools prepared and QC'ed. The mPCR protocol was optimized for each pool. DNA from tumor and matched germline samples was purified, and quantified prior to mPCR and sequencing on HiSeq 2500. Further orthogonal validation was performed for 1073 (535 non-silent and 538 silent) mutations selected from 11 patients (CRUK0076, CRUK0063, CRUK0077, CRUK0065, CRUK0004, CRUK0098, CRUK0069,

CRUK0018, CRUK0079, CRUK0008, CRUK0029). An Ion AmpliSeq™ custom panel (Life Technologies) was designed using the online designer (www.ampliseq.com). Multiplex PCRs were performed on DNA from each region of the relevant tumor, a total of 49 regions, according to the manufacturer's protocol. Barcoded sequencing libraries were constructed, which were sequenced with 200bp read length on the Ion Torrent PGM sequencer (Life Technologies). The Ion Torrent Torrent Suite™ software was used to perform sequence alignment to target regions from the hg19 genome. A median coverage of 815 (range 31 – 61,720) was obtained across the target regions.

Utilizing SNP frequencies to prevent inter-patient sample swaps

All tumor regions and the associated germline sample from a single patient should, in the absence of somatic changes, have highly similar SNP profiles thus demonstrating a common patient of origin. The variant allele frequencies of 24 SNPs previously identified by Pengelly et al[7], were routinely extracted for each sequenced tumor region and compared in order to confirm that no tumor DNA samples swaps had occurred between patients.

SNV, INDEL and Dinucleotide substitution calling from multi-region whole exome sequencing

Raw paired end reads (100bp) in FastQ format generated by the Illumina pipeline were aligned to the full hg19 genomic assembly (including unknown contigs) obtained from GATK bundle 2.8[8], using bwa mem (bwa-0.7.7)[9]. Picard tools v1.107 was used to clean, sort and merge files from the same patient region and to remove duplicate reads (http://broadinstitute.github.io/picard). A combination of picard tools (1.107), GATK (2.8.1) and FastQC (0.10.1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) were used to produce quality control metrics.

SAMtools mpileup (0.1.19)[10] was used to locate non-reference positions in tumor and germline samples. Bases with a phred score of <20 or reads with a mapping-quality <20 were omitted. BAQ computation was disabled and the coefficient for downgrading mapping quality was set to 50. VarScan2 somatic (v2.3.6)[11] utilized output from SAMtools mpileup in order to identify somatic variants between tumor and matched germline samples. With the exception of minimum coverage for the germline sample that was set to 10, minimum variant frequency was changed to 0.01 and tumor purity was set to 0.5, default parameters were used. VarScan2 processSomatic was used to extract the somatic variants. The resulting single nucleotide variant (SNV) calls were filtered for false positives using Varscan2's associated fpfilter.pl script, initially with default settings then repeated again with min-var-frac = 0.02, having first run the data through bam-readcount (0.5.1) (https://github.com/genome/bam-readcount). All insertion/deletion (INDEL) calls in reads classed as 'high confidence' by VarScan2 processSomatic were recorded for further downstream filtering.

MuTect (1.1.4)[12] was also used to detect SNVs utilising annotation files contained in GATK bundle 2.8. Following completion, variants called by MuTect were filtered according to the filter parameter 'PASS'.

To further reduce false positive variant calls, additional filtering was performed. An SNV was considered a true positive if the variant allele frequency (VAF) was greater than 2% and the mutation was called by both VarScan2, with a somatic p-value <=0.01, and MuTect. Alternatively, a frequency of 5% was required if only called in VarScan2, again with a somatic p-value <=0.01. Additionally, sequencing depth in each region was required to be >=30 and >=5 sequence reads had to support the variant call. In contrast, the number of reads supporting the variant in the germline data had to be <5 and the VAF <=1%. In addition to these sample specific measures, we also utilized the cohort to reduce SNP contamination through two independent means. Firstly, all variants designated as 'germline' by VarScan2, from all regions

in all 100 patients within the cohort, were combined so that every germline variant detected in the cohort had an associated TRACERx population frequency. SNVs were filtered if they were found to have >1% frequency in the TRACERx cohort. Finally, in an effort to reduce the impact of direct sample to sample contamination, the SNVs from every patient were compared against the germline SNPs in every other patient independently. If >5% SNVs were identified as SNPs in another patient, the sample was flagged as contaminated and any such variant that matched a SNP was removed from further analysis. Finally, a blacklist filter, relating to the genomic location of the variant, was applied. The blacklisted genomic regions were obtained from UCSC Genome Table Browser[13] and include regions excluded from the Encode project (both DAC and Duke list), simple repeats, segmental duplications and microsatellite regions.

Additionally M-seq provides the opportunity to increase the sensitivity to detect low frequency mutations. By sharing the independently called mutations across the multiple regions and reassessing the reads at each position for each tumor region, it is possible to call more mutations and reduce the possibility of over-representing the mutational heterogeneity. Where a somatic variant was not called ubiquitously across tumor regions but was called in one or more region, read information was extracted from the original alignment file using bam-readcount (0.5.1) (https://github.com/genome/bam-readcount). In such cases, VAF restrictions were reduced to VAF >=1% allowing for the positive identification of low frequency variants that would otherwise have been missed.

INDELS were filtered using the same parameters as described above, with the exception of the requirement of >=10 reads supporting the variant call, a somatic p-value of <=0.001 and a sequencing depth of >=50. Occasionally, when attempting to identify INDELS across multiple tumor regions, discrepancies in the start position, end position or length of the INDEL were identified. In such cases, the longest predicted INDEL was reported and the maximum sequence related values reported.

Dinucleotide substitutions (dinucleotide variants, DNV) were identified in cases where two adjacent SNVs were called. In such cases, a proportion test was performed to provide an indication as to whether the frequency of the two SNVs was significantly similar and thereby indicative of a single mutational event. In such cases, the start and stop position was corrected to represent a dinucleotide substitution and sequence related values were recalculated to represent the mean of the SNVs.

Mutations were classified as clonal or subclonal using PyClone[14](see below). In cases where it was not feasible to use PyClone, mutations were classified based on presence or absence. Those mutations found ubiquitously in all tumor regions were classified as clonal. Heterogeneous mutations were classed as subclonal. To ascertain the likelihood that a heterogeneous mutation was absent from a region rather than missed due to a lack of sampling depth, we determined the maximum variant allele frequency compatible with the data, as described by Yates et al[15].

Variants were annotated using Annovar[16] and COSMIC (v75) (cancer.sanger.ac.uk)[17].

All mutations used in analysis can be found in Table S3 in the Supplementary Appendix.

Identification and classification of driver mutations

All non-silent variants were compared against a list of potential driver genes (n=641). The driver gene list was comprised of all genes identified in the COSMIC cancer gene census (v75)[17], supplemented with those identified in large-scale pan-cancer analyses (using $q < 0.05$ as cut-off)[18], and previous large-scale non-small cell lung cancer sequencing studies[19-21]. Any

variants that were located within one of these genes underwent categorization based on pre-set criteria. If the gene was annotated as being recessive by COSMIC (tumor suppressor), and the variant was deemed to be deleterious (either a stop-gain or predicted deleterious in two of the three computational approaches applied – Sift[22], Polyphen[23] and MutationTaster[24], then the specific variant would be classed as a driver mutation. Alternatively, if the variant was found in a gene annotated by COSMIC as dominant (oncogene), then we sought to identify exact matches to the specific variant in COSMIC. If an exact match was found ≥3 times, the variant was classed as a driver mutation.

Variant Call Validation

In order to determine the accuracy of the somatic variant calls, the previously described validation sequence datasets (mPCR-NGS and AmpliSeq) were utilized. Firstly in order to calculate an error profile for the validation data, the frequency of non-reference non-variant base calls was determined for each mutation. This provided a median sequence error rate <1%. The variant allele frequency of each mutation submitted for validation was calculated for each tumor region to enable comparison to the exome data. Two questions were posed. Firstly does the exome call from a tumor region match the validation data for that region? Secondly, does the validation data agree with the variant being classified as somatic rather than germline? A call was classified as being a false positive if the exome VAF was >1% whilst the upper 95th confidence interval from a binomial test of the validation data, providing variant read count and total depth as input, was <1%. A false negative was called if the exome VAF <1% whilst the lower 95th confidence interval was >1%. Finally a false positive somatic call was declared if a binomial test of the associated germline validation data provided a lower 95th confidence interval >1% VAF. This methodology was used on both the mPCR-NGS and the AmpliSeq validation datasets, results of which can be found in Table S4 in the Supplementary Appendix. Notably, only 3/ 2442 (0.12%) clonal and 16/5573 (0.29%) subclonal SNVs and indels were classified as false positives, while only 1/2442 (0.04%) clonal and 33/5573 (0.59%) subclonal SNVs and indels were classified as false negatives.

Additionally a strong relationship was seen between the VAF determined from the exome-sequencing and the VAF calculated from the validation sequencing datasets (Pearson correlation co-efficient for exome SNV vs mPCR-NGS = 0.96, exome SNV vs AmpliSeq = 0.95, exome INDEL vs AmpliSeq = 0.87, Fig.S2 and Table S4 in the Supplementary Appendix)

Identification and verification of ALK, ROS1 and RET status using whole-exome sequencing

Due to the extension of the Agilent Human All Exome V5 kit to cover intronic regions in genomic areas linked with fusion events in NSCLC (ALK, EML4, FGFR1, FGFR2, FGFR3, KIF5B, NTRK1, RET, ROS1, TCF3), it was possible to attempt to identify fusions in these areas.

Sequence reads were filtered to produce a file containing only discordant or chimeric reads. Discordant reads were obtained by selecting mapped read pairs with an insert size greater than 10000. Reads annotated with a secondary alignment were selected from the alignment file to represent chimeric DNA. The chimeric reads were merged with the discordant read set to create a single alignment file. This was performed independently for all tumor regions plus the germline sample.

The depth of the filtered fusion reads was counted in 1000bp windows across the genes of interest using bedtools (version 2.25)[25]. The counts from the tumor regions were summed and normalized. Normalized counts were also derived from the germline sample and were subtracted from the tumor counts. To ascertain whether signal being detected was significantly higher than expected by chance, a random set of 1000 genes was selected and fusion read depth calculated as described above. 10,000 permutations were performed on this gene set, with the same number of windows selected as in the tumor set. The maximum fusion

score for each permutation was obtained and used to generate an empirical p-value for the scores in the fusion list. Significant fusions ($p<0.01$) were recorded.

Reference genomic DNA for ALK, ROS1 and RET fusion events derived form cell lines were purchased from Horizon Diagnostics (catalogue numbers HD664 for EML4/ALK, HD-C105 for SLC34A2/ROS1 and HD-C108 for RET/CCDC6). The DNA was sequenced and analyzed as described above. Significant fusion events were identified covering all 3 of the known events in 5 of the 6 genes (Fig.S28 in the Supplementary Appendix).

TCGA exome data sets

Tumor and matched germline exome sequencing BAM files for both lung adenocarcinoma (LUAD, n = 397) and lung squamous cell carcinoma (LUSC, n = 350), were obtained from the Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/) via https://cghub.ucsc.edu. The BAM files were converted to FASTQ using bedtools bamtofastq (v2.25)[25]. The resulting sequence files were processed as described for the TRACERx cohort, with a set of annotated variants being generated as the output. Clinical data for the TCGA patients was accessed through http://www.cbioportal.org using the R package 'cgdsr'[26].

Copy number analysis

Processed sample exome copy number data from paired tumor-normal was generated using VarScan2 (v2.3.6). Varscan2 copynumber was run using default parameters except min-coverage = 8, min-segment-size = 50. The data-ratio parameter was calculated on a per-sample basis as described by Koboldt and colleagues[11]. The Varscan2 copynumber produced per region LogR values. LogR values were subsequently GC corrected using a wave-pattern GC correction method based work by Cheng and colleagues[27].

Homozygous and heterozygous single nucleotide polymorphisms (SNPs), were called in the germline sample using Platypus v0.8.1[28] with default parameters apart from the genIndels flag set to FALSE. Tumor regions from the same patient were then genotyped based on the variants identified in the germline. Only SNPs with a minimum coverage of 20x in the germline and all tumor regions from the same patient were taken forward for copy number analysis. The B-allele frequency (BAF) of each SNP was calculated as the proportion of reads at that position that contained the reference base versus the variant. All SNPs were filtered by a list of previously classified poor quality SNPs constructed from the TRACERx100 germline samples. A SNP was classified as poor quality if it was found in at least 20 samples, and the number of times a given SNP showed a BAF value in the germline that was between 0.1—0.32 or 0.68 – 0.9 was higher than the number of times the same SNP showed a BAF of either <0.1 or > 0.9, multiplied by 0.13469. For SNPs that had undergone allelic imbalance, it was possible to track if the parental origin of the major allele relative to the minor allele swapped between different tumor regions, indicating mirrored subclonal allelic imbalance events as described in detail below.

LogR and BAF values for each tumor region were processed with ASCAT v2.3[29] using default parameters except "gamma" set to 1, in order to provide segmented allele-specific copy number data plus cellularity and ploidy estimates for all samples. Manual verification was performed of the automatically selected models for ploidy and cellularity using an orthogonal measure of tumor cellularity based on mutation variant allele fraction, as described below. Per region copy number data called by ASCAT is available in Table S10 in the Supplementary Appendix. Floating point copy number values were used for all copy number analysis.

Gene-level amplification was called by mean gene copy number >= 2x ploidy + 1 copy. Gene level deletion was called by mean gene copy number < 0.5 copies. For both amplification and deletion events, the nature of the relatively low resolution of the exome sequence data (restricted to exonic regions) makes it difficult to call very focal events. This was in particular apparent with regard to deep CDKN2A deletions, which are likely under called in the TRACERx data. To determine intratumor heterogeneity (ITH) status of amplified and deleted genes, we called amplifications and deletions across all regions from each tumor. If at least one region showed an amplified or deleted mean copy number, we called a gene as clonally amplified if all other regions showed a copy number gain of ploidy +1 copy for amplification or a copy number loss of ploidy –1 copy for deletion. The gene would be called as subclonally amplified or deleted if at least one region showed no gain or loss, or if a clonal event overlapped with mirrored subclonal allelic imbalance. In this manner, by allowing a lower threshold across regions if at least one region show high amplification/deep deletion, ITH status is called conservatively, biased towards clonal. In order to associate amplifications and deletions called as subclonal with specific mutation clusters from PyClone (see below), we first identified all clusters present at >= 50% CCF in each region. We then identified all clusters present in the same regions as a given somatic copy number aberration (SCNA), and associated the SCNA with the cluster closest to the trunk. If no clusters matched a given SCNA, we re-called the SCNA but only calling the amplification or deletion using the stricter threshold (amp: 2x ploidy + 1, del: 0.5 copies), and repeated the association test above. If an SCNA could still not be associated with a mutation cluster, it was annotated as subclonal with no known cluster associated.

To determine genome-wide copy number gain and loss, copy number data for each sample was divided by the sample mean ploidy, and log2 transformed. Gain and loss were defined as log2(2.5/2) and log2(1.5/2), respectively. ITH status of chromosomal arm gain and loss were defined on a per tumor basis, by requiring at least one region to show at least 98% gain or loss. Clonal arm gain or loss was then called if the same chromosomal arm showed at least 75% gain or loss across all remaining regions. Subclonal arm gain or loss was called if at least one region showed less than 75% gain or loss of the chromosomal arm or if a chromosomal arm was subjected to mirrored subclonal allelic imbalance. To determine global SCNA ITH, we defined gain and loss as described above. All parts of the genome were considered independently and split into minimum consecutive segments of overlap within each tumor across all regions. Any segment of gain or loss that overlapped across all regions was defined as clonal and all other segments of copy number aberrations as subclonal. Within each tumor, we then summarized the percent of the genome subjected to SCNA in any region (total SCNA), the percent of the genome subjected to clonal SCNA (SCNA shared by all tumor regions), and the percent of the genome subjected to subclonal SCNA (SCNA found in some but not all tumor regions). The proportion subclonal copy number alterations was then defined as the percent of the genome subjected to subclonal SCNA divided by the percent of the genome subjected to SCNA in any region (total SCNA). To determine a static measure of SCNA as shown in Fig.10 in the Supplementary Appendix, for each tumor we determined the total amount of SCNA in each tumor region, then calculated the mean across all regions.

Orthogonal validation of copy number tumor cellularity estimates

To assess (and ensure) putative copy number solutions were compatible with mutation variant allele frequencies (VAFs) a measure of tumor cellularity based directly on the mutation VAFs was devised. We reasoned that the majority of ubiquitous mutations (identified in all tumor regions) are likely clonal, and, as such, their VAFs (after correcting for copy number) should reflect the tumor purity. Therefore, after correcting for the copy number state of mutations (see below), we determined a mutation based purity estimate based on the copy number adjusted VAF distribution.

For each sample, we assessed whether the tumor estimate derived from copy number analysis was consistent with our VAF estimate. For any sample where the purity estimates were not concordant, we considered whether an alternative copy number solution might provide a better fit of the data.

Removing samples of low tumor cellularity

The median number of mutations called independently in each tumor region was determined. A region was flagged for manual inspection, and possibly removed from further analysis, if less than 30% of the median mutation count was identified. This process was repeated but using all mutations called, therefore including those obtained by sharing mutational information across regions. Further sample quality control occurred following the orthogonal validation of copy number tumor cellularity estimates. If the purity estimate from copy number analysis was greater than 1 standard deviation away from the mean of the orthogonal VAF estimate, the sample was manually assessed. If no concordant solution was found, the sample was removed from further analysis.

Subclonal deconstruction

In order to estimate whether mutations were clonal or subclonal, and the clonal structure of each tumor, a modified version of PyClone was used. For each mutation, two values were calculated, obsCCF and phyloCCF. obsCCF corresponds to the observed cancer cell fraction (CCF) of each mutation. Conversely, phyloCCF corresponds to the phylogenetic CCF of a mutation. To clarify the difference between these two values, consider a mutation present in every cancer cell within a tumor. A subclonal copy number event in one tumor region may lead to loss of this mutation in a subset of cancer cells. While, the obsCCF of this mutation is therefore below 1, from a phylogenetic perspective the mutation can be considered 'clonal' as it occurred on the trunk of the tumor's phylogenetic tree, and, as such, the phyloCCF may be 1.

To calculate the obsCCF of each mutation, local copy number (obtained from ASCAT), tumor purity (also obtained from ASCAT), and variant allele frequency were integrated. In brief, for a given mutation we first calculated the observed mutation copy number, n_{mut}, describing the fraction of tumor cells carrying a given mutation multiplied by the number of chromosomal copies at that locus using the following formula:

$$n_{mut} = VAF \frac{1}{p}[pCN_t + CN_n(1-p)]$$

where *VAF* corresponds to the variant allele frequency at the mutated base, and p, CN_t, CN_n are respectively the tumor purity, the tumor locus specific copy number, and the normal locus specific copy number (CN_n was assumed to be 2 for autosomal chromosomes). We then calculated the expected mutation copy number, n_{chr}, using the *VAF* and assigning a mutation to one of the possible local copy numbers states using maximum likelihood. In this case only the integer copy numbers were considered.

All mutations were then clustered using the PyClone Dirichlet process clustering[14]. For each mutation, the observed variant count was used and reference count was set such that the VAF was equal to half the pre-clustering CCF. Given that copy number and purity had already been corrected, we set the major allele copy numbers to 2 and minor allele copy numbers to 0 and purity to 0.5; allowing clustering to simply group clonal and subclonal mutations based on their pre-clustering CCF estimates. We ran PyClone with 10,000 iterations and a burn-in of 1000, and default parameters, with the exception of --var_prior set to 'BB' and –ref_prior set to 'normal'.

To determine the phyloCCF of each mutation, a similar procedure to that described above was implemented, with the exception that mutations were corrected for subclonal copy number events. Specifically, if the observed variant allele frequency was significantly different from that expected (P<0.01, using prop.test in R) given a clonal mutation, we determined whether a subclonal copy number event could result in a non-significant (P>0.01) difference between observed and expected VAFs. The pre-clustering CCF for each mutation was then calculated by dividing n_{mut} by n_{chr}. Subclonal copy number events were estimated using the raw values from ASCAT output.

Finally, to ensure potentially unreliable VAFs of indels did not lead to separate mutation clusters, each estimated indel CCF was multiplied by a region specific correction factor. Assuming the majority of ubiquitous mutations, present in all regions, are clonal, the region specific correction factor was calculated by dividing the median mutation CCF of ubiquitous mutations by the median indel CCF of ubiquitous indels.

Detection of mirrored subclonal allelic imbalance

Breakpoints from each individual tumor sample's ASCAT segmentation profile were combined to create a single patient specific consensus segmentation profile. This profile was then applied to all tumor samples while maintaining their original integer allele specific copy number calls at all genomic positions. For each consensus segment with allelic imbalance in at least one region, the B-allele fraction (BAF) was then determined for all SNPs as the ratio of the minor allele to the total allele count. If a given copy number event is shared between regions, then the minor allele must always be from the same parental allele across tumor regions (e.g., if a SNP A/T is represented by A/TT in a triploid context, it must be A/TT across regions, never AA/T). By keeping track of the individual alleles, it is possible to determine if the minor allele is different between regions, indicating two independent copy number events. Thus mirrored subclonal allelic imbalance detection involves comparing the BAF of the same heterozygous SNPs across multiple samples and testing whether the BAF values always follow the same distribution or if their positions (i.e. high and low frequency) are reversed (mirrored).

Mirrored subclonal allelic imbalance can only be detected where allelic imbalance is present in a minimum of two samples. One sample is chosen as the reference for all consensus segments in each chromosome. The reference sample is that with the greatest amount of total allelic imbalance across the entire chromosome, and, which also has allelic imbalance at the genomic position of the specific segment being investigated.

BAFs of heterozygous SNPs present in each segment for each sample were compared to the BAFs of the reference sample. Briefly, two log-truncated normal distributions were created with their μ set as the theoretical BAF of SNPs positioned on the chromosome corresponding to the ASCAT derived major CN and minor CN from that sample. The SNPs were then assigned a log likelihood of belonging to either the major or minor CN equivalent chromosome using these distributions. These values were then compared and the probabilities of a reversal (or mirroring) having occurred for each SNP's BAF are calculated. Finally, these were examined using a one tailed t-test to determine whether the log likelihood ratio values of these SNP BAFs constituting a mirrored subclonal allelic imbalance event was significantly different from zero.

Mirrored subclonal allelic imbalance arm gains and losses

Arm gains and losses were called as described in the Copy Number Analysis section. A mirrored subclonal allelic imbalance arm gain or loss was defined as when the opposite parental alleles were affected in at least 75% of a chromosome arm in a minimum of two regions.

These results were then compared to significant arm gains and losses from the previous TCGA GISTIC2 analyses (Broad Institute TCGA Genome Data Analysis Center (2016): SNP6 Copy number analysis (GISTIC2). Broad Institute of MIT and Harvard. doi:10.7908/C1348JSB) for the lung adenocarcinoma and lung squamous cell carcinoma histological subtypes.

Mirrored subclonal allelic imbalance amplifications

Amplification was called as described in the Copy Number Analysis section. A mirrored subclonal allelic imbalance amplification was defined as an amplification meeting this definition as well as meeting the criteria for detection of mirrored subclonal allelic imbalance in at least two regions from the same tumor.

Comparing mirrored subclonal allelic imbalance breakpoint

Mirrored subclonal allelic imbalance events describe genomic segments of copy number change that shows inverse ratios of parental alleles between different tumor regions. As mirrored subclonal allelic imbalance spans a single genomic region, it may have occurred as a single event, with a cancer cell undergoing genomic rearrangement resulting in two daughter cells with different parental allele loading at a given segment. If so, genomic breakpoint locations between tumor regions showing mirrored subclonal allelic imbalance at a given location should be close to identical. To test this hypothesis, the minimum distance between genomic breakpoints for clonal copy number gains and losses were determined, providing a control group representing single event copy number changes. This distribution was compared to the minimum distance of mirrored subclonal allelic imbalance breakpoints when comparing the two groups of tumor regions showing inverse parental allele ratios (Fig.S15 in the Supplementary Appendix).

Timing of mutations

Individual mutations (SNVs, DNVs and indels) were timed as clonal or subclonal through PyClone clustering as described above. For clonal timing, the values from phyloCCF were used. Clonal mutations were then further timed as early, late or untimed clonal. This was performed first within each tumor region relative to overlapping copy number events as previously described[30]. In brief, mutations in parts of the genome with at least two copies of the major allele were preliminarily classified as early if the mutation copy number (described above) was > 1, and late if it was <= 1. Temporal analysis was then performed across all regions. Mutations classified as clonal were called as “clonal early” if a given mutation was early across all regions, or, if the mutation was early in the majority of regions. A clonal mutation was called as “clonal late” if it was called as late across all regions, or if it was late across the majority of regions. Any clonal mutations that could not be timed as either early or late, were classified as “clonal untimed”.

Timing of chromosomal arm somatic copy number aberrations

Whole chromosomal arm gain and loss were called as clonal and subclonal as described above. Clonal chromosomal arm gain was further timed by the mutation copy number of all mutations on a given arm. All mutations mapping to the arm gain were identified, and the across-region average ratio of mutations with a mutation copy number within one copy of the major allele was determined. If the majority of mutations showed a copy number within one copy of the major allele, indicating that the mutations occurred prior to increase in copy number and were gained with the chromosomal arm, the chromosomal arm gain was called as “clonal late”. If the majority of mutations showed a lower copy number, indicating that the mutations occurred after chromosomal arm gain, the copy number gain was called as “clonal early.” If there were not enough mutations (minimum 5) mapping to regions of chromosomal arm gain, the chromosomal arm gain was called as “clonal untimed”.

For clonal chromosomal arm loss, we timed the losses relative to loss of heterozygosity (LOH) status in genome doubled tumors. Any chromosomal arm loss showing LOH in a genome doubled context is likely to have occurred prior to genome doubling as it would otherwise require two independent hits. Thus, chromosomal arm loss in a genome doubled context showing LOH were called as "clonal early", chromosomal arm loss in a genome doubled context without LOH were called as "clonal late", and chromosomal arm loss not in a genome doubled context were called as "clonal untimed".

Identification of subclonal mutations driven by copy number loss

Mutations were investigated in order to identify those whose absence, or low CCF values, may be driven by copy number loss events. For each tumor we identified any SNV residing in genomic segments of copy number heterogeneity across tumor regions, with minor and major copy number aberrations considered separately. For each chromosome, we grouped mutations into non-contiguous genomic segments with consistent copy number states within tumor regions and within SNV clusters defined above. To restrict our analysis to mutations lost in at least one tumor region, we determined the median CCF value of each SNV group, and only considered SNV groups where the median CCF value was <=0.25 in at least one tumor region. We then evaluated whether copy number loss coincided with lower CCF levels using a one-sided Wilcoxon test or, if more than two copy number states were present across tumor regions, a one-sided Cochrane Armitage trend test. To ensure the lower CCF value was driven by copy number and not tumor region, we also implemented a regression analysis, including both copy number and region in the model. If over 85% of mutations within a given pyclone cluster were determined to be driven by copy number, then the entire cluster was classified as copy number driven. Finally, to avoid overestimating copy number driven losses of mutations, only losses occurring in <=75% of tumor regions were considered.

In addition, comparisons were made between the results of each mutations obsCCF and phyloCCF. Given that the only difference between the calculation of the two is that obsCCF does not correct for subclonal copy number events, mutations which appear clonal by phyloCCF but subclonal by obsCFF may reflect copy number driven heterogeneity. To avoid overestimating copy number driven heterogeneity only mutations with a VAF of at least 1% were considered potentially to reside on a subclonal copy number and thereby considered as potentially driven by subclonal copy number loss.

Phylogenetic tree construction

Trees were constructed using the published tool CITUP (0.1.0)[31], using the command, 'run_citup_qip.py', and limiting the maximum and minimum number of clusters to those identified with PyClone phylo (see above). As input, CITUP requires mutation clusters, as well as their mean cancer cell prevalence values. Only tumors with phyloCCF Pyclone output from at least two tumor regions and containing at least two mutation clusters were included in the analysis (292 tumor regions from 91 tumors). To ensure accurate tree construction, mutation clusters were first filtered to ensure no violation of evolutionary principles. In brief, two principles were considered. First, the pigeonhole principle, which states that two mutation clusters cannot be considered independent and on separate branches of an evolutionary tree if the sum of the cancer cell prevalence values of the two clusters exceeds 100% within a single tumor region. Second, a descendent clone must exhibit a smaller cellular prevalence than its ancestor within each and every tumor region, referred to as the 'crossing rule'. Using these principles it can be determined whether particular mutation clusters conflict with each other and cannot be fitted to the same evolutionary tree. For instance, if the cellular prevalence of mutation cluster 1 in tumor region 1 is 80% and mean cellular prevalence of mutation cluster 2 in tumor region 1 is 60%, then, by the pigeonhole principle, cluster 2 must be a descendent of

cluster 1. However, if, in a different tumor region, the cellular prevalence of cluster 2 is greater than cluster 1, it can be said that cluster 1 and 2 conflict due to 'crossing rule'.

To ensure accurate tree construction, only clusters with at least 5 mutations were included. For the majority of tumors, all subsequent clusters were used as input to CITUP. However, for a subset of tumors evolutionary conflicts were identified prior to running CITUP, and a small number of mutation clusters were therefore removed. In total, 155/680 mutation clusters were removed, 89 of these containing less than 5 mutations. This removed 818 mutations from a total 59,440 mutations, representing less than 1.4% of all clustered mutations. For six tumors, manual tree construction was required. For tumors CRUK0032, CRUK0062 and CRUK0065, this was due to the number of clusters exceeding the phylogenetic space that was practically possible for CITUP to exhaustively explore. For CRUK0004, CRUK0017 and CRUK0069, erroneous copy number correction of certain mutation clusters in certain regions led to inflated CCF values and phylogenetic trees that would breach evolutionary principles. For these samples, manual trees were generated by adjusting the cluster CCF values. In cases where multiple evolutionary trees were deemed equally likely by CITUP (as measured by BIC score), every tree was reported in Fig.S12 in the Supplementary Appendix. Detailed description on every tree including notes on manual review can be found in Table.S7 in the Supplementary Appendix.

Mutational Signature Analysis

Mutational signatures were estimated using the deconstructSigs package in R[32]. Signature 1A, 2, 4, 5, 13 were considered. For temporally dissected mutational signatures, mutational signature analysis was only applied if at least 15 mutations were present. Similarly, subclone specific mutational signature analysis was restricted to subclones with at least 15 mutations. Given that not all mutations can be described as an amalgamation of previously identified signatures using deconstructSig[32], a small proportion of mutations for each tumor were necessarily classified as 'other'.

dN/dS analysis

To estimate dN/dS values for missense mutations (ωmis) and nonsense mutations (ωnon) exome-wide and for a cohort of cancer genes, we used an adaptation of a previously described method[33]. In brief, as outlined by Martincorena et al[34] and Hodgkinson et al[35], to obtain accurate estimates of dN/dS, a model accounting for any context dependence effect by 1-nucleotide upstream and downstream using 192 substitution rates was used. Further, two selection parameters were incorporated to measure the observed-over-expected ratio of missense (ωmis) and nonsense (ωnon) mutations.

Maximum-likelihood estimates and confidence intervals for the 194 parameters were obtained using Poisson regression. To obtain accurate estimates of the 192 substitution rate parameters, they were estimated from the exome-wide set of temporally and clonally dissected mutations. These 192 rates were then assumed constant across all genes and mutation set maximum-likelihood estimates for ωmis and ωnon were obtained for the combined set of mutations residing in cancer genes. Maximum likelihood estimation was performed using the bbmle package in R, function mle2.

Stratifying TRACERx patients according to NLMT and MATCH clinical trials

At the time of the submission of this manuscript, two ongoing clinical trials, UK based NLMT and US based MATCH, were testing genomic-based stratification to select patients for targeted therapies. One of the recruitment criteria for NLMT and MATCH were that patients must have failed standard of care, and as such no TRACERx patient would formally be eligible. However,

TRACERx data can be used to assess to what degree genomic stratification of lung cancer patients for targeted therapies may be based on subclonal events, given the stratification criteria. Genomic stratification criteria of NLMT was acquired from Middleton et al[36]. Genomic stratification criteria for MATCH was acquired from https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match [2016-12-19]. These were combined into Table.S11 in the Supplementary Appendix, which was used to stratify TRACERx tumors. The NLMT criteria are subtype-specific, however, when both LUAD and LUSC were tested for a specific aberration, we expanded this to include all NSCLC subtypes (thereby including the "other" NSCLC subtypes). To call an aberration, we used our definition for driver mutation, deletion, and amplification as described above with the exception that a loss event could also be called if a specific gene harbored a driver mutation and a loss resulting in no remaining wild type copies. Fusion events were not included.

Statistical analysis

All analysis was performed in the R statistical environment version >=3.2.1. All statistical tests were two-sided unless expressly stated. All heterogeneity measures were defined without prior knowledge of patient outcome. Survival analysis was performed using the Kaplan-Meier method based on the median fraction of either subclonal mutation or subclonal SCNA, p-value determined by a log-rank test. Recurrence-free survival was defined as time to recurrence or relapse, or if a patient died without a recurrence, time to death. Hazard ratio was determined through a Cox proportional hazards model. Multivariate Cox regression was performed with recurrence-free survival versus subclonal SNV or subclonal SCNA above the median, with stage, age, pack-years, histological subtype and whether they had adjuvant therapy or not included in the model.

Data availability

The sequencing data has been deposited at the European Genome-phenome Archive (EGA)[37], which is hosted by the EBI and the CRG, under accession number EGAS00001002247.

Further information about EGA can be found on https://ega-archive.org (https://ega-archive.org/).

Supplementary figures
Figure S1

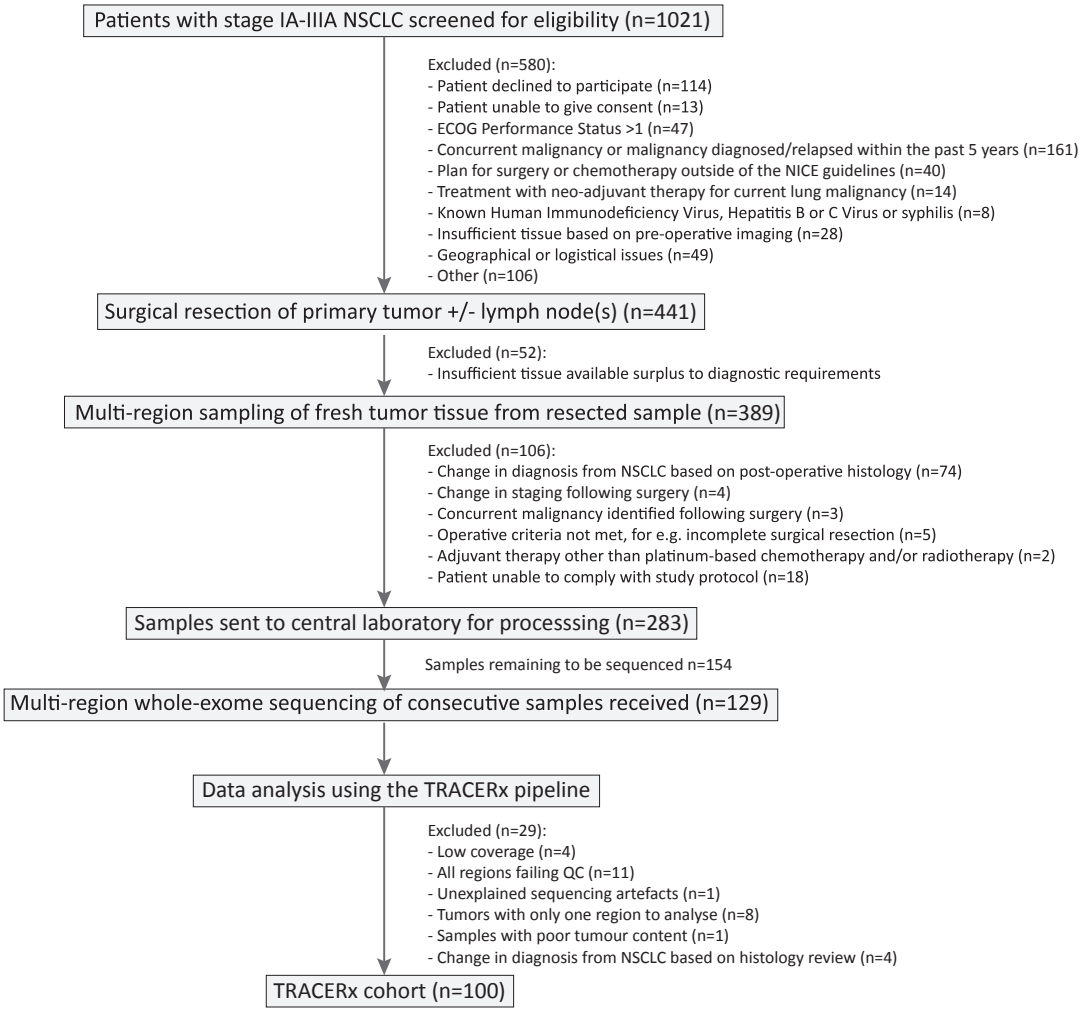


Figure S1. CONSORT diagram for patient recruitment into the TRACERx study and eventual selection of the 100 patient cohort.

Figure S2

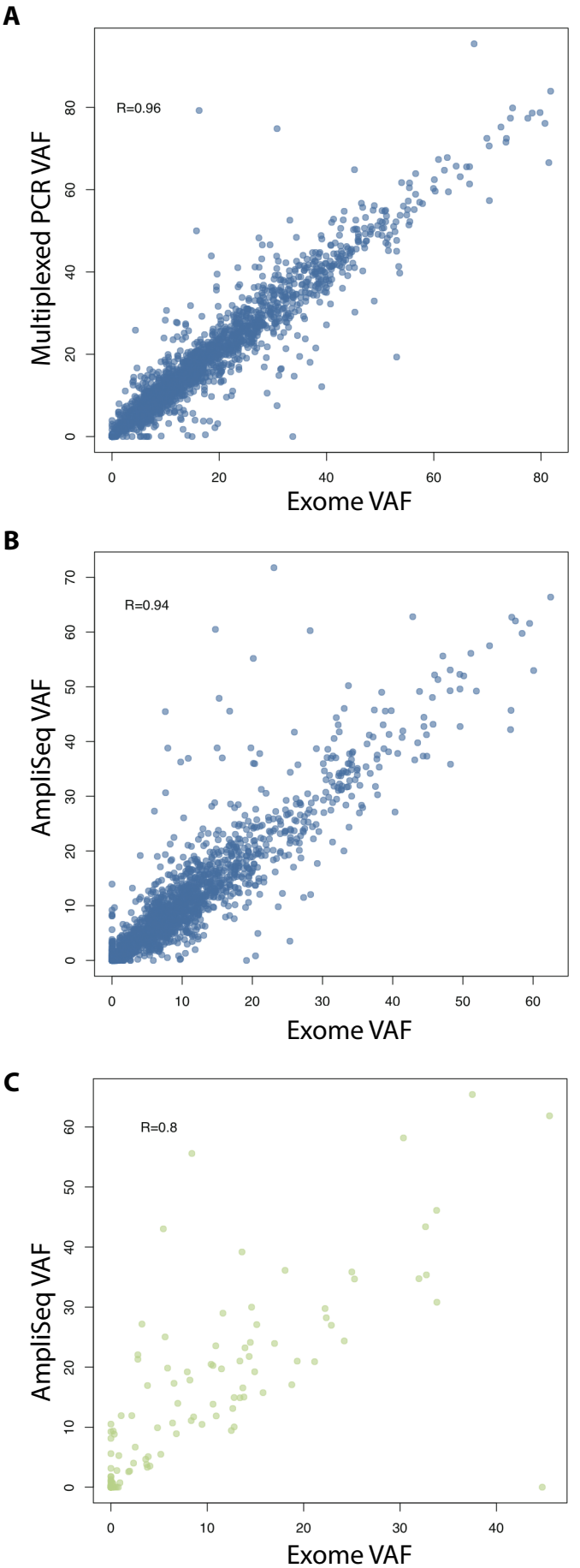


Figure S2. Orthogonal validation
Correlation of variant allele frequencies obtained from exome sequencing and orthogonal validation by A) multiplexed PCR and subsequent illumina sequencing of SNVs, B) sequencing of SNVs on the Ion Torrent PGM sequencer and C) sequencing of INDELs on the Ion Torrent PGM sequencer.

Figure S3

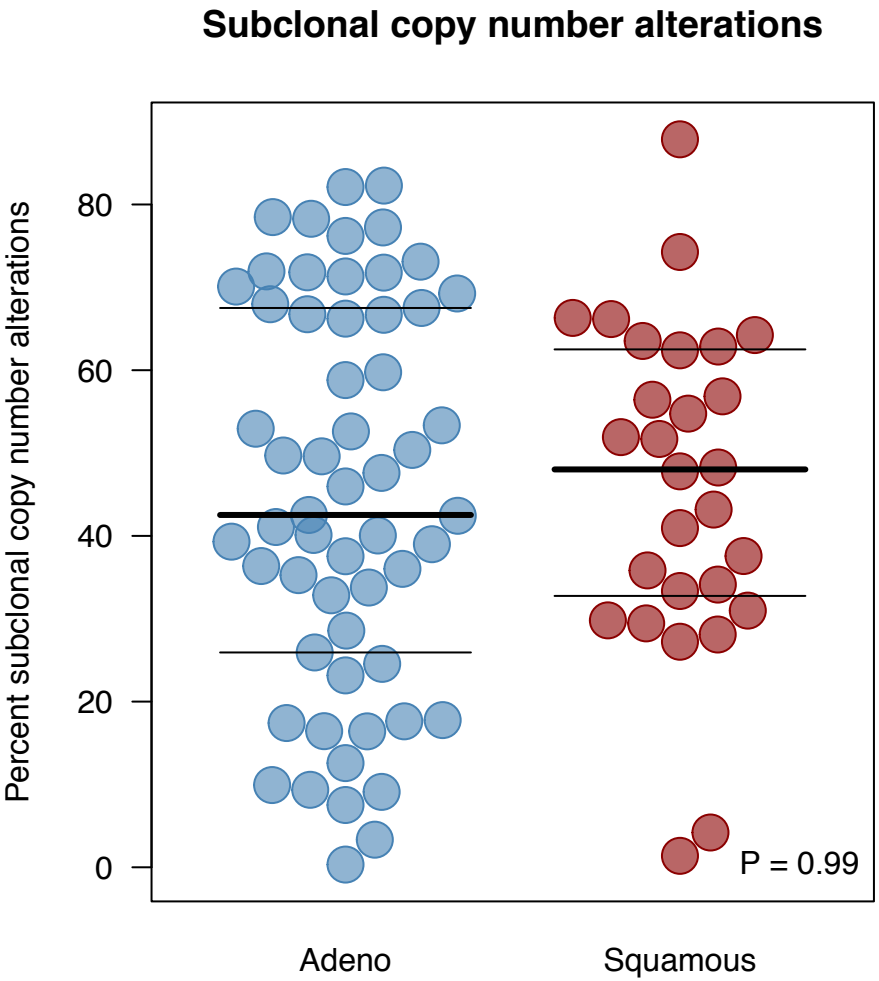


Figure S3. Percent of aberrant cancer genome subject to subclonal copy number alteration in adenocarcinoma and in squamous cell carinoma.

Median all samples, 48% (range 0.3-88%). Median adenocarcinoma, 43% (range 0.3-82%). Median squamous cell carcinoma 48% (range 1.4-88%).

Figure S4

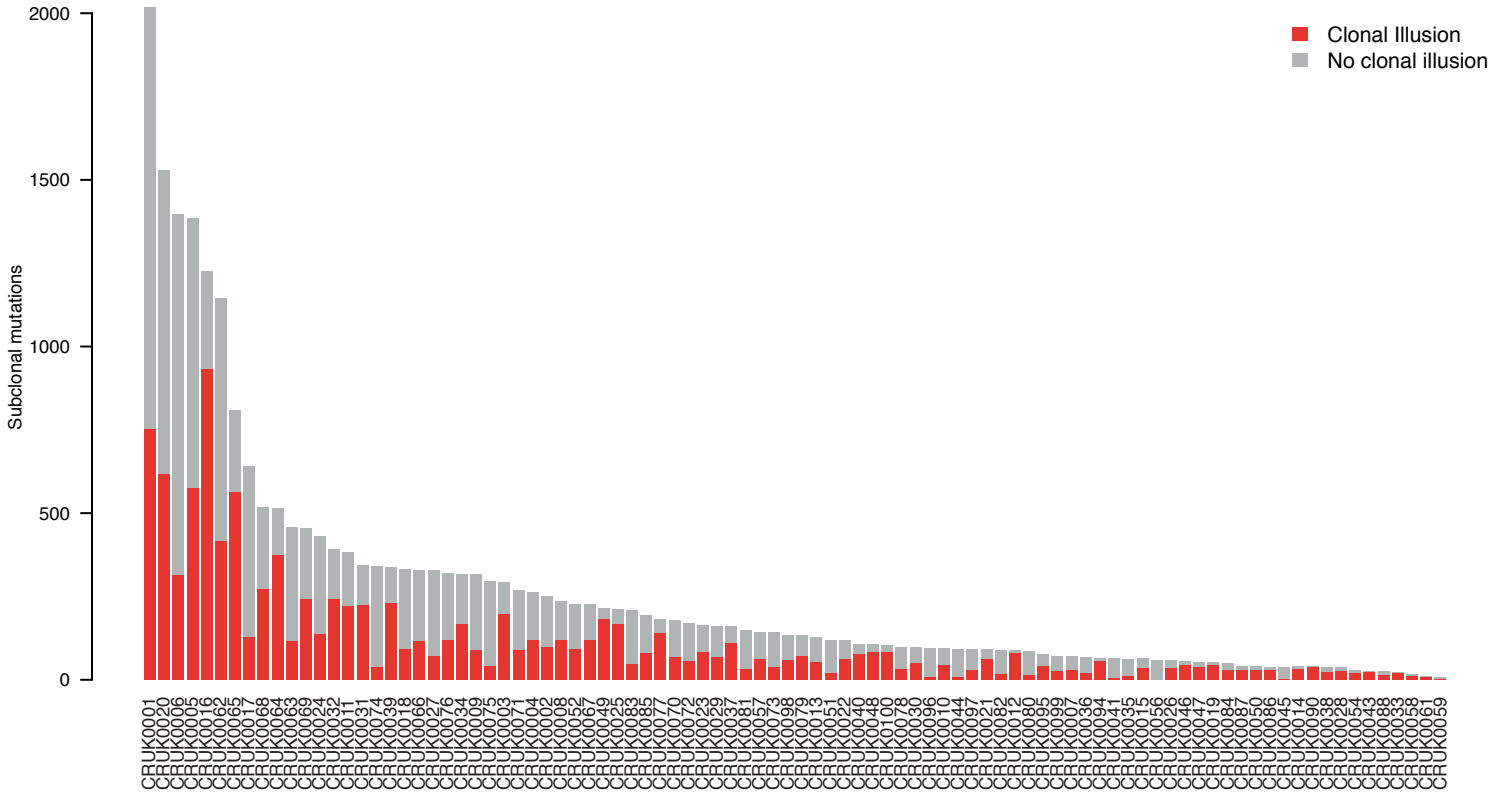


Figure S4. Illusion of clonality
Barplot depicts all subclonal mutations, and whether these exhibit an illusion of clonality; whereby the mutation may appear clonal in at least one tumor region.

Figure S5

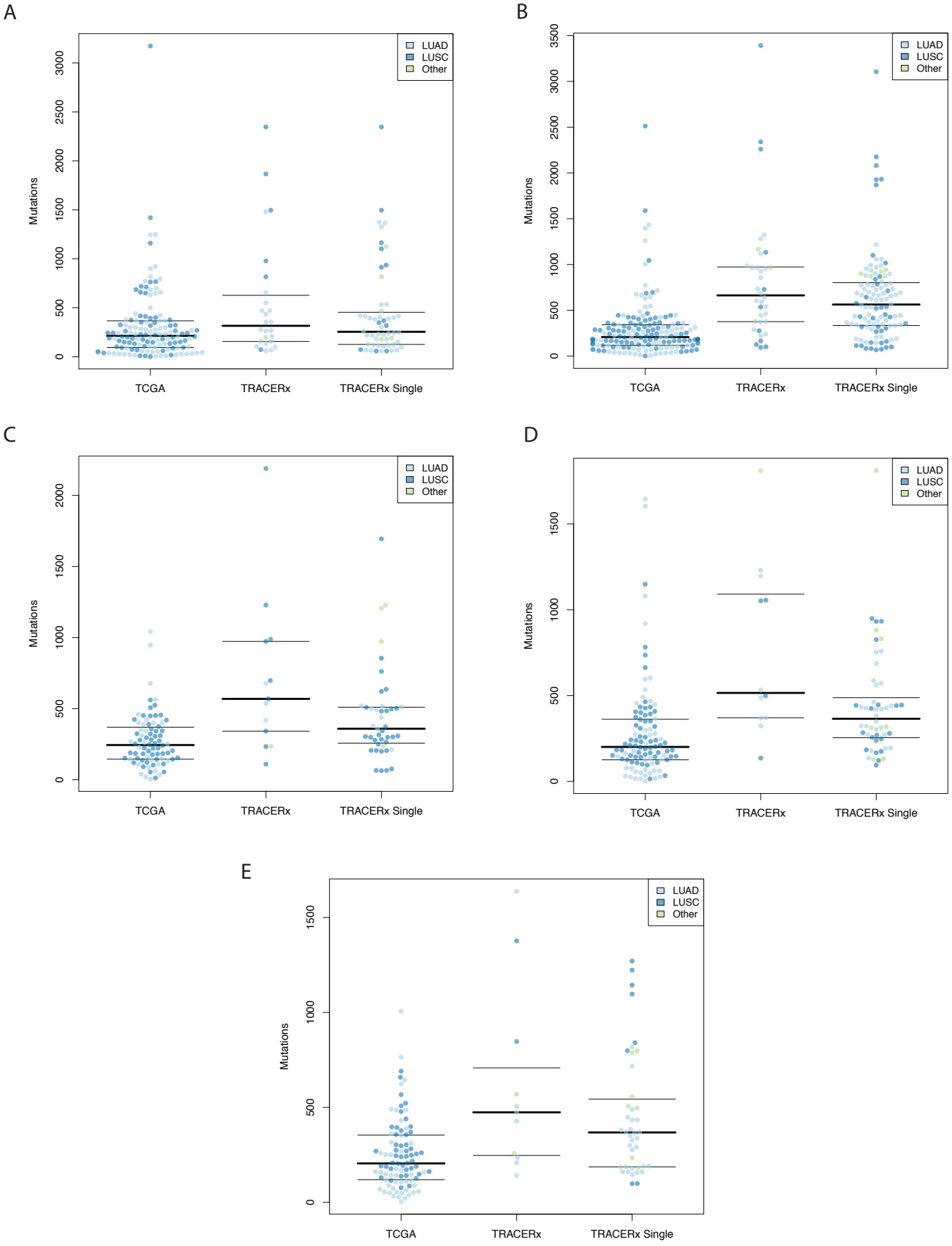


Figure S5. Comparison of TCGA and TRACERx mutation numbers
Mutation count in TCGA adenocarcinoma and squamous cell carcinoma, TRACERx multi-region and TRACERx analyzed as single samples in A) Stage 1A tumors, B) Stage 1B tumors, C) Stage 2A tumors, D) Stage 2B tumors and E) Stage 3A tumors. LUAD: Lung adenocarcinoma. LUSC: Lung squamous cell carcinoma.

Figure S6

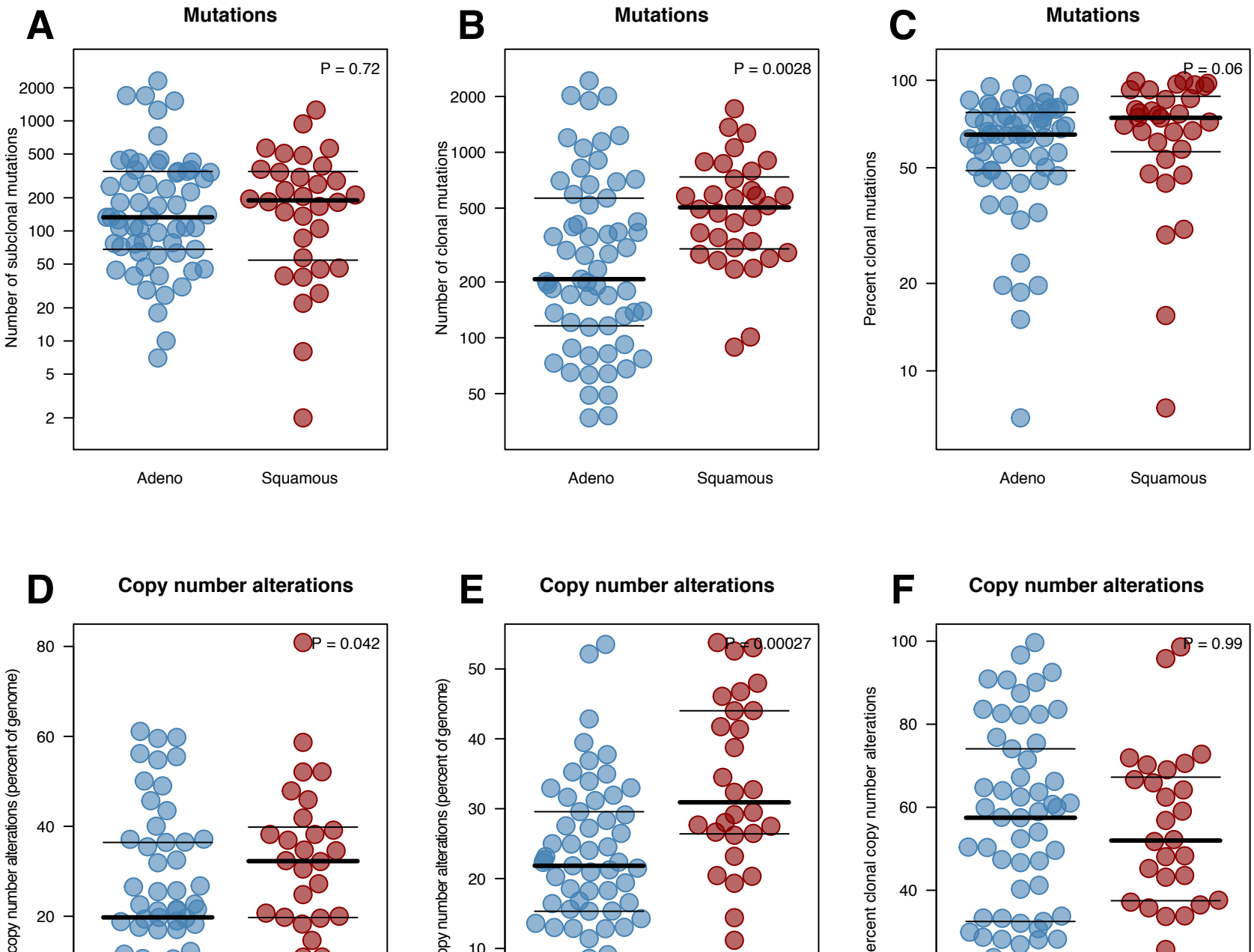


Figure S6. Mutation & copynumber intratumor heterogeneity.
Clonal, subclonal & fraction clonal mutations (A-C) or copy number alterations (D-F) by histology. Clonal, subclonal & fraction clonal mutations versus copy number alterations in adenocarcinoma (G-I) and squamous cell carcinoma (J-L). Clonal and subclonal copy number alterations defined as percent of the genome subjected to either clonal or subclonal copy number alterations. Percent clonal copy number alterations defined as percent of genome subjected to clonal copy number alterations divided by total percent of genome subjected to copy number alterations.

Figure S7

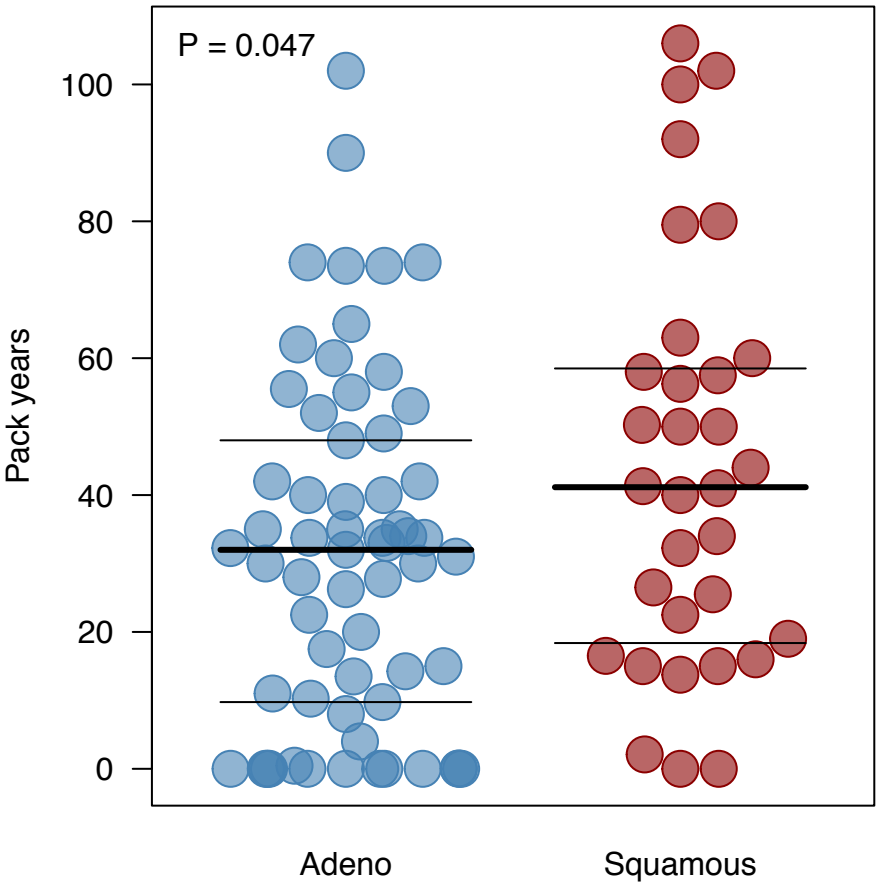


Figure S7. Pack years by histopathological subtype. Comparing the number of pack years in adenocarcinoma and squamous cell carcinoma. Never-smokers are assigned a value of 0 pack years.

Figure S8

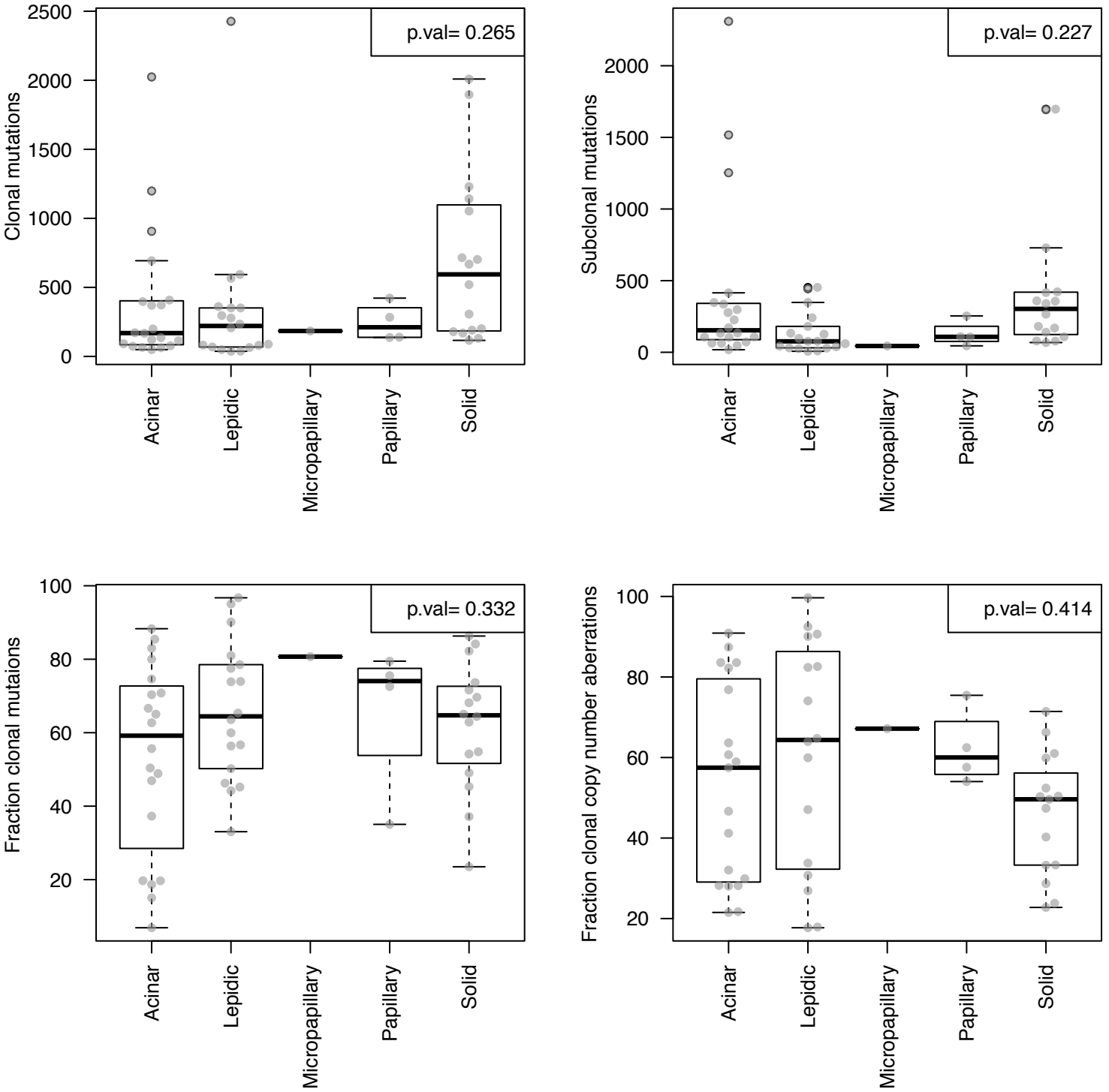


Figure S8. Mutation and copy number intratumor heterogeneity in adenocarcinoma by histopathological subtype

Comparing intratumor heterogeneity by mutation and copy number across adenocarcinoma subtypes, showing no significant subtype-dependent variation.

Figure S9

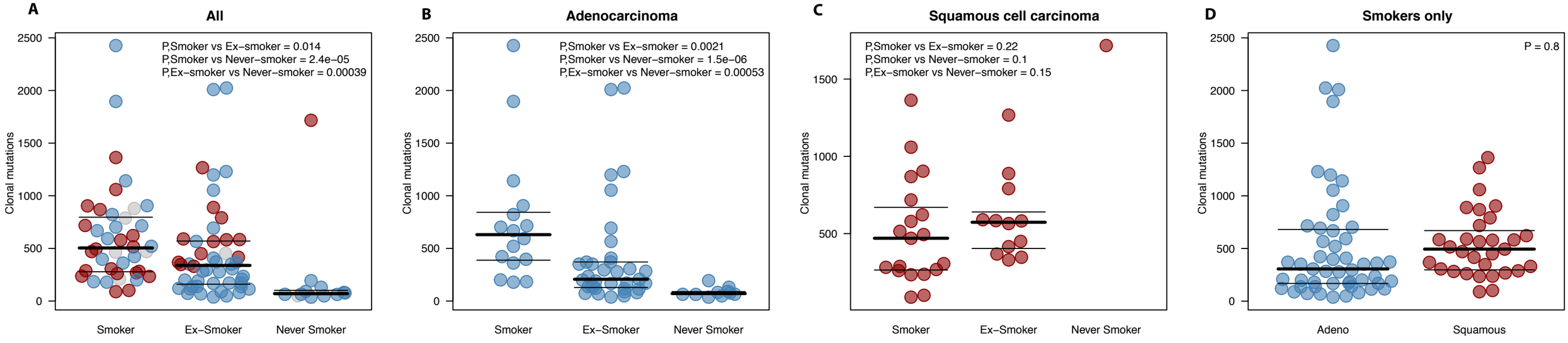


Figure S9. Clonal and subclonal mutations by smoking status.
Number of clonal (A-D) and subclonal mutations (E-H) by smoking status in All (A,E), adenocarcinoma (B,F), squamous cell carcinoma (C,G), and clonal and subclonal mutations in smokers only in adenocarcinoma and squamous cell carcinoma.

Figure S10

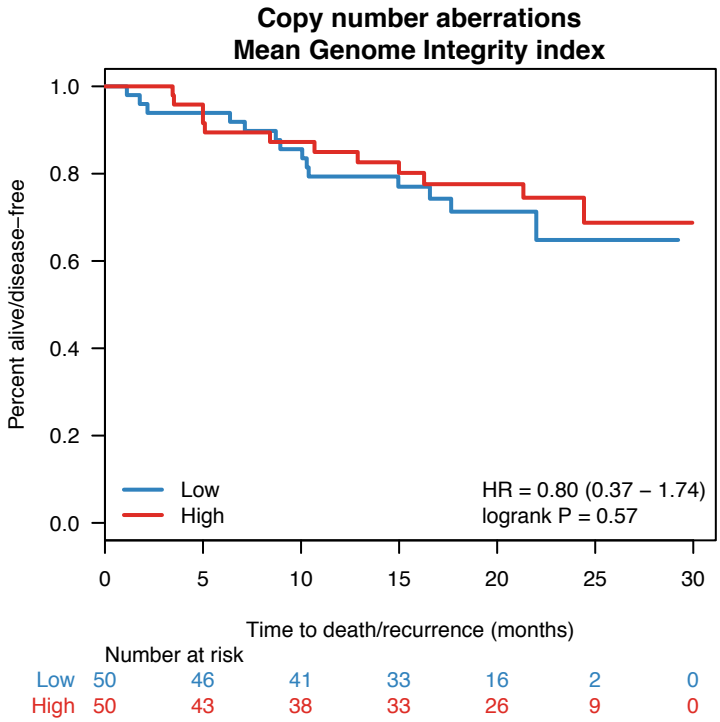
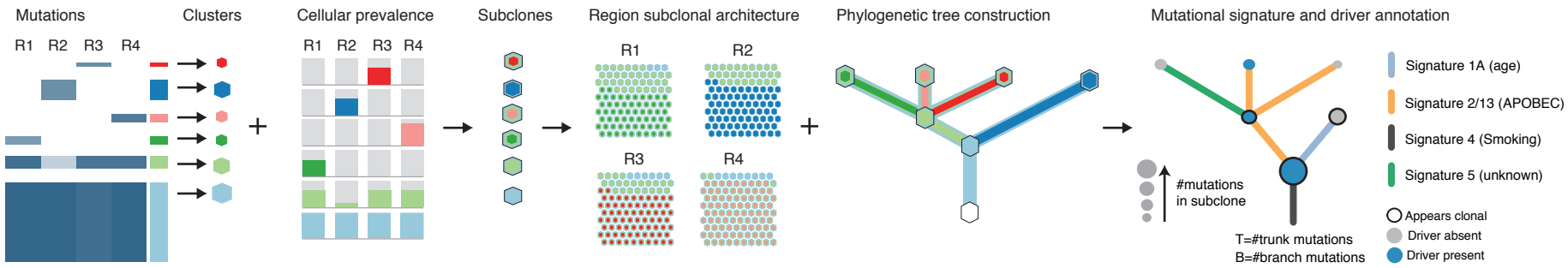


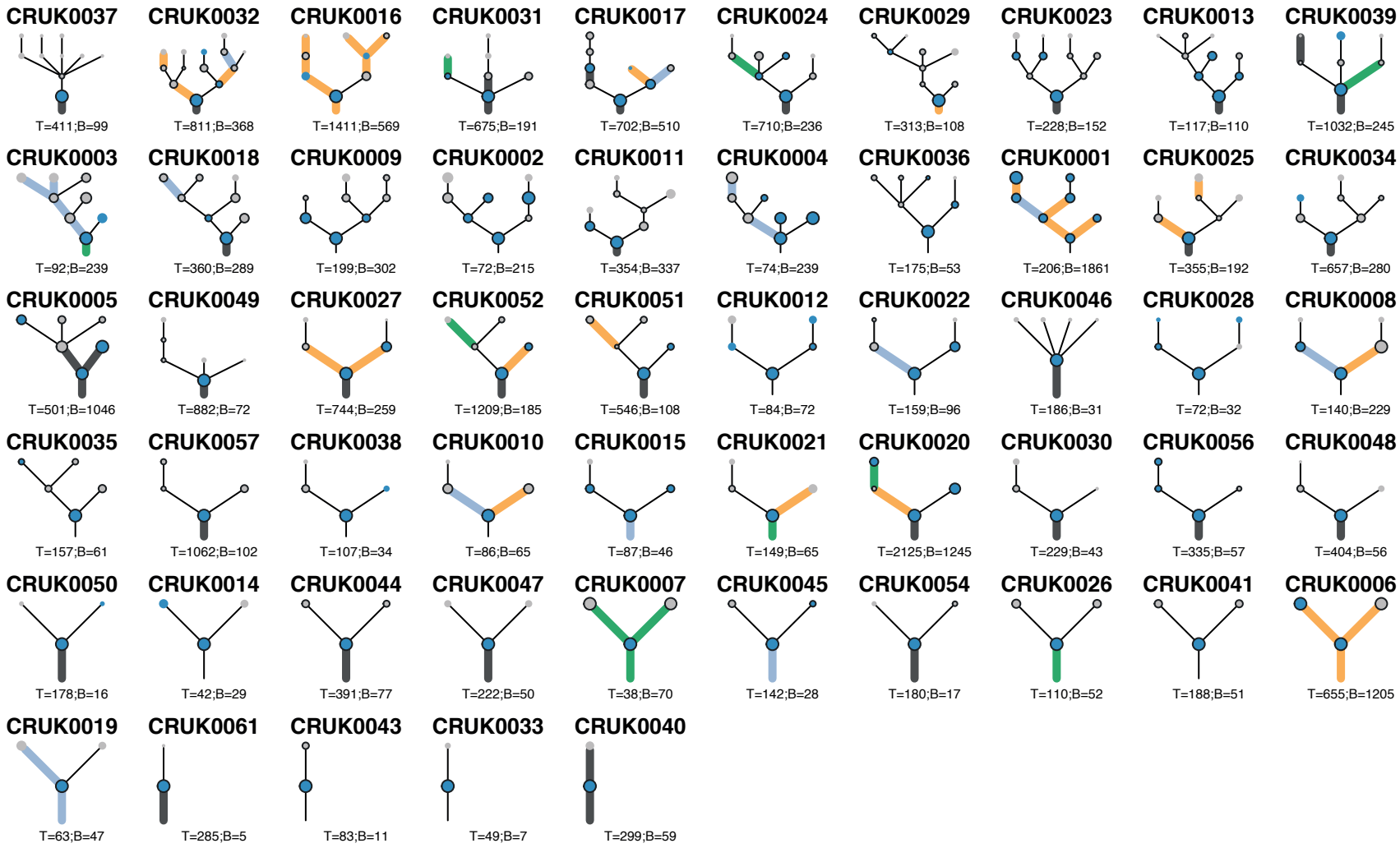
Figure S10. Survival analysis based on overall copy number aberration burden

Kaplan-Meier curve showing recurrence-free survival versus overall burden of copy number aberrations as defined by mean across tumor regions of the fraction of the genome deviating from the tumor ploidy status (genome integrity index, GII), divided into high and low based on the median.

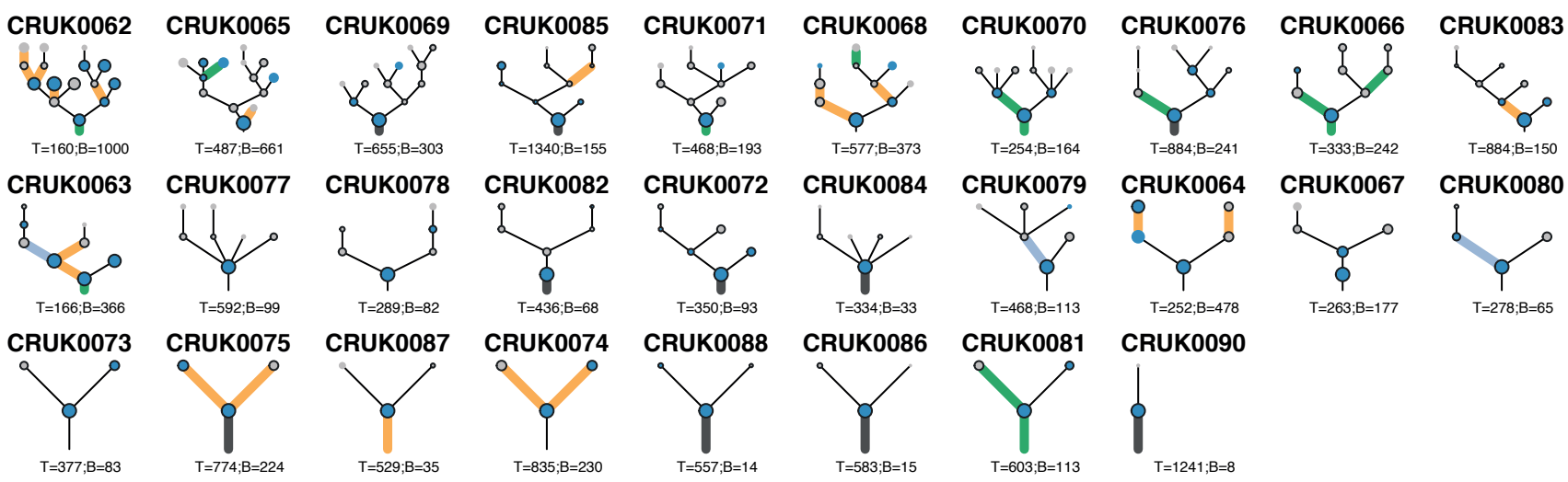
Figure S11



B Adenocarcinoma



C Squamous cell carcinoma



D Other NSCLC

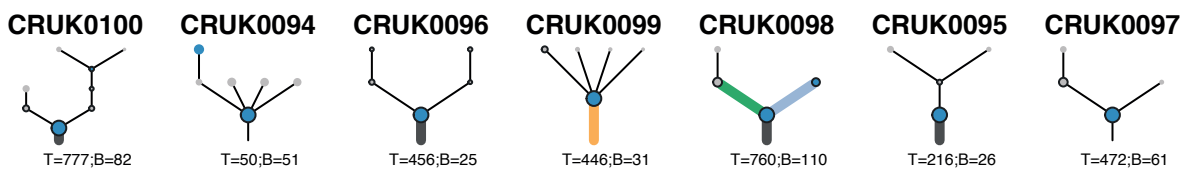


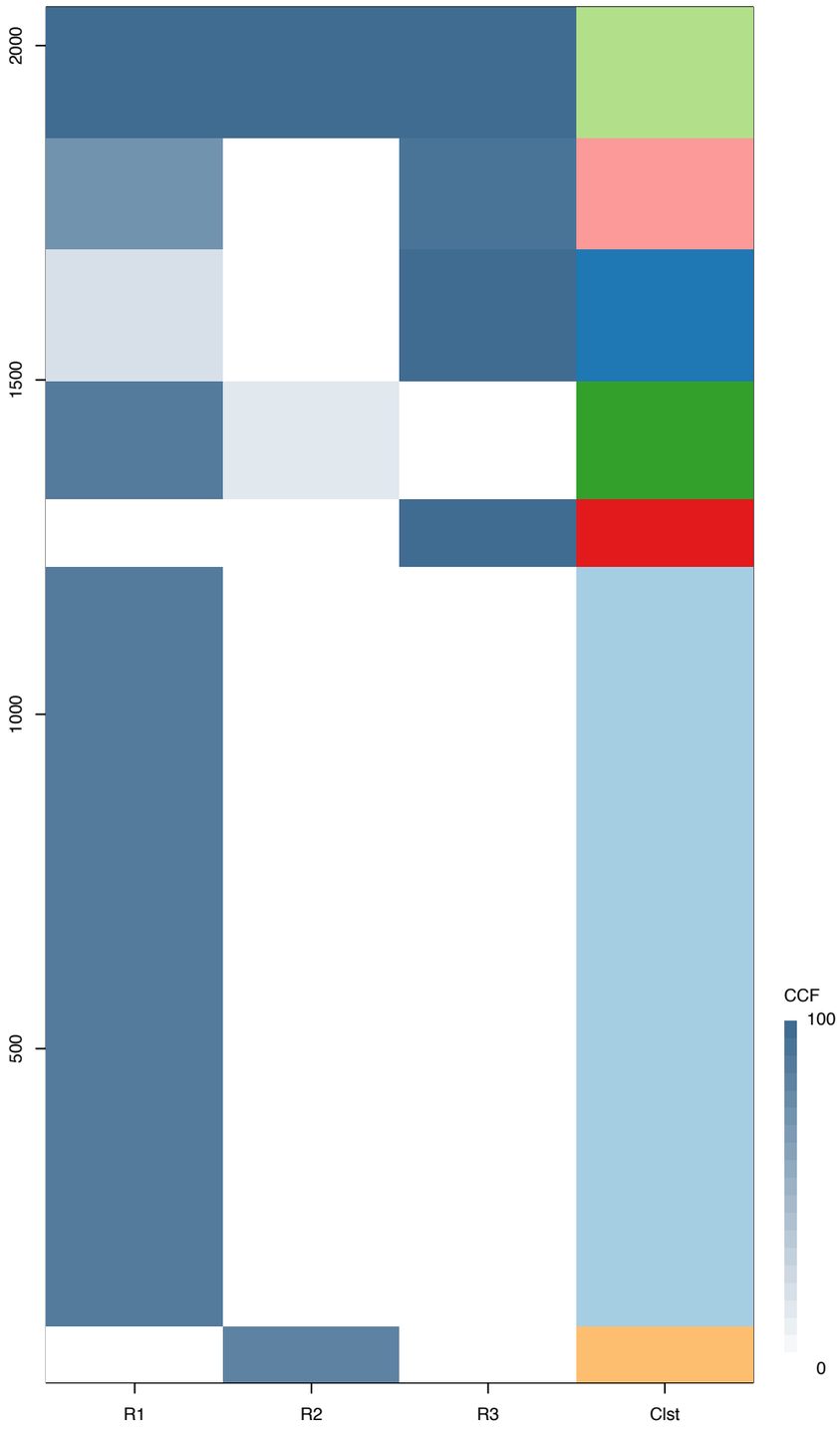
Figure S11. Phylogenetic trees

A) Schematic illustrating how the evolutionary history of each tumor is inferred based on M-seq from a single time-point. Based on their prevalence in the sequenced cancer cell population, mutations can be clustered to infer the founding clone and subclonal clusters. The mutations in the founding clone represent the trunk of the tumor's phylogenetic tree and are found in every cancer cell. The cellular prevalence of the subclonal clusters across all regions of the tumor provides the regional subclonal architecture and enables the construction of the phylogenetic tree. Finally, the mutational processes and driver events are annotated. B-D) Evolutionary trees of NSCLC tumors. For each tumor the relationships between identified subclones is depicted, with size of circle reflecting number of mutations in subclone relative to largest clone. Color of lines connecting subclones indicates dominant mutational process. Black lines either do not have a dominant mutational process or lack sufficient mutations for accurate mutational process decomposition. Length of lines connecting tumor subclones do not carry information. Subclones colored dark blue harbor a cancer driver event.
T: number of trunk mutations. B: number of branch mutations.

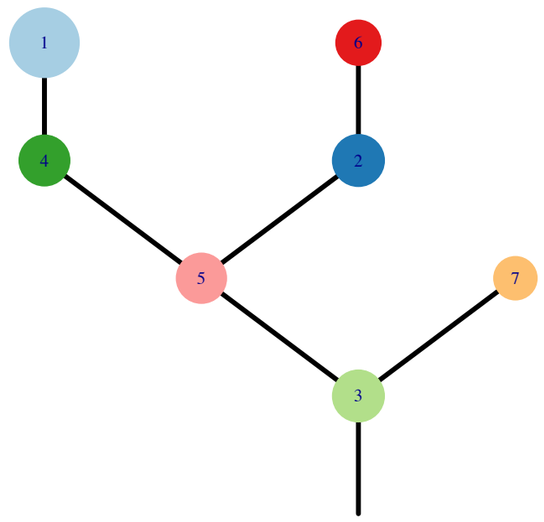
Figure S12. Phylogenetic trees

Phylogenetic trees for 91 tumors. For each tumor there may be multiple equally likely phylogenetic trees. In these cases, all possible trees are shown. For each multi-pane figure, top-left shows cancer cell fractions (CCF) as a heatmap for all clustered mutations. CCF value for each mutation represents the mean of the mutation cluster CCF values, with darker color indicating higher CCF value. Regions indicated below heatmap, "R" represents regions from primary tumor, "LN" represents lymph node metastasis regions. "Clst" is a colored bar representing individual mutation clusters. Middle top pane shows the complete phylogenetic tree as constructed based on the mutation clusters. Each cluster is numbered, with largest cluster labeled "1". Patient clinical information is shown below the tree. Cancer driver genes found in the tumor, with mutation cluster, cytoband, and type (mutation [SNV]/copynumber aberration [Amp/Del]) indicated on the right. Driver alterations considered targetable by either the MATCH or NLMT clinical trials (shown in supplementary figure 27) are highlighted with a red box. Note that some alterations are histology-specific, or only targetable with specific wild-type variants, thus specific driver alteration may not always be targetable in different samples. Below are phylogenetic trees drawn for each region, with clusters not found in a given region shown in grey. Grid of 100 representative cells are shown beneath the tree. The nested colors within each cell represent the mutational clusters present.

Fig.S12A



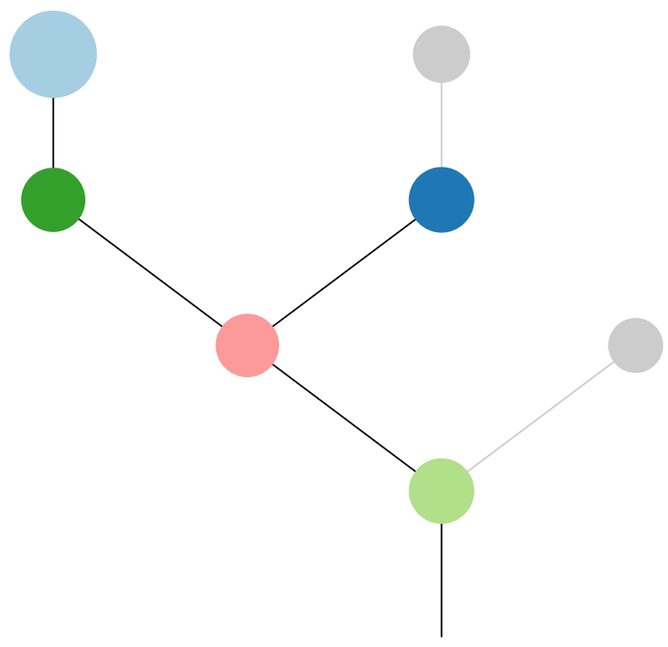
CRUK0001



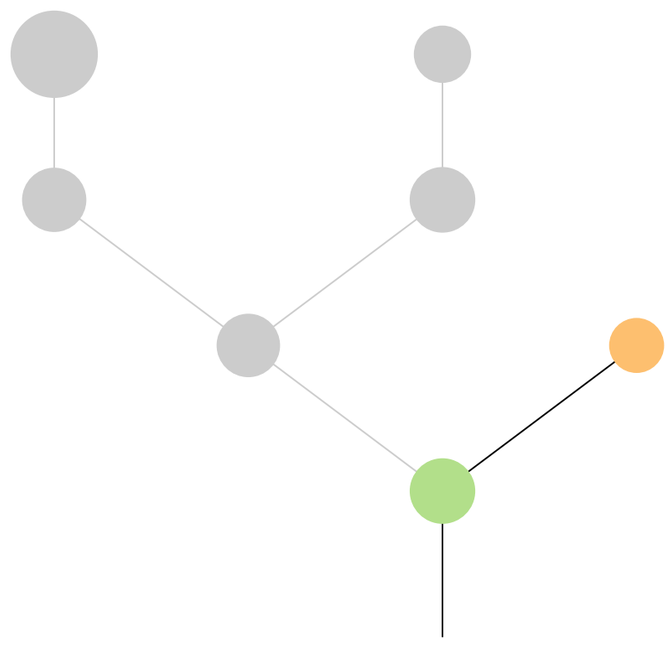
Histology:Adeno, Age:85, PackYears:31, Size:26
Stage:2a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
CDKN2C	1	1p32.3	SNV
CDH1	1	16q22.1	SNV
NF1	1	17q11.2	SNV
CDK4	2	12q14.1	Amp
LRIG3	2	12q14.1	Amp
ARHGAP35	2	19q13.32	SNV
CARD11	3	7p22.2	Amp
RAC1	3	7p22.1	Amp
ETV1	3	7p21.2	Amp
HNRNPA2B1	3	7p15.2	Amp
HOXA9	3	7p15.2	Amp
HOXA11	3	7p15.2	Amp
HOXA13	3	7p15.2	Amp
JAZF1	3	7p15.2	Amp
EGFR	3	7p11.2	Amp
WRN	3	8p12	SNV
MGA	3	15q15.1	SNV
TP53	3	17p13.1	SNV
CD79A	3	19q13.2	Amp
ZNF180	3	19q13.31	Amp
BCL3	3	19q13.32	Amp
CDK4	4	12q14.1	Amp
LRIG3	4	12q14.1	Amp
MGA	4	15q15.1	SNV
PASK	5	2q37.3	SNV
WIF1	6	12q14.3	Amp
HMGA2	6	12q14.3	Amp
CDK4	7	12q14.1	Amp
LRIG3	7	12q14.1	Amp

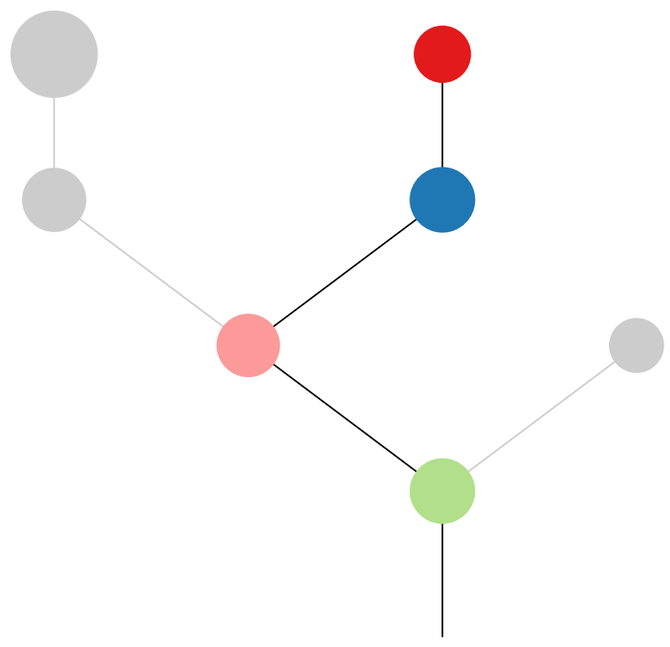
R1

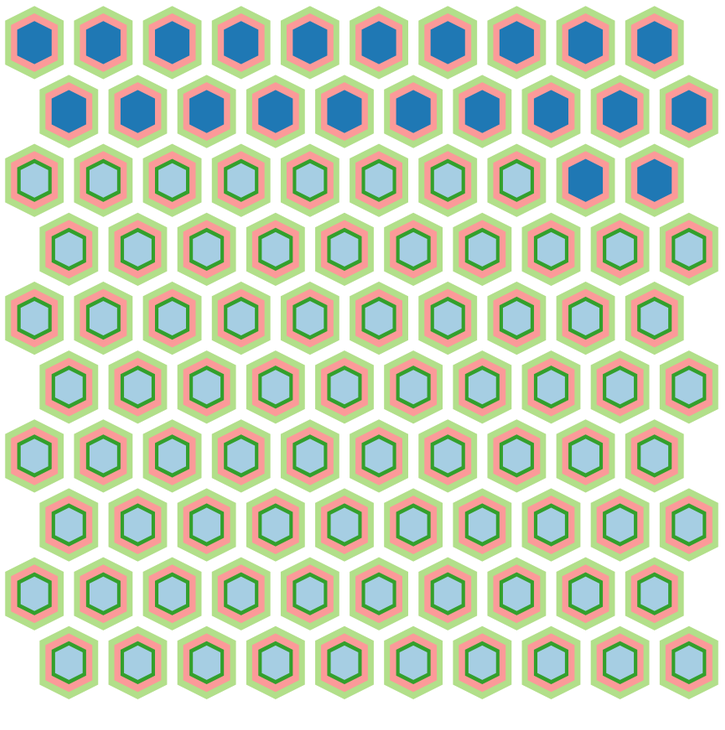


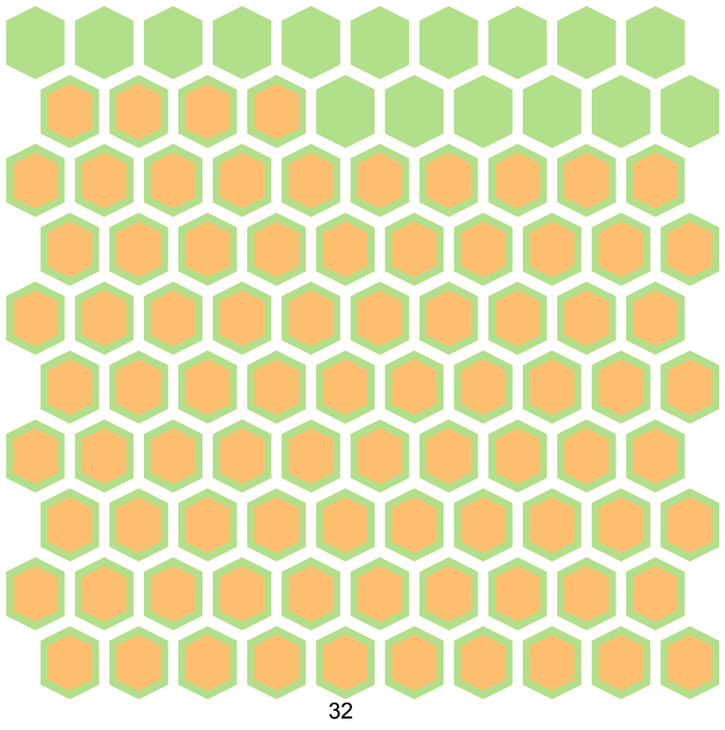
R2



R3







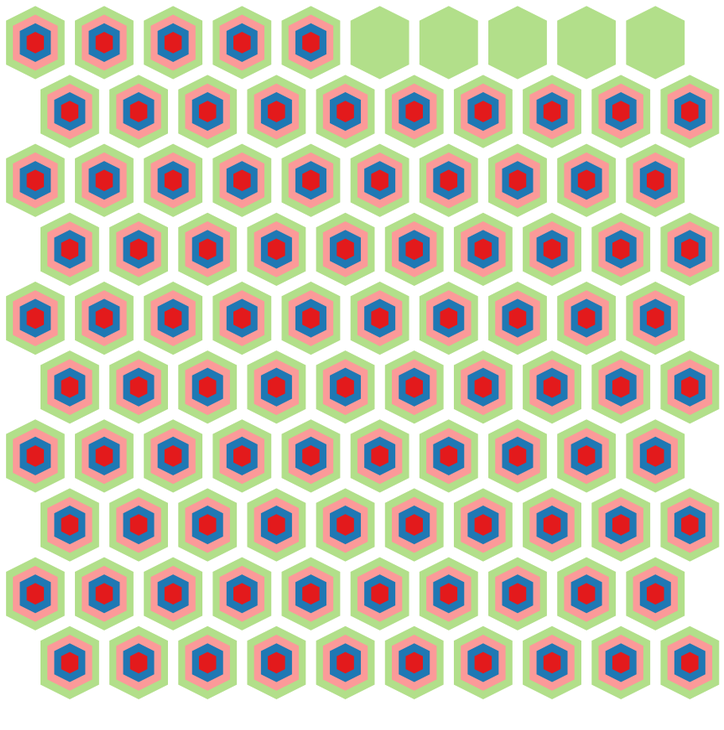
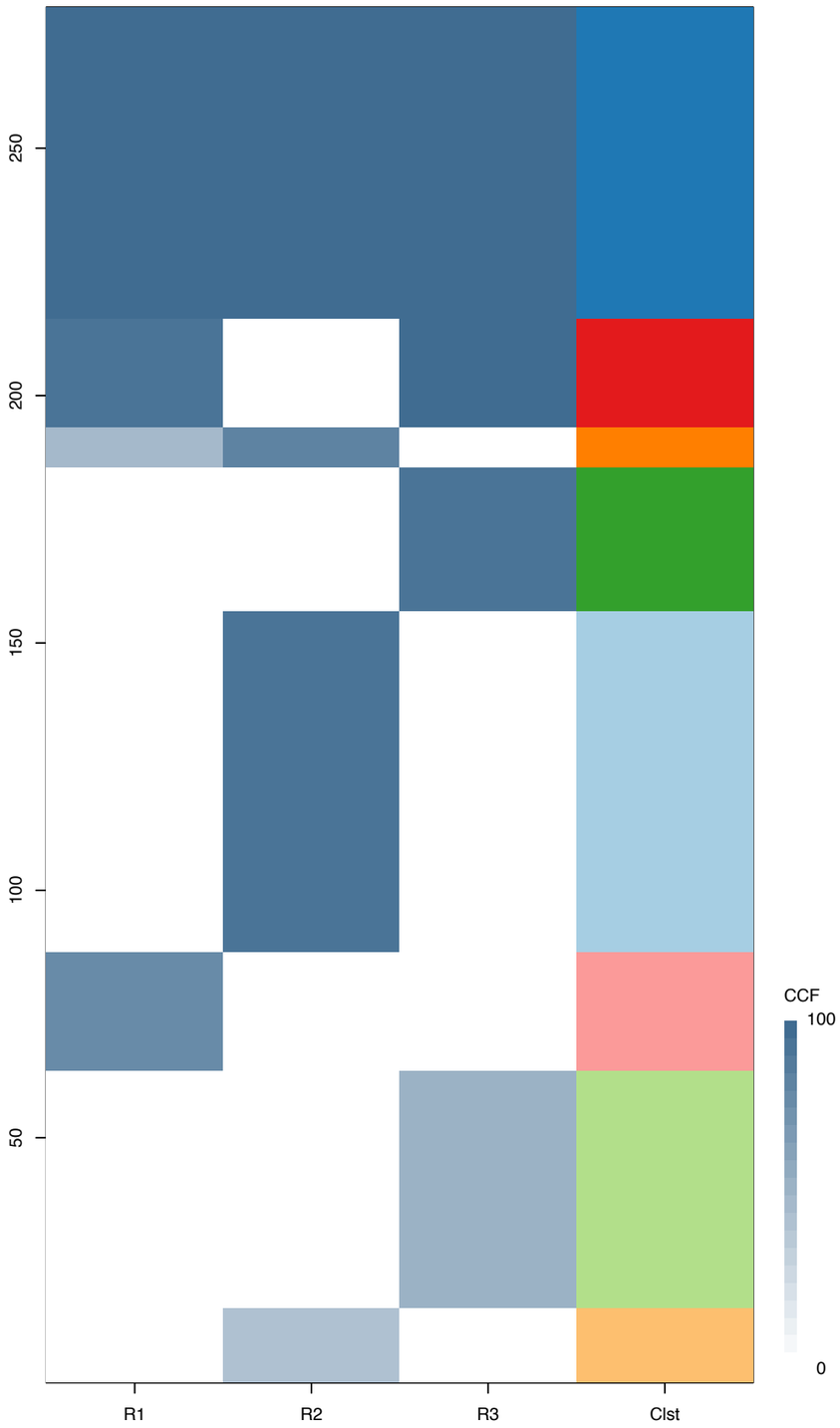
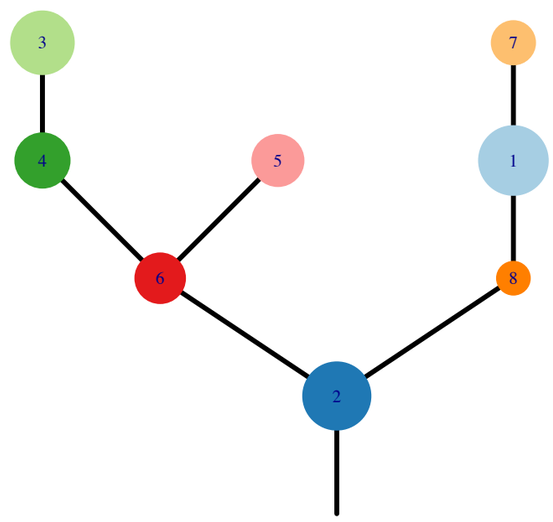


Fig.S12B



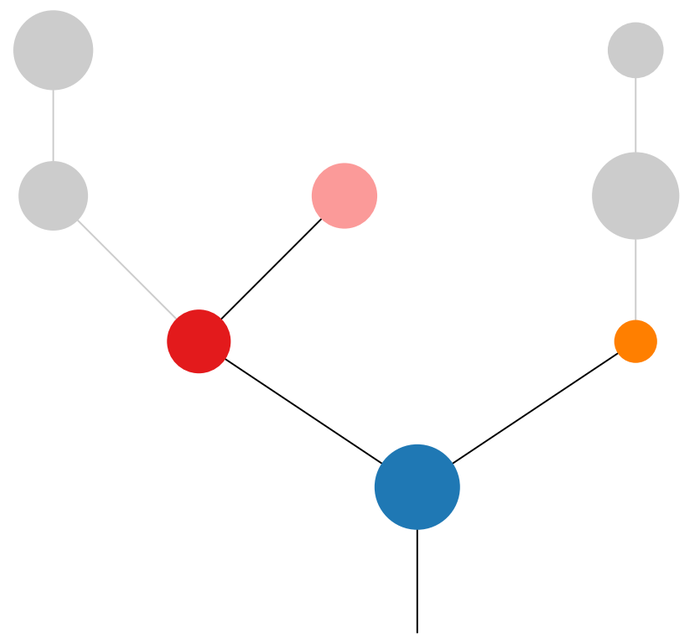
CRUK0002



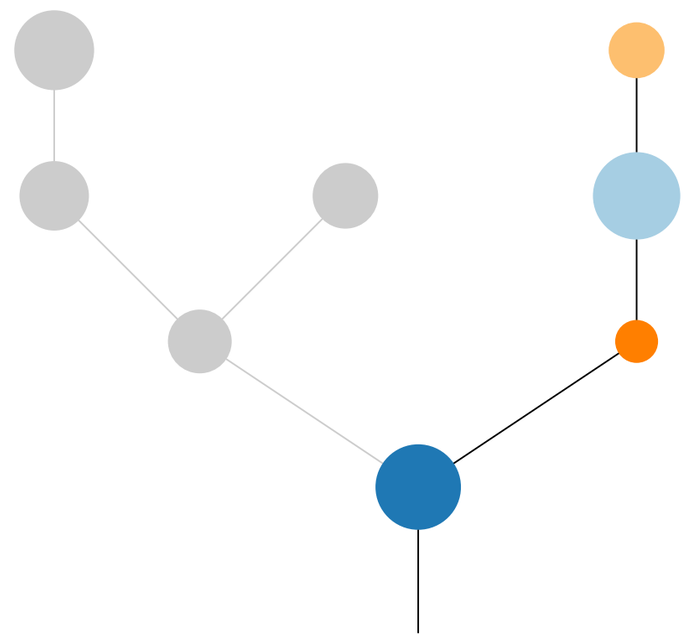
Histology:Adeno, Age:83, PackYears:8, Size:35
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
IKZF1	1	7p12.2	SNV
ETNK1	1	12p12.1	Amp
KRAS	1	12p12.1	Amp
PPFIBP1	1	12p11.23	Amp
RB1	1	13q14.2	SNV
TERT	2	5p15.33	Amp
MET	2	7q31.2	SNV
DDIT3	2	12q13.3	Amp
CDK4	2	12q14.1	Amp
LRIG3	2	12q14.1	Amp
WIF1	2	12q14.3	Amp
MDM2	2	12q15	Amp
NF1	5	17q11.2	SNV
HMGA2	6	12q14.3	Amp
EP300	6	22q13.2	SNV

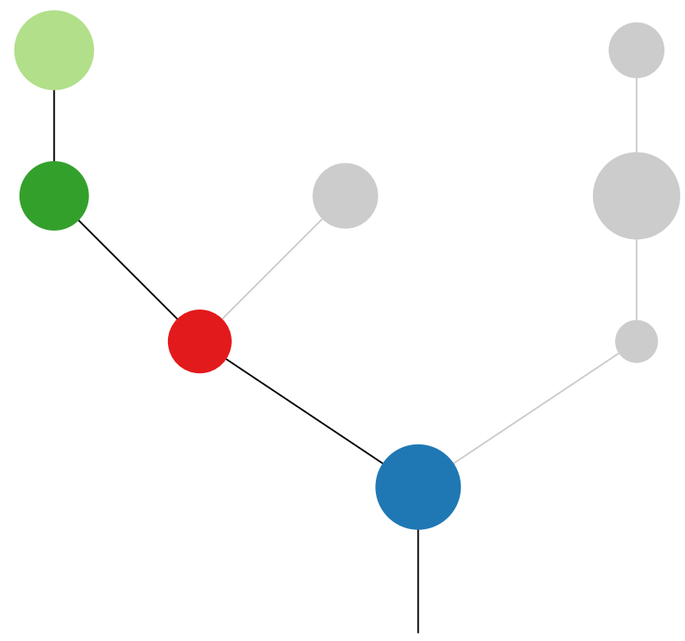
R1

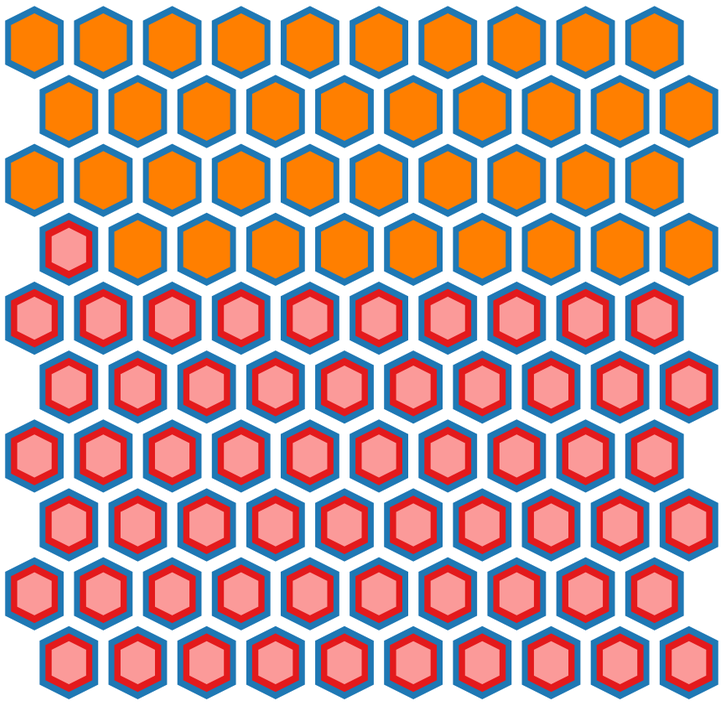


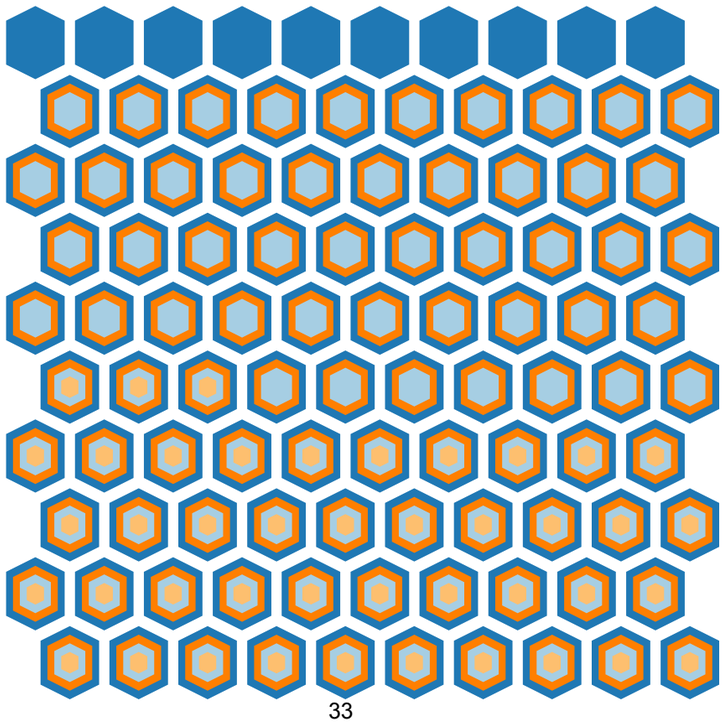
R2



R3







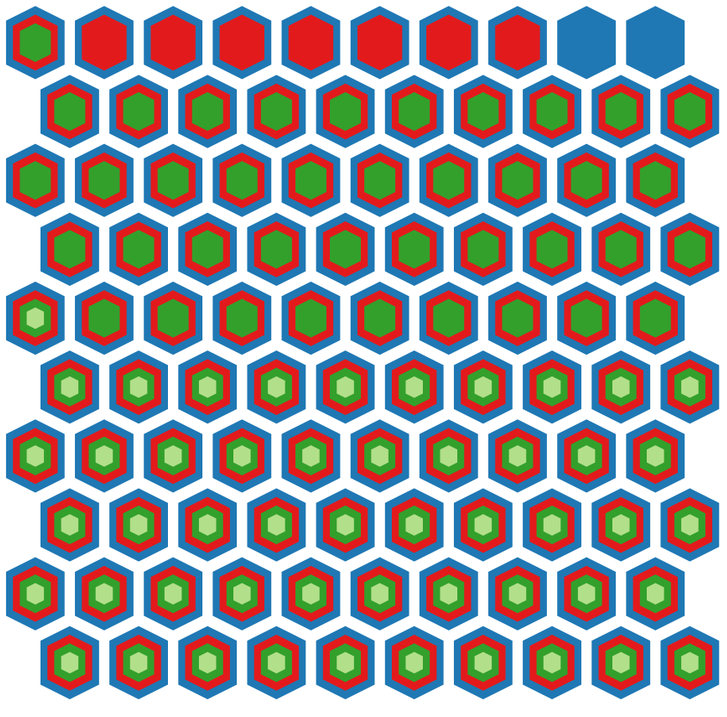
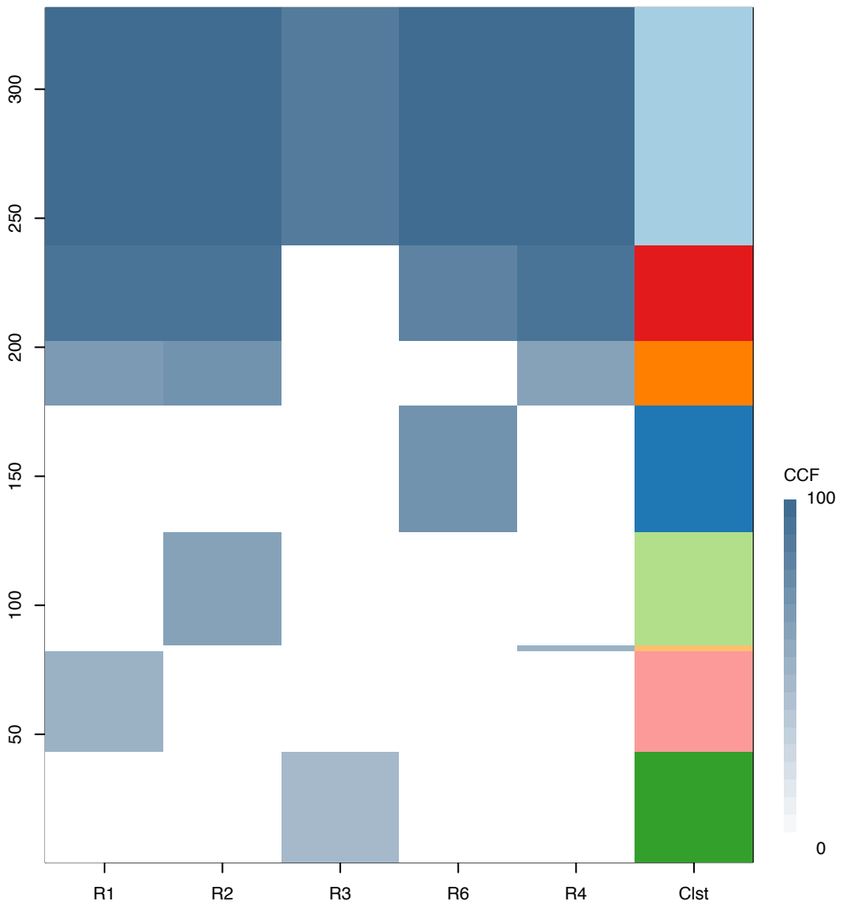
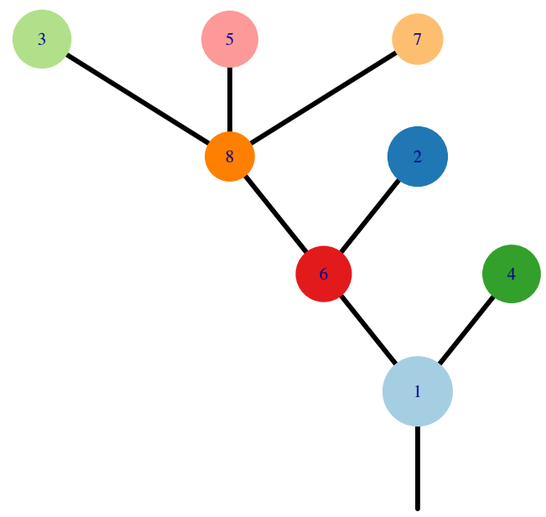


Fig.S12C



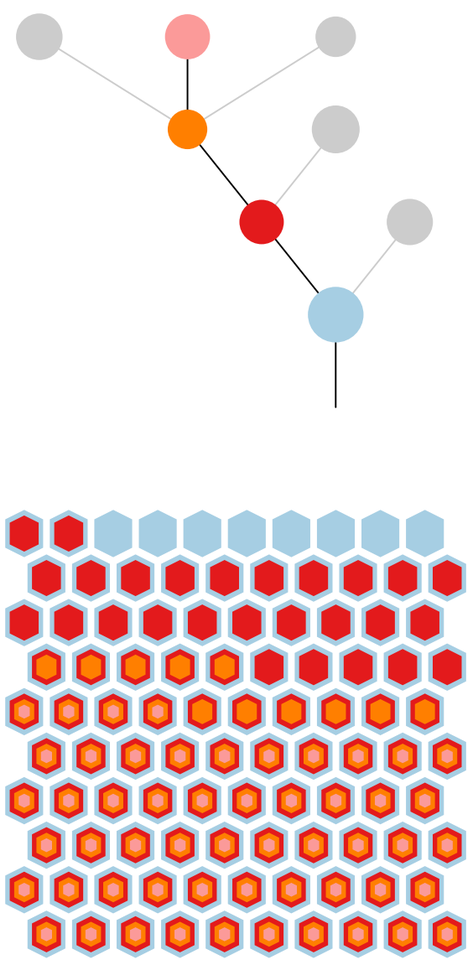
CRUK0003



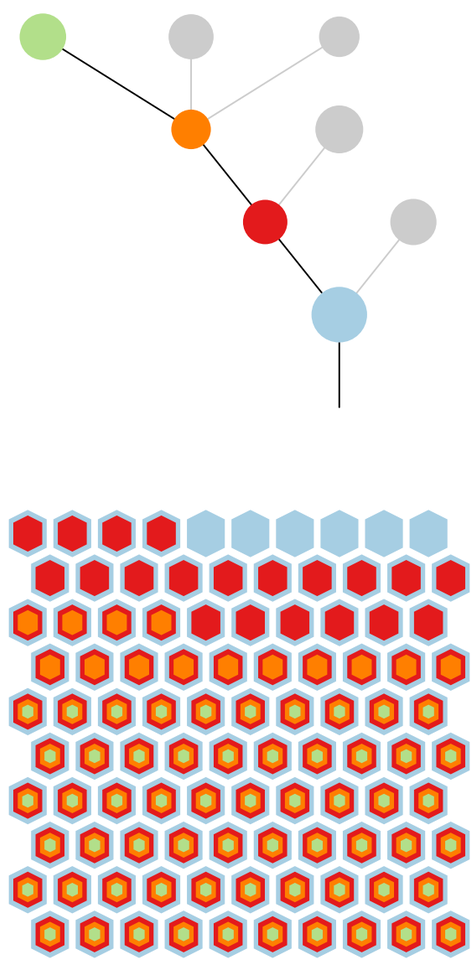
Histology:Adeno, Age:79, PackYears:NA, Size:52
Stage:3a, Gender:Female, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
PIK3CA	1	3q26.32	SNV
EGFR	1	7p11.2	SNV
CDKN2A	1	9p21.3	Del
CTNNB1	4	3p22.1	SNV

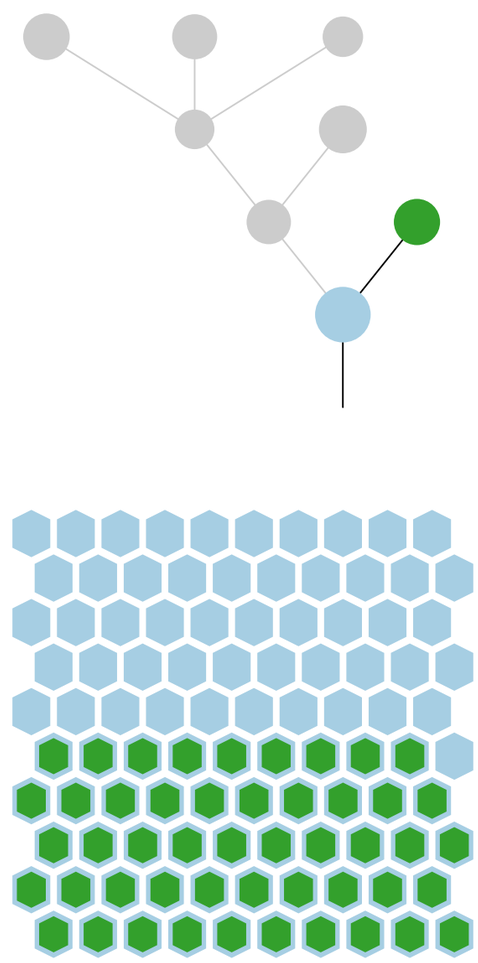
R1



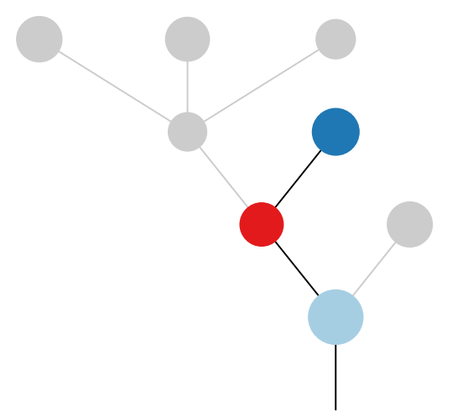
R2



R3



R6



R4

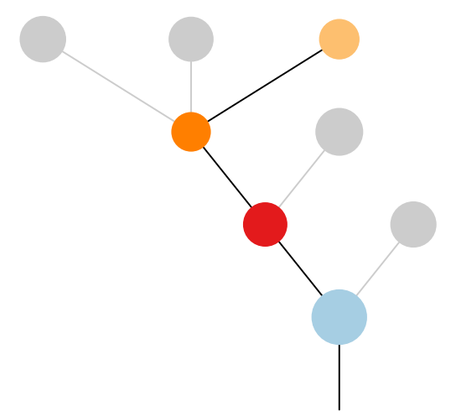
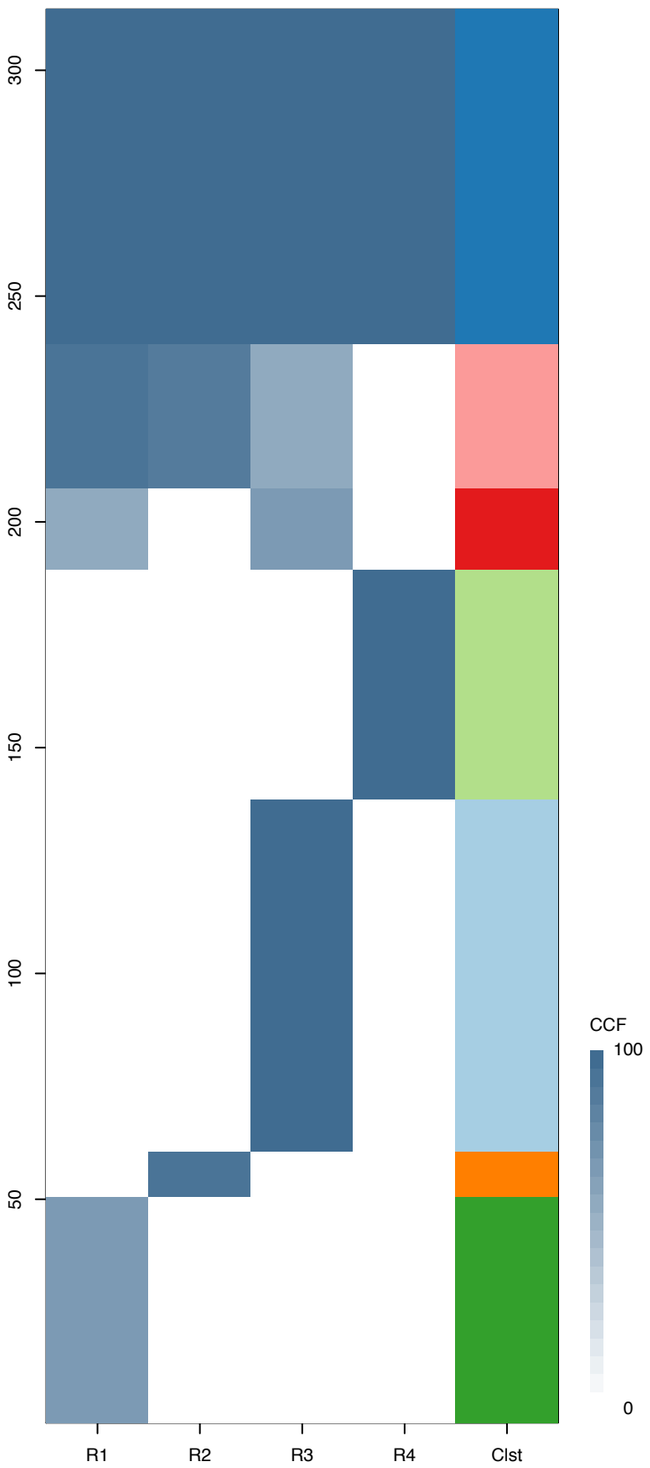
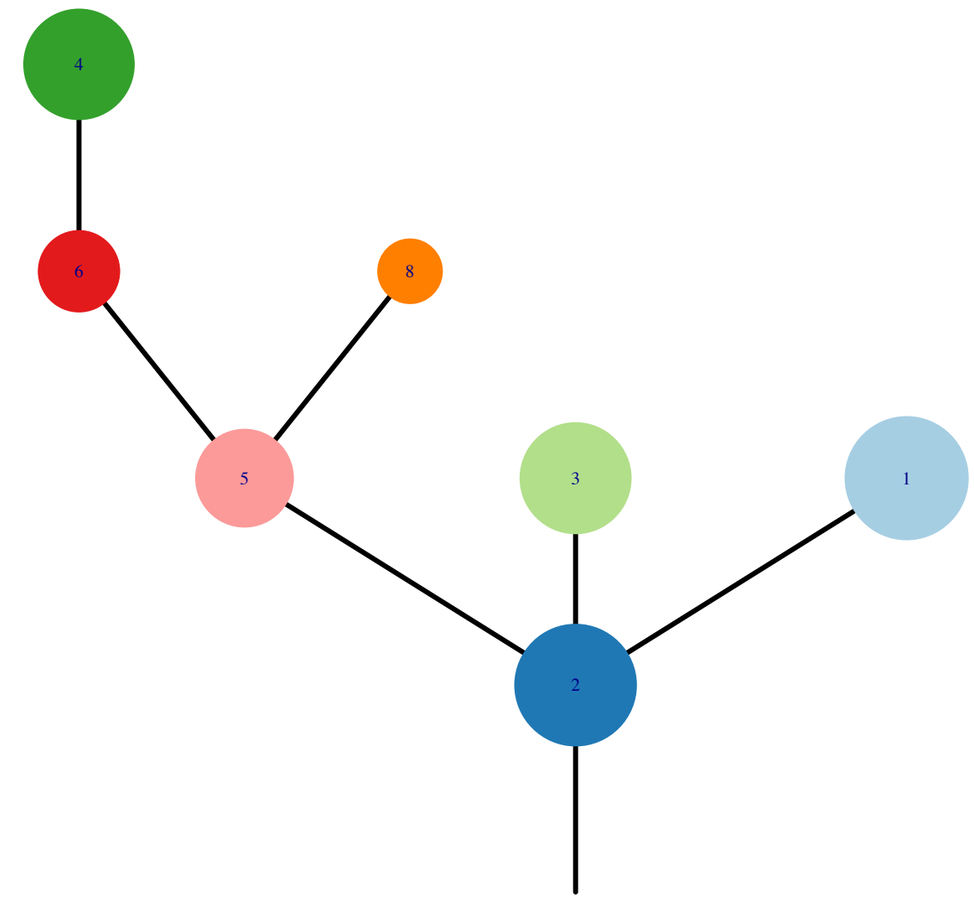


Fig.S12D



CRUK0004_A



Histology:Adeno, Age:70, PackYears:13.5, Size:47
Stage:2b, Gender:Female, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
NOTCH1	1	9q34.3	SNV
SMAD4	1	18q21.2	SNV
EGFR	2	7p11.2	Amp
EGFR	2	7p11.2	SNV
FUS	2	16p11.2	Amp
TP53	2	17p13.1	SNV
TFEB	3	6p21.1	Amp
CCND3	3	6p21.1	Amp
HSP90AB1	3	6p21.1	Amp
MLLT6	3	17q12	Amp
LASP1	3	17q12	Amp
ERBB2	3	17q12	Amp
AKAP9	8	7q21.2	Amp
CDK6	8	7q21.2	Amp

R1 R2 R3 R4

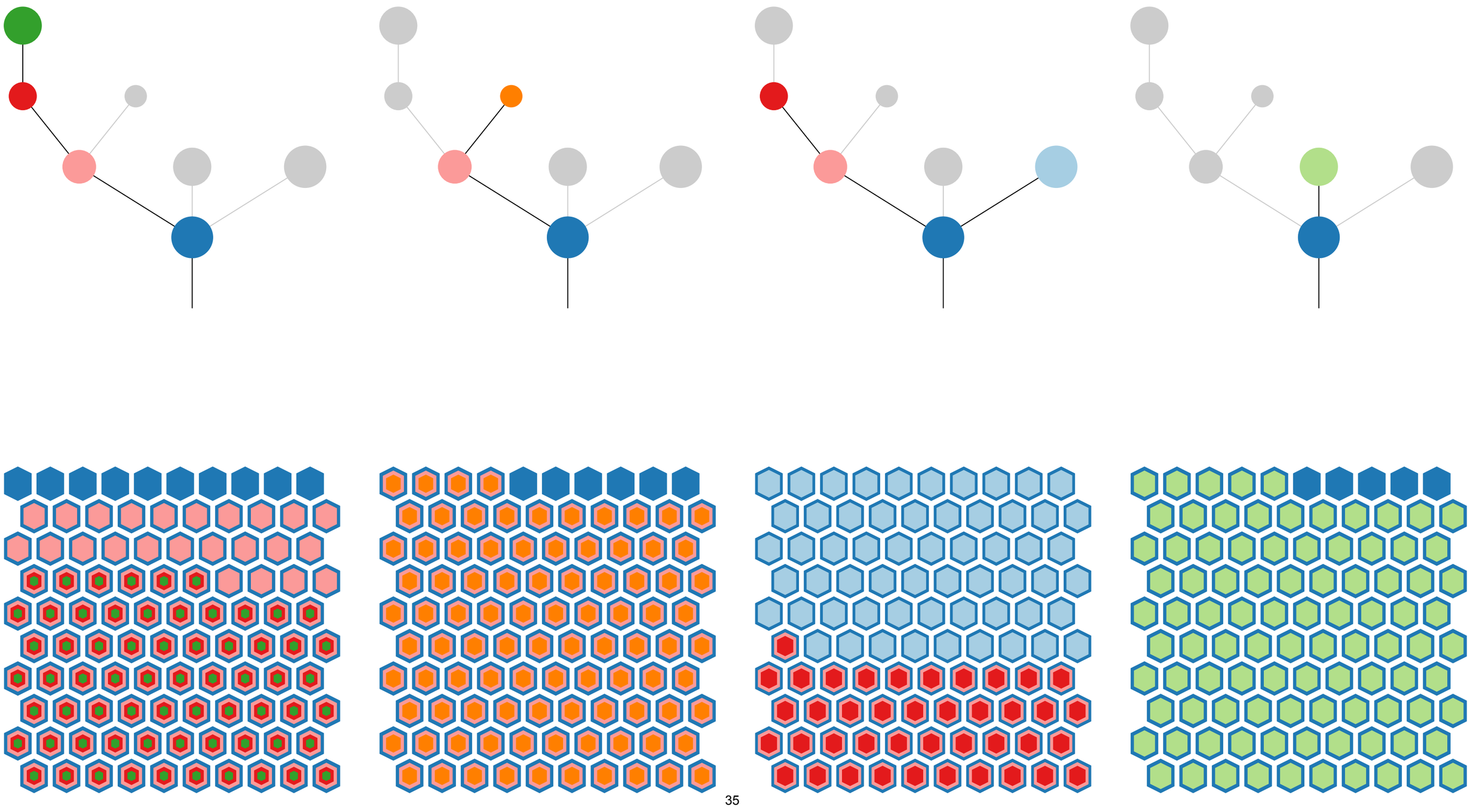
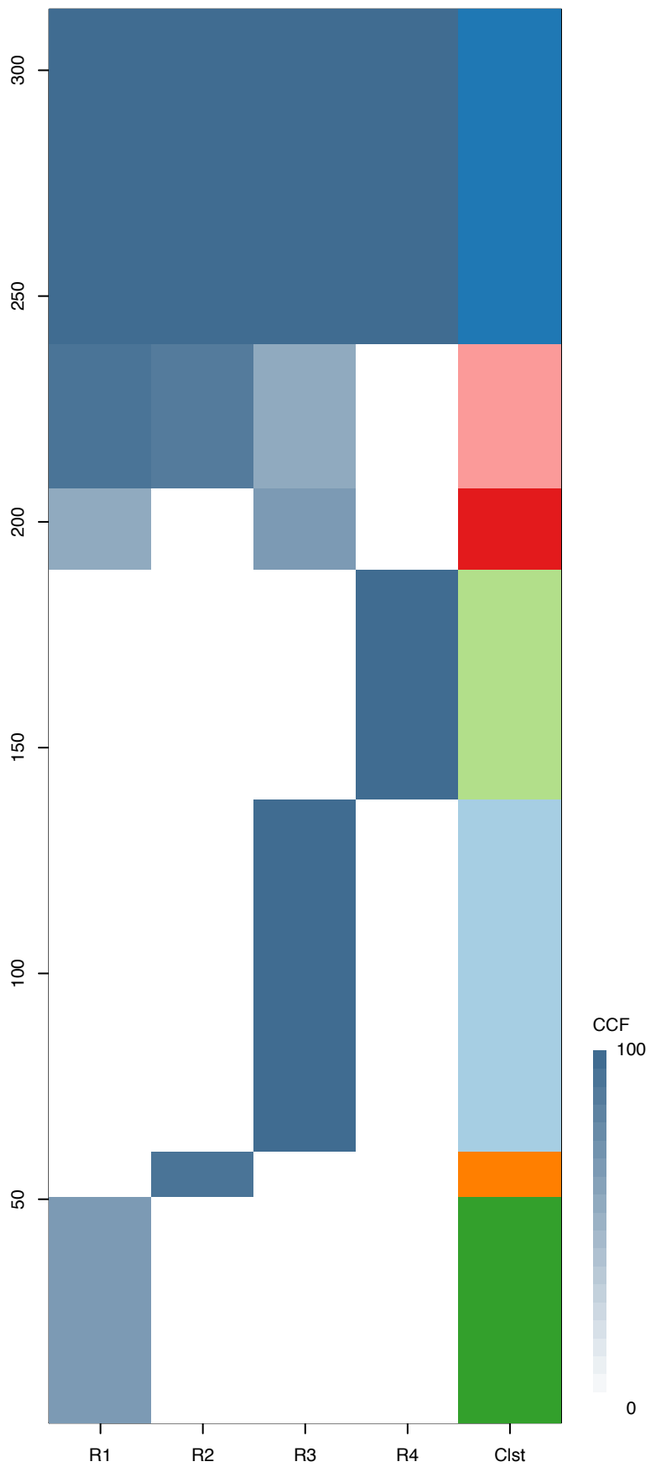
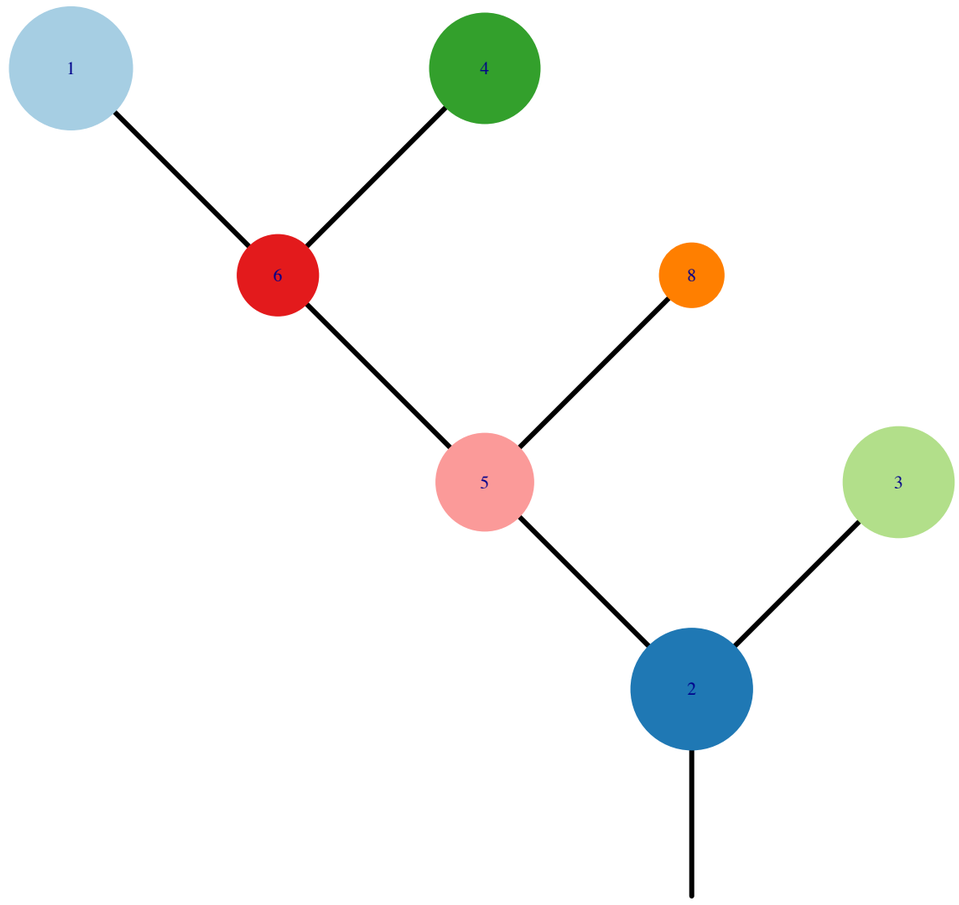


Fig.S12D



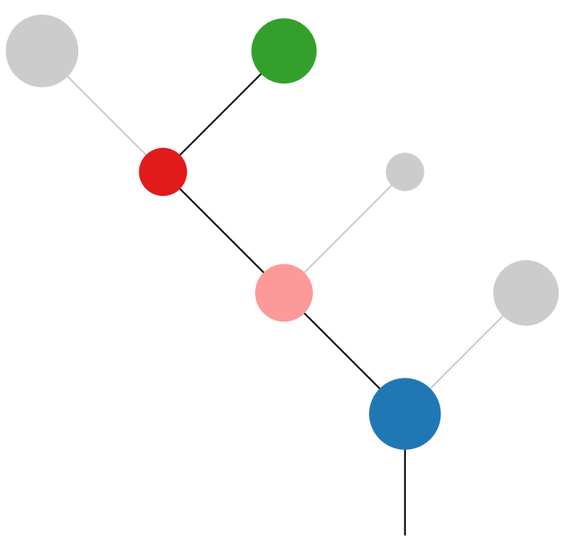
CRUK0004_B



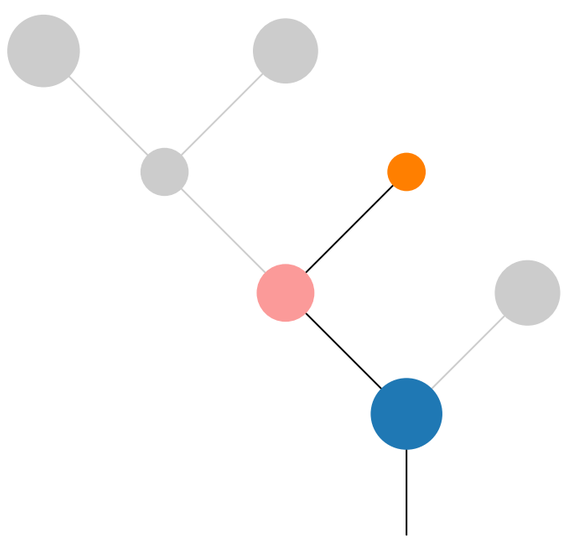
Histology:Adeno, Age:70, PackYears:13.5, Size:47
Stage:2b, Gender:Female, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
NOTCH1	1	9q34.3	SNV
SMAD4	1	18q21.2	SNV
EGFR	2	7p11.2	Amp
EGFR	2	7p11.2	SNV
FUS	2	16p11.2	Amp
TP53	2	17p13.1	SNV
TFEB	3	6p21.1	Amp
CCND3	3	6p21.1	Amp
HSP90AB1	3	6p21.1	Amp
MLLT6	3	17q12	Amp
LASP1	3	17q12	Amp
ERBB2	3	17q12	Amp
AKAP9	8	7q21.2	Amp
CDK6	8	7q21.2	Amp

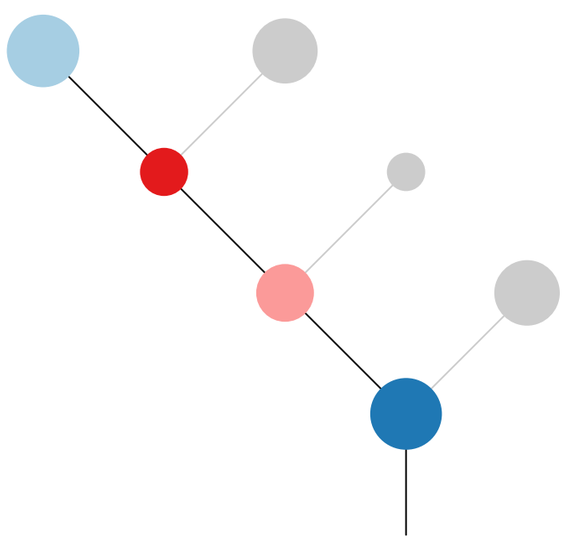
R1



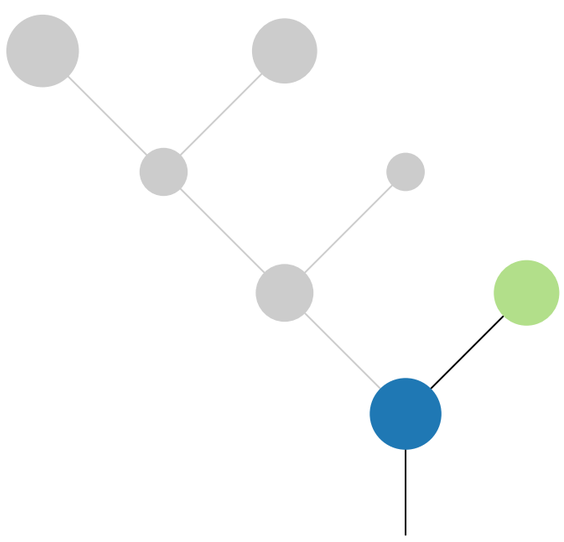
R2



R3



R4



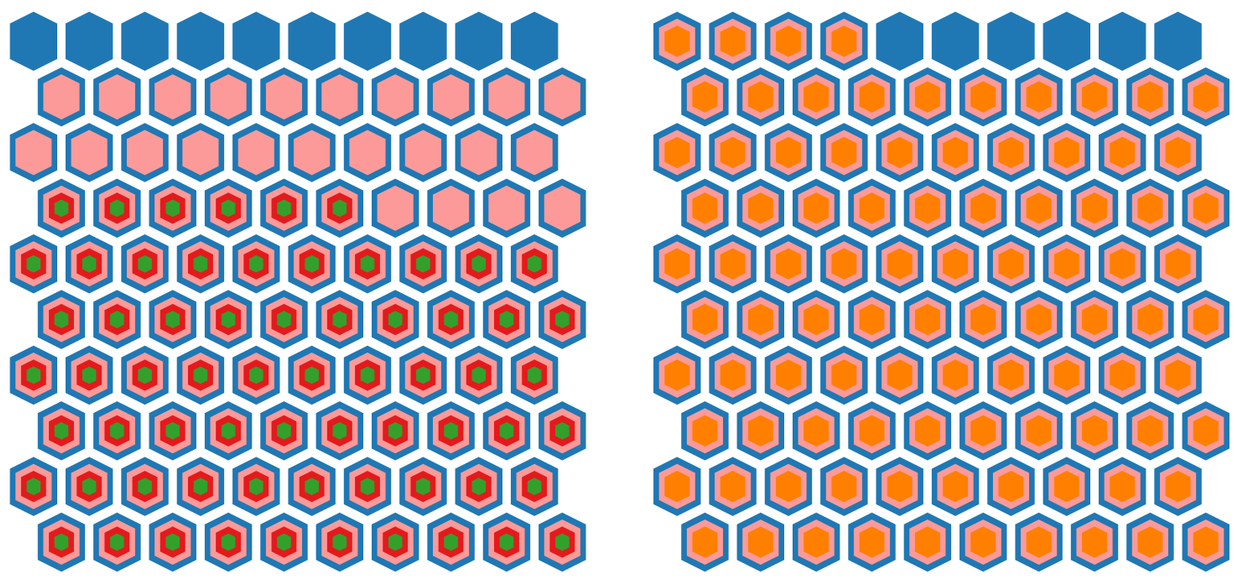
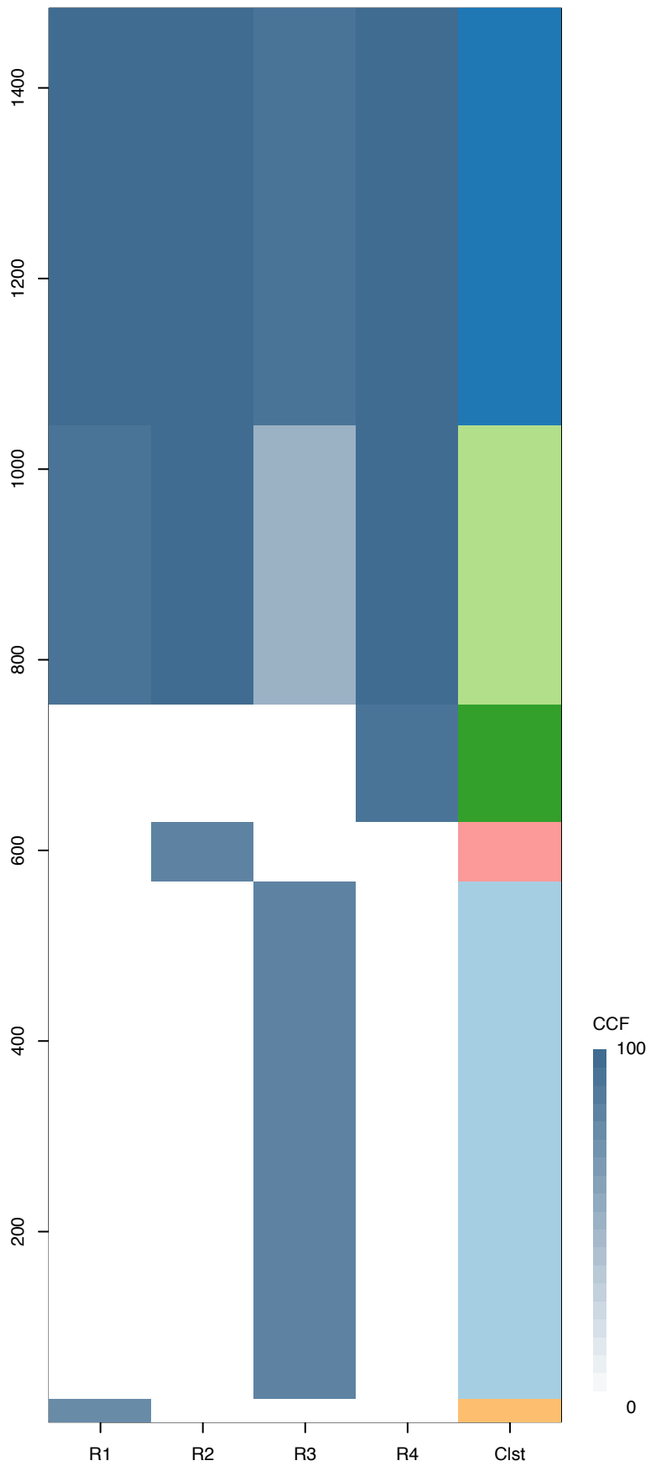
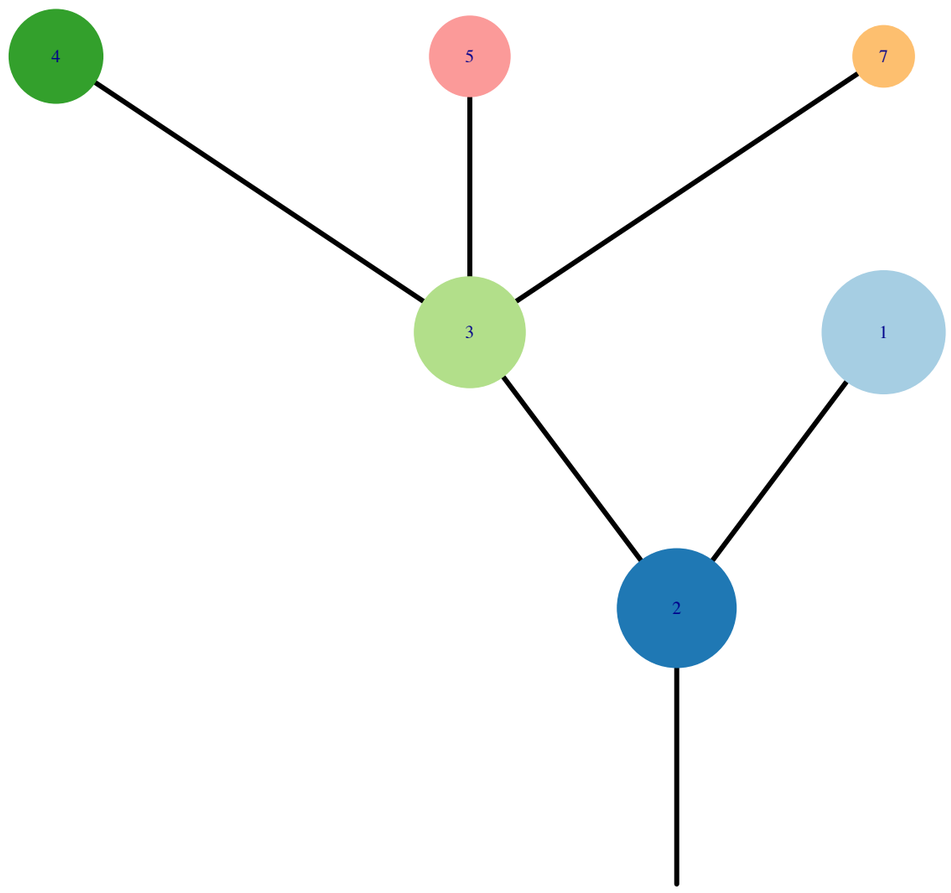


Fig.S12E



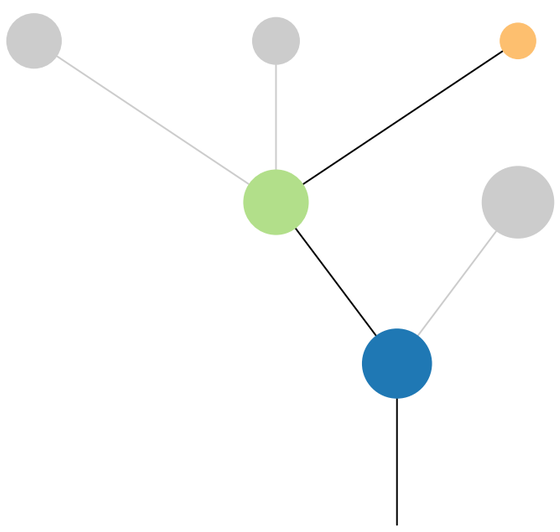
CRUK0005_A



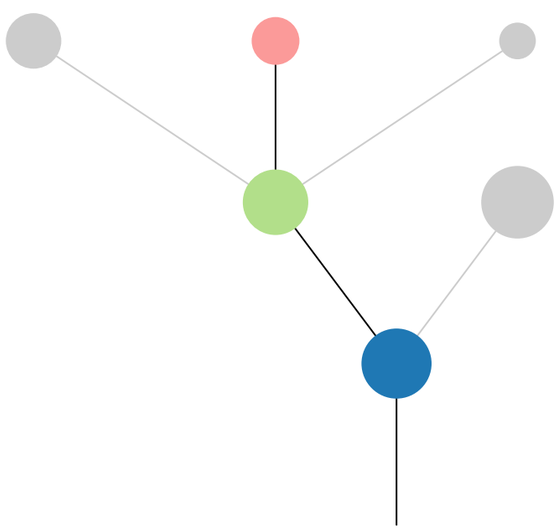
Histology:Adeno, Age:76, PackYears:55, Size:43
Stage:3a, Gender:Female, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
NRAS	1	1p13.2	SNV
BRE	2	2p23.2	Amp
PASK	2	2q37.3	SNV
TERT	2	5p15.33	Amp
IL7R	2	5p13.2	Amp
LIFR	2	5p13.1	Amp
PMS2	2	7p22.1	SNV
BRAF	2	7q34	SNV
CDK4	2	12q14.1	Amp
CMTR2	2	16q22.2	SNV
TP53	2	17p13.1	SNV
LSM14A	4	19q13.11	Amp
CLTCL1	?	22q11.21	Amp
SEPT5	?	22q11.21	Amp
ATRX	?		SNV

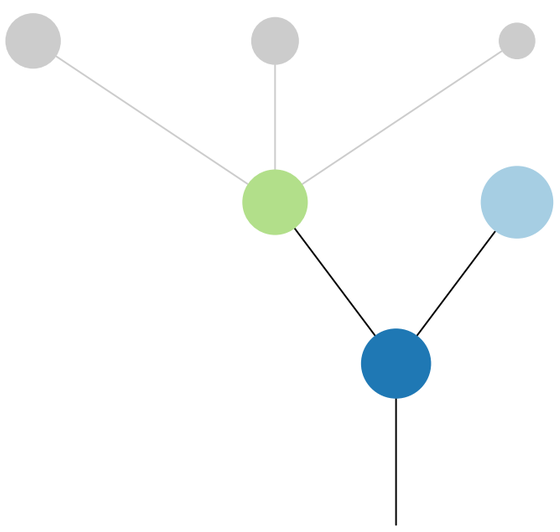
R1



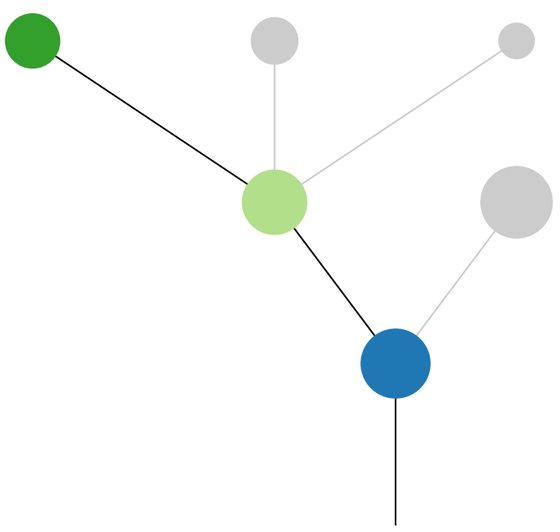
R2



R3



R4



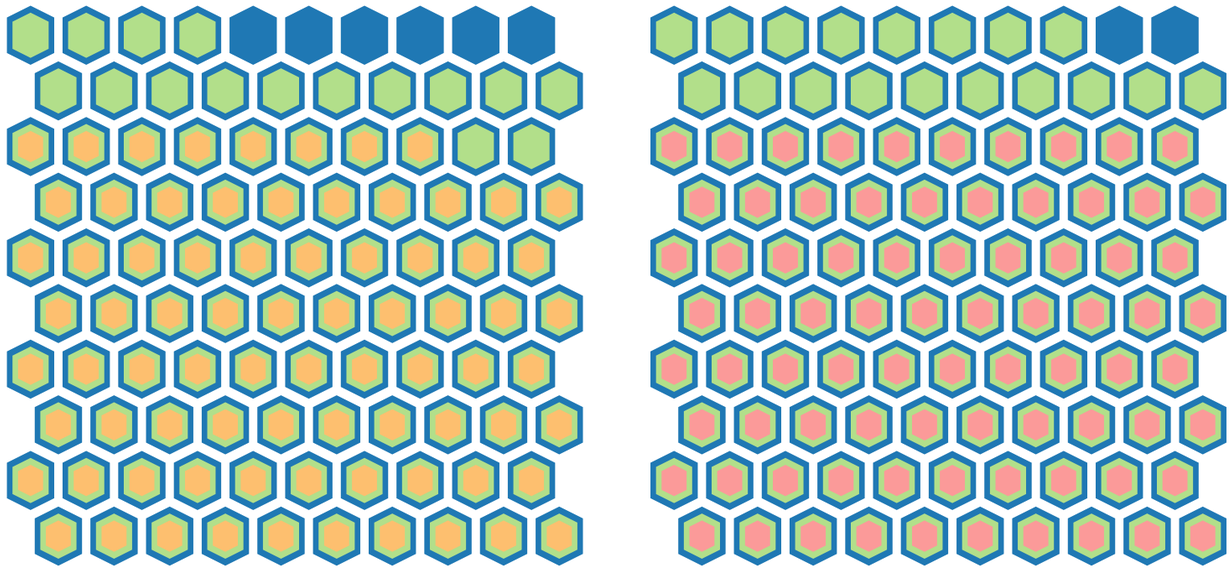
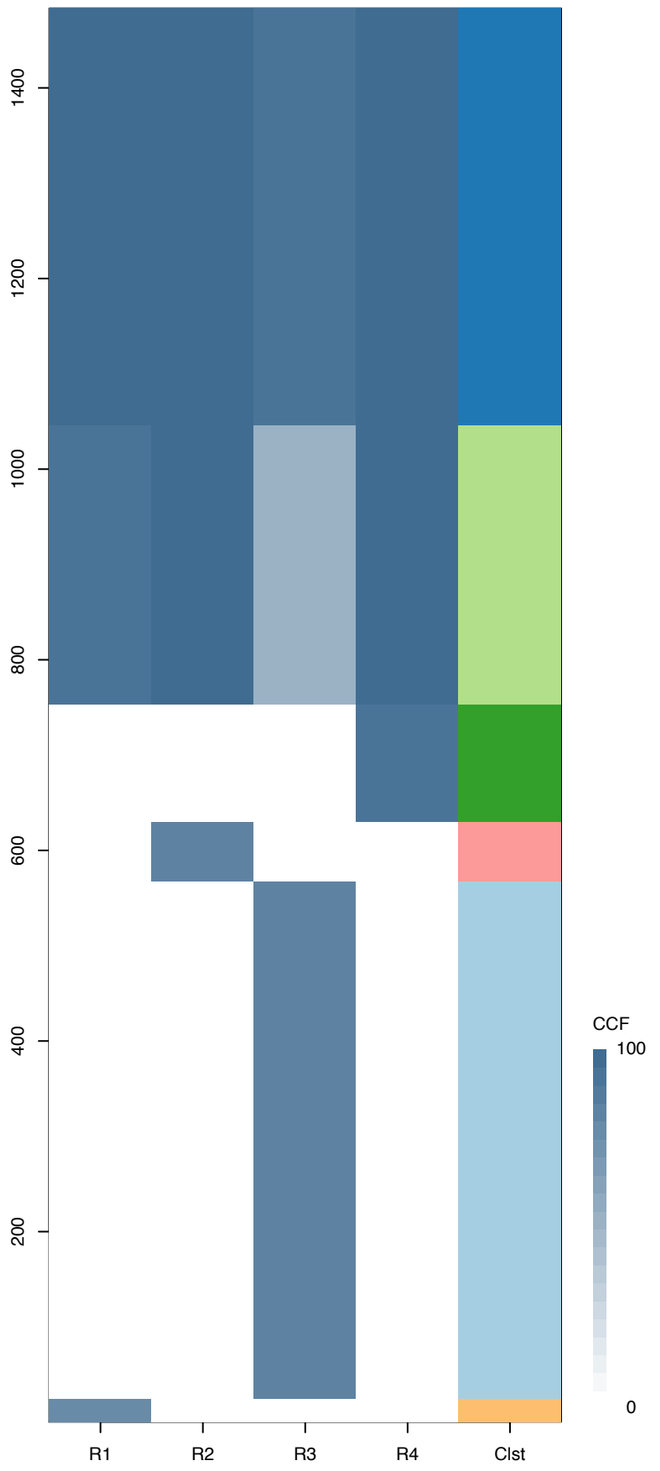
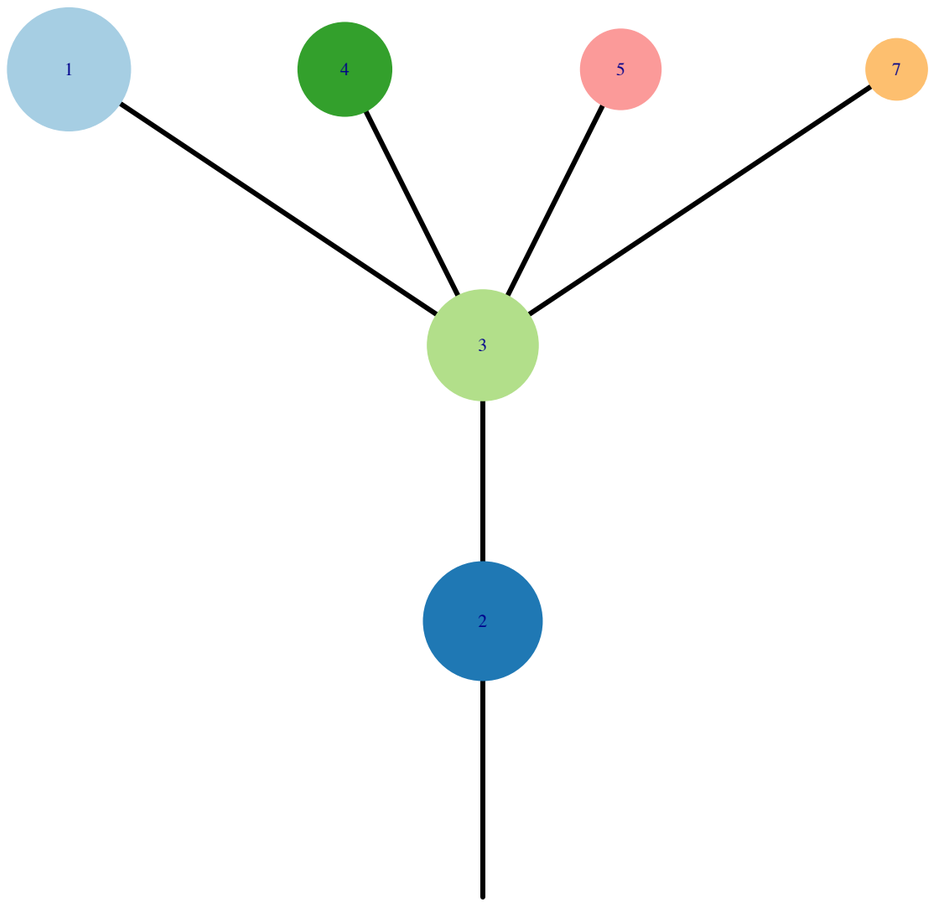


Fig.S12E



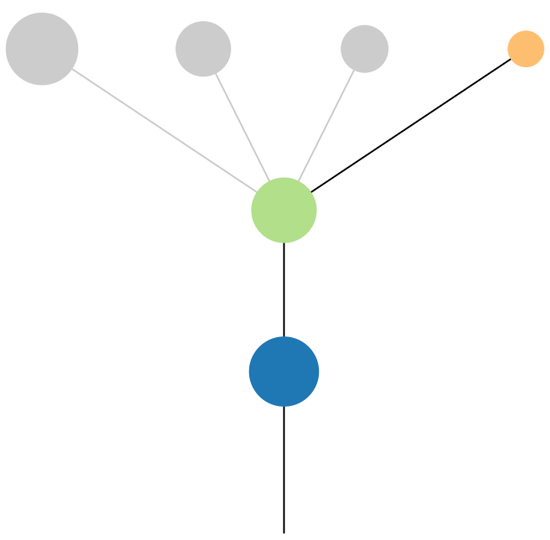
CRUK0005_B



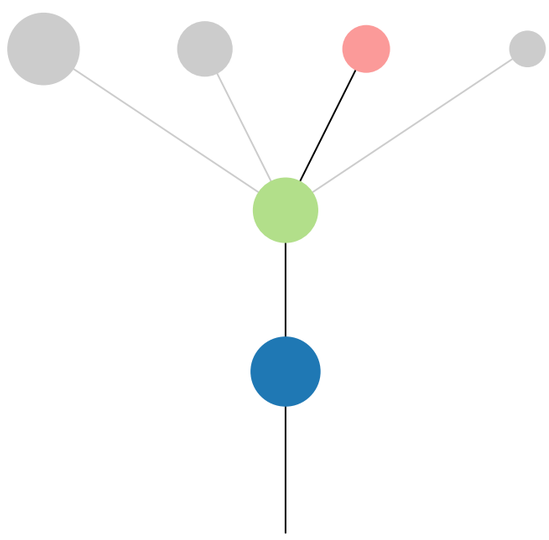
Histology:Adeno, Age:76, PackYears:55, Size:43
Stage:3a, Gender:Female, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
NRAS	1	1p13.2	SNV
BRE	2	2p23.2	Amp
PASK	2	2q37.3	SNV
TERT	2	5p15.33	Amp
IL7R	2	5p13.2	Amp
LIFR	2	5p13.1	Amp
PMS2	2	7p22.1	SNV
BRAF	2	7q34	SNV
CDK4	2	12q14.1	Amp
CMTR2	2	16q22.2	SNV
TP53	2	17p13.1	SNV
LSM14A	4	19q13.11	Amp
CLTCL1	?	22q11.21	Amp
SEPT5	?	22q11.21	Amp
ATRX	?		SNV

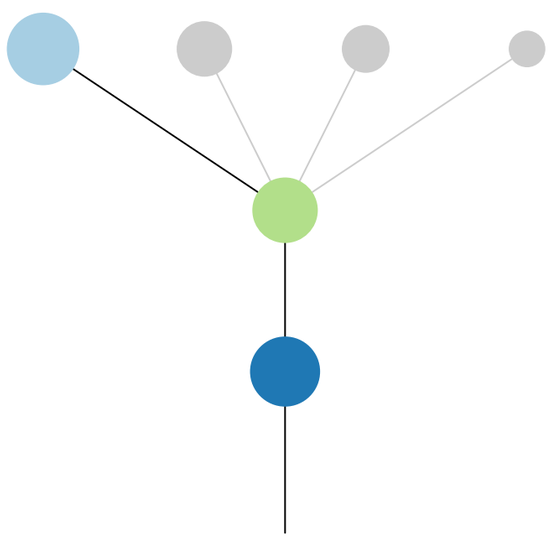
R1



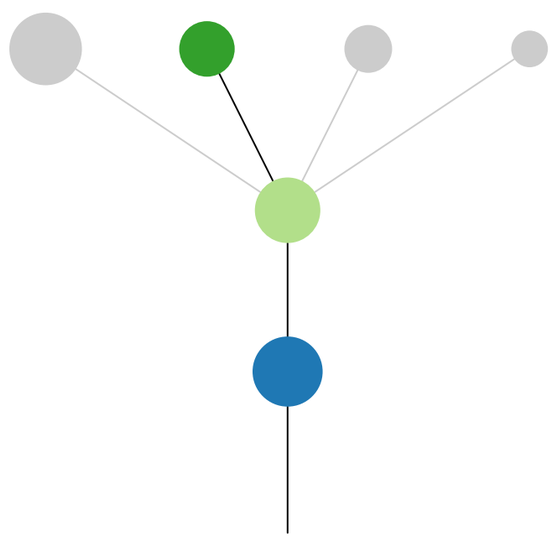
R2

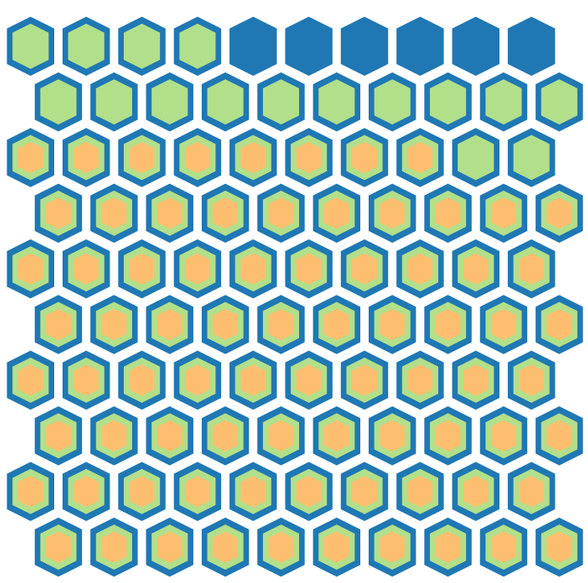


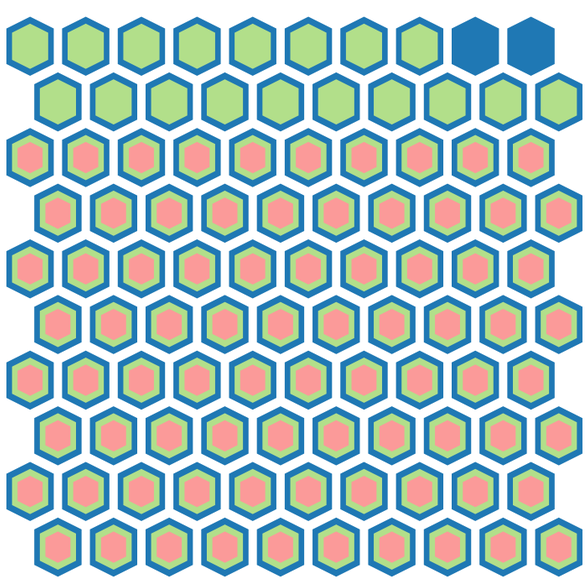
R3

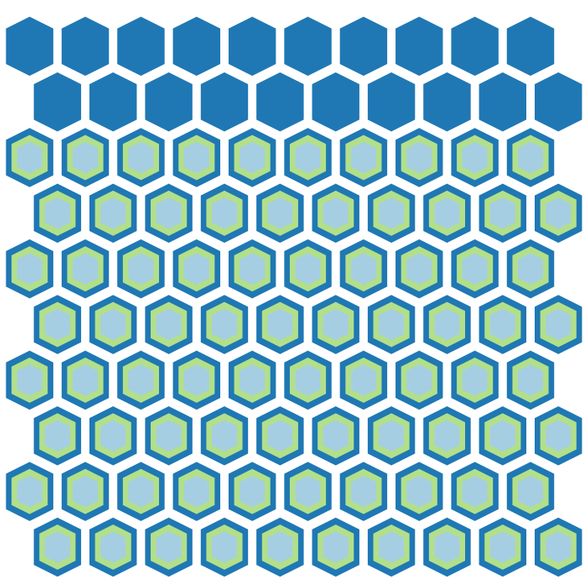


R4









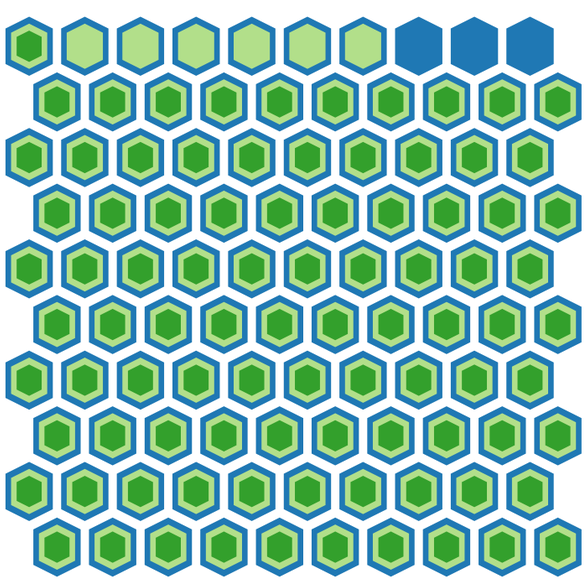
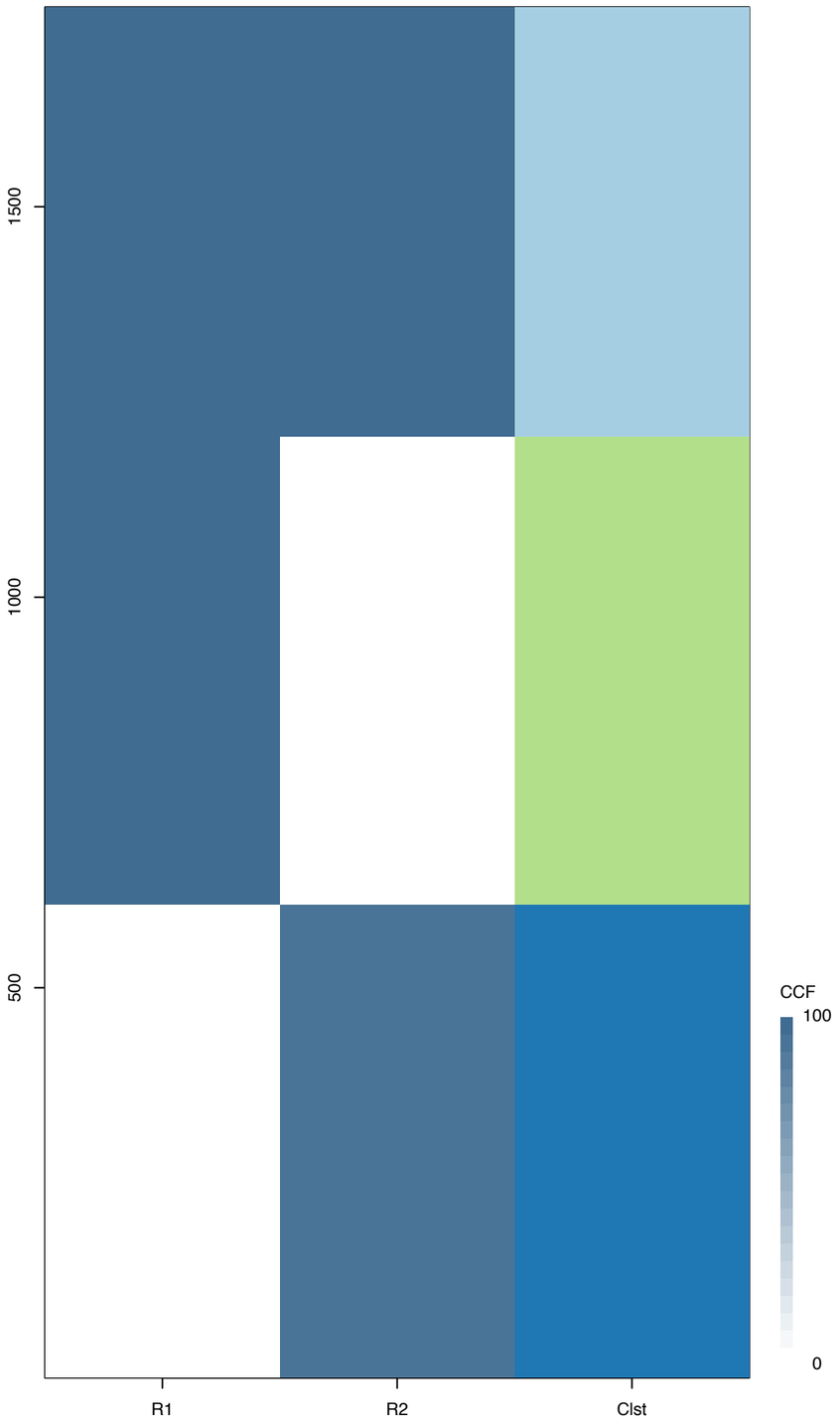
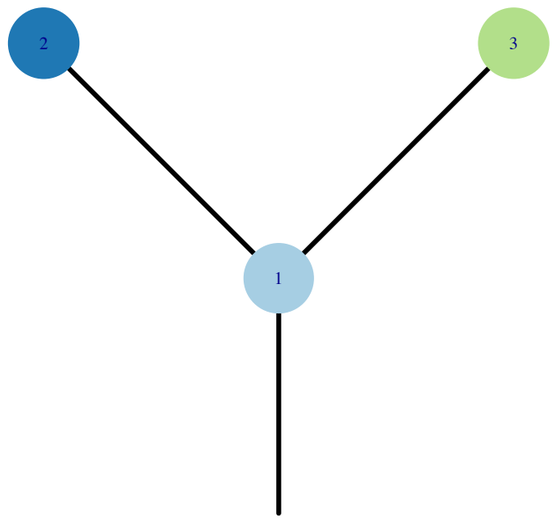


Fig.S12F

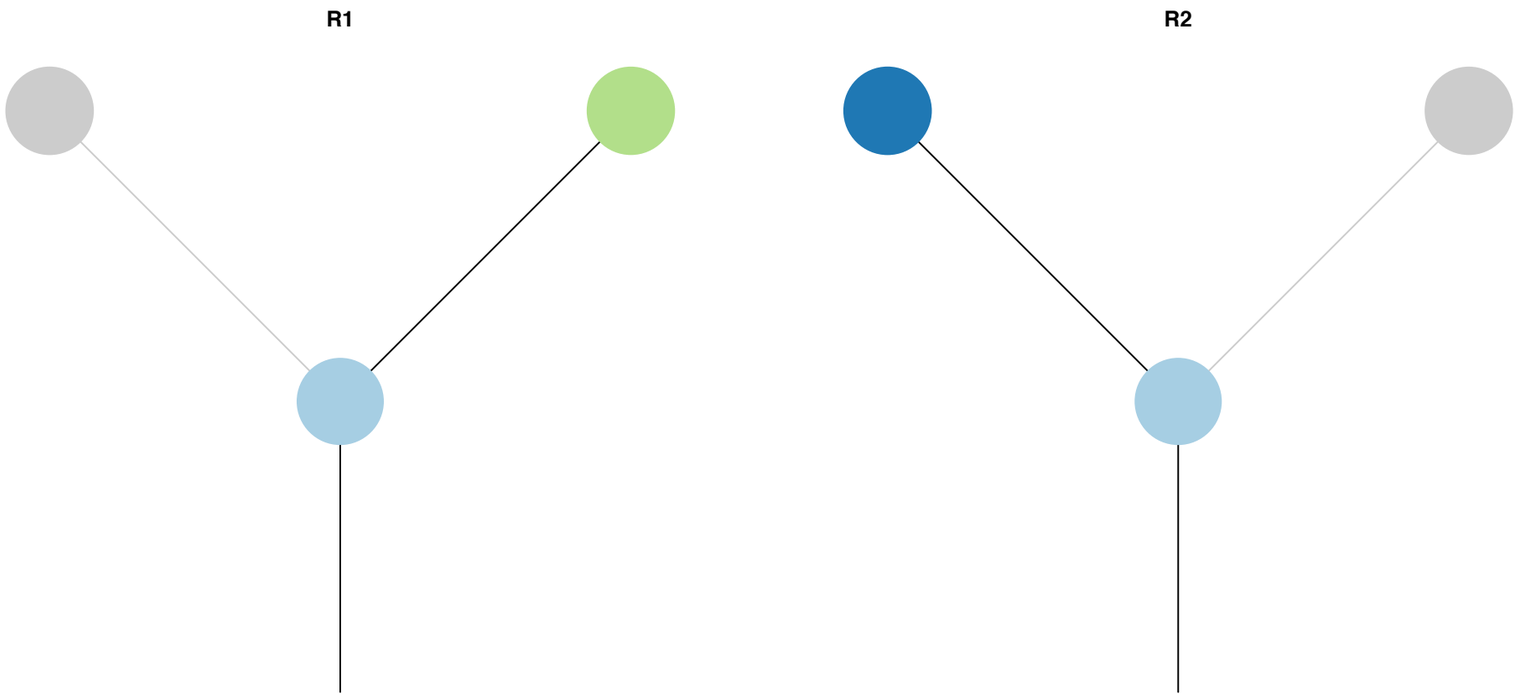


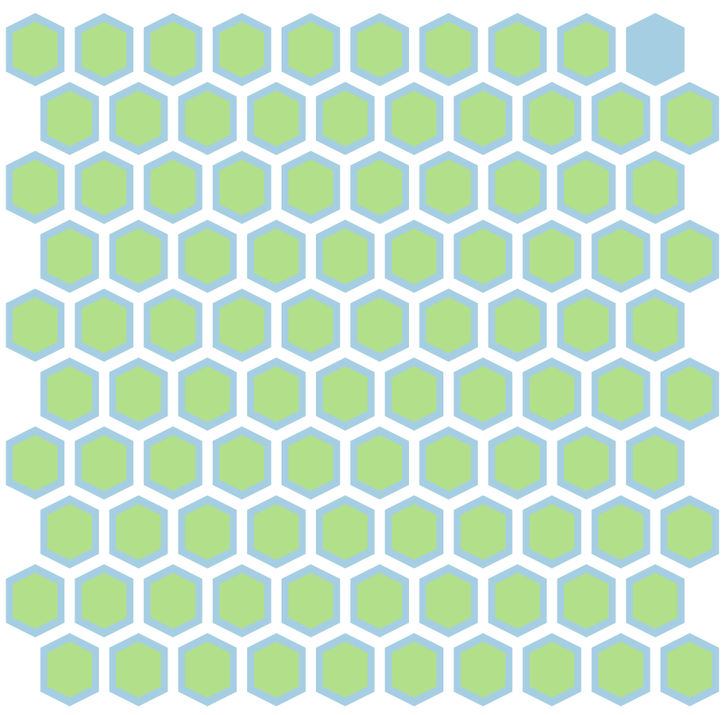
CRUK0006



Histology:Adeno, Age:67, PackYears:NA, Size:18
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
TERT	1	5p15.33	Amp
TP53	1	17p13.1	SNV
KEAP1	1	19p13.2	SNV
PLXNB2	1	22q13.33	SNV
MUTYH	2	1p34.1	SNV
MAP3K1	2	5q11.2	SNV
FANCC	?	9q22.32	SNV
KDM5C	?		SNV
ATRX	?		SNV





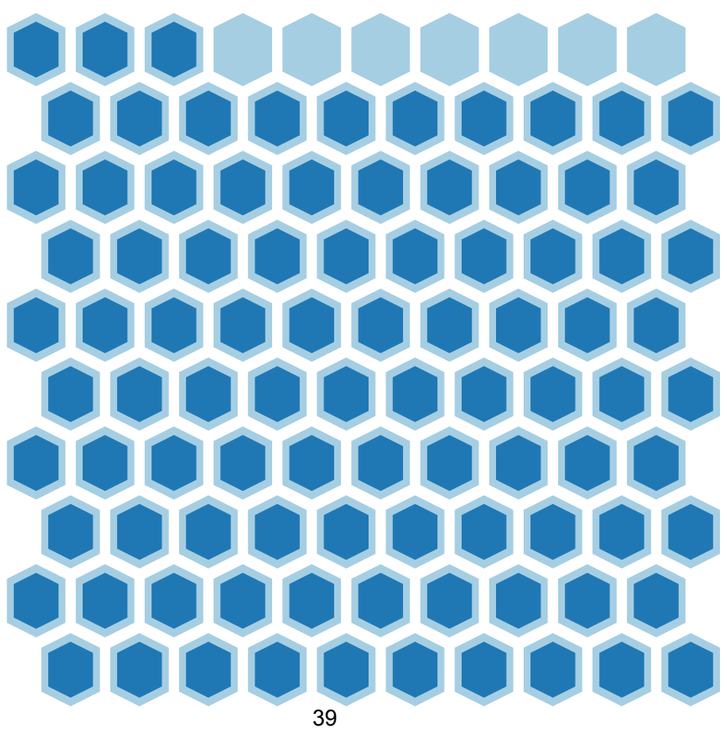
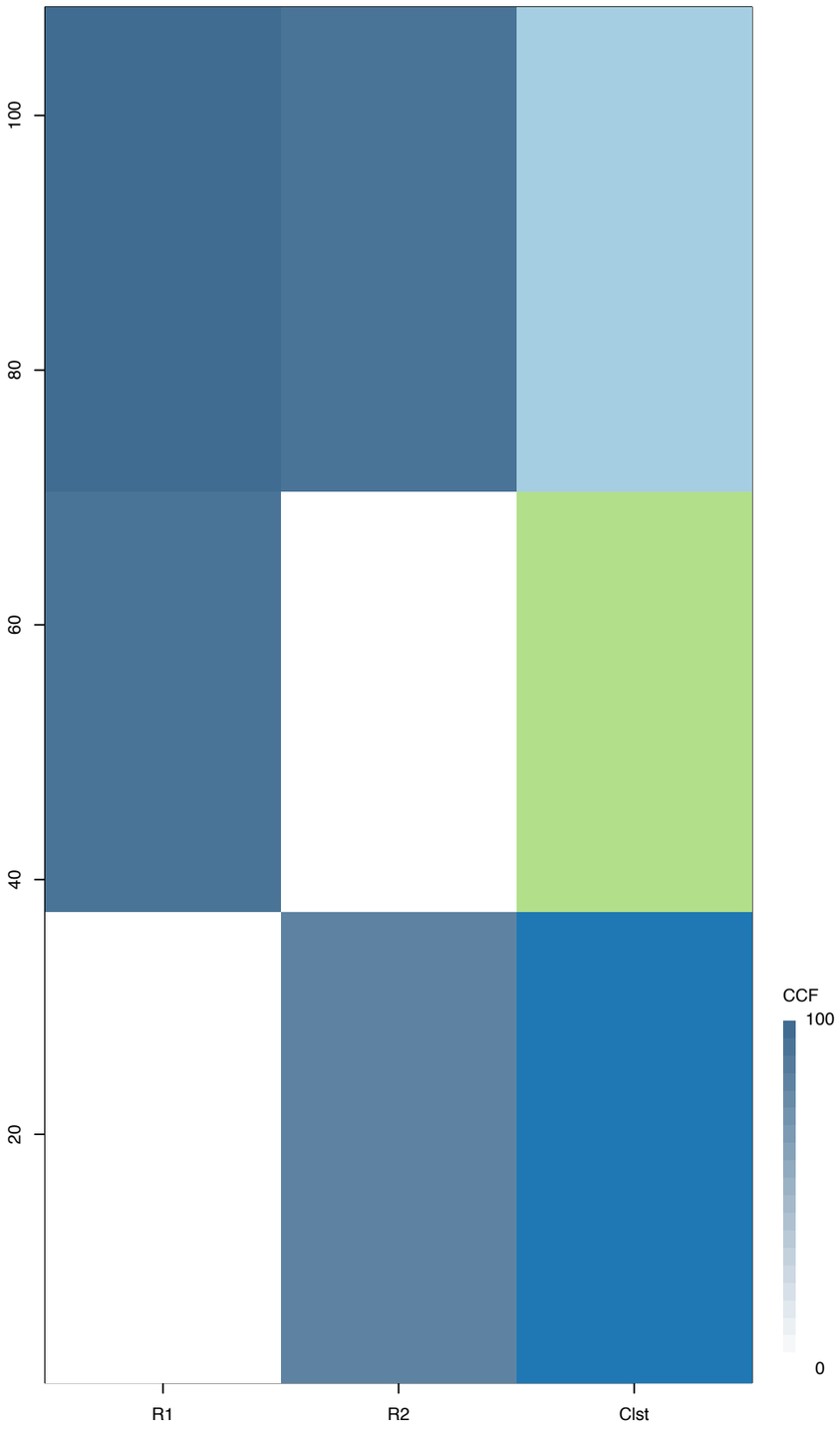
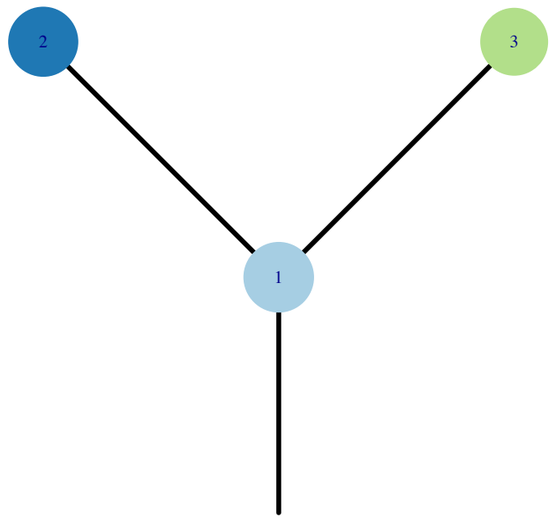


Fig.S12G



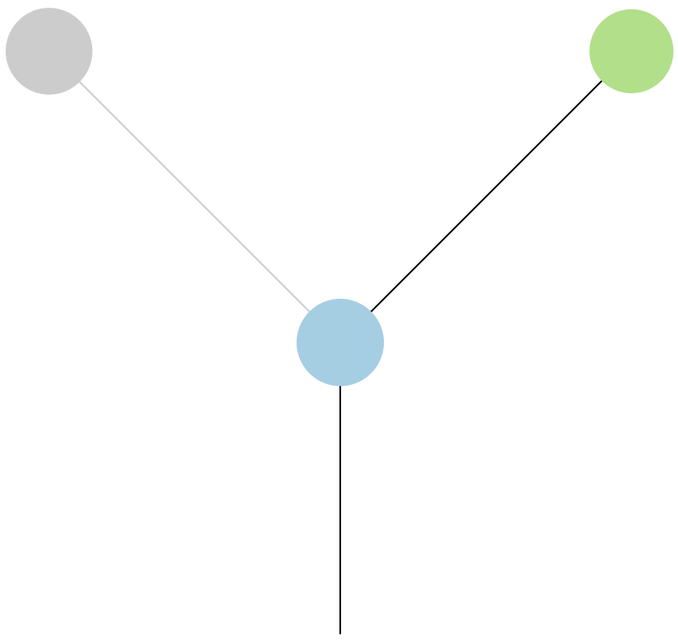
CRUK0007



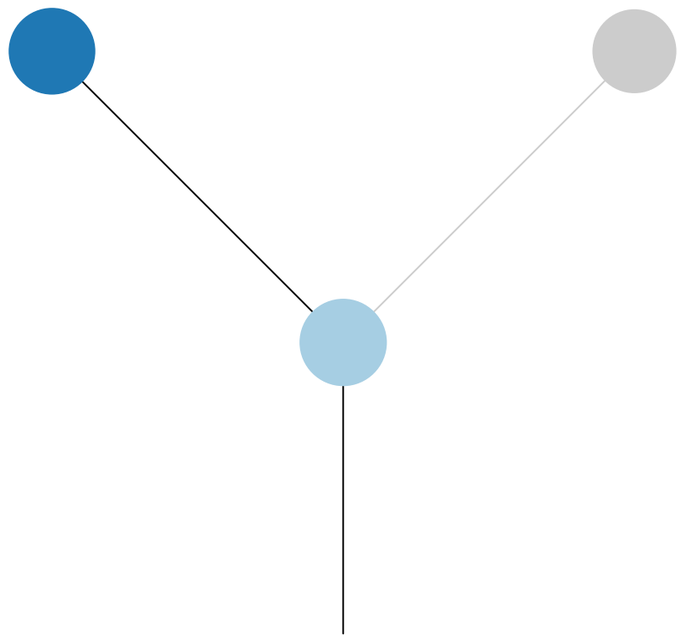
Histology:Adeno, Age:70, PackYears:0.45, Size:18
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

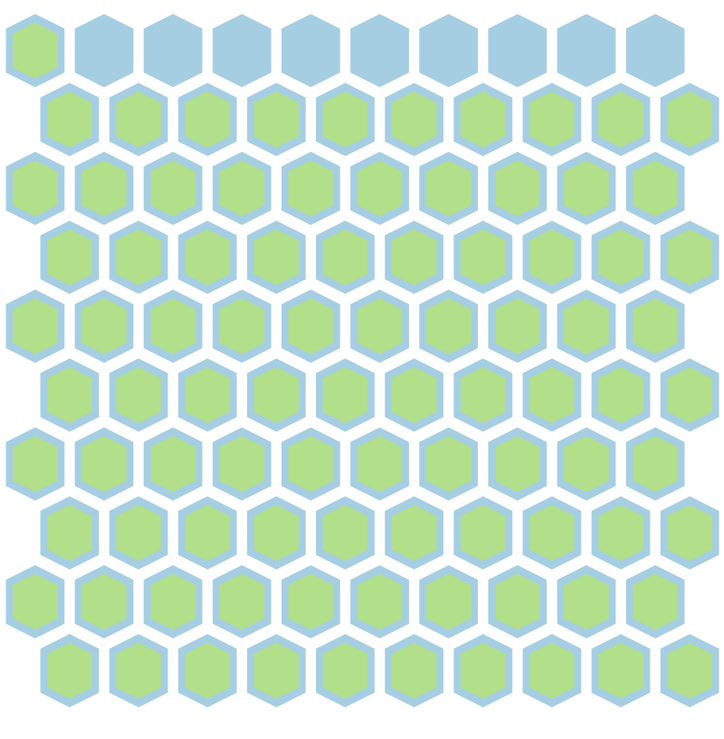
Gene	Cluster	Cytoband	Type
PIK3CA	1	3q26.32	SNV
EGFR	1	7p11.2	SNV
SGK223	1	8p23.1	Amp

R1



R2





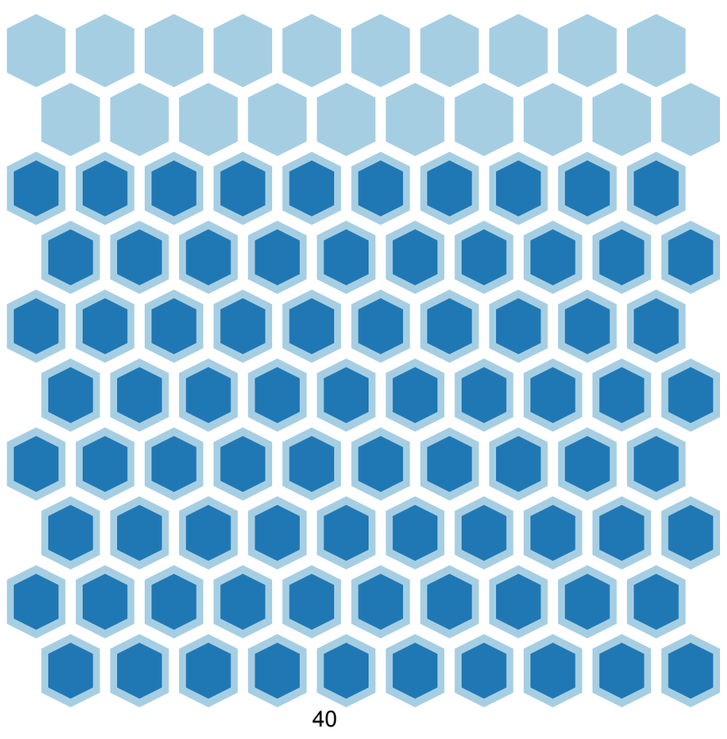
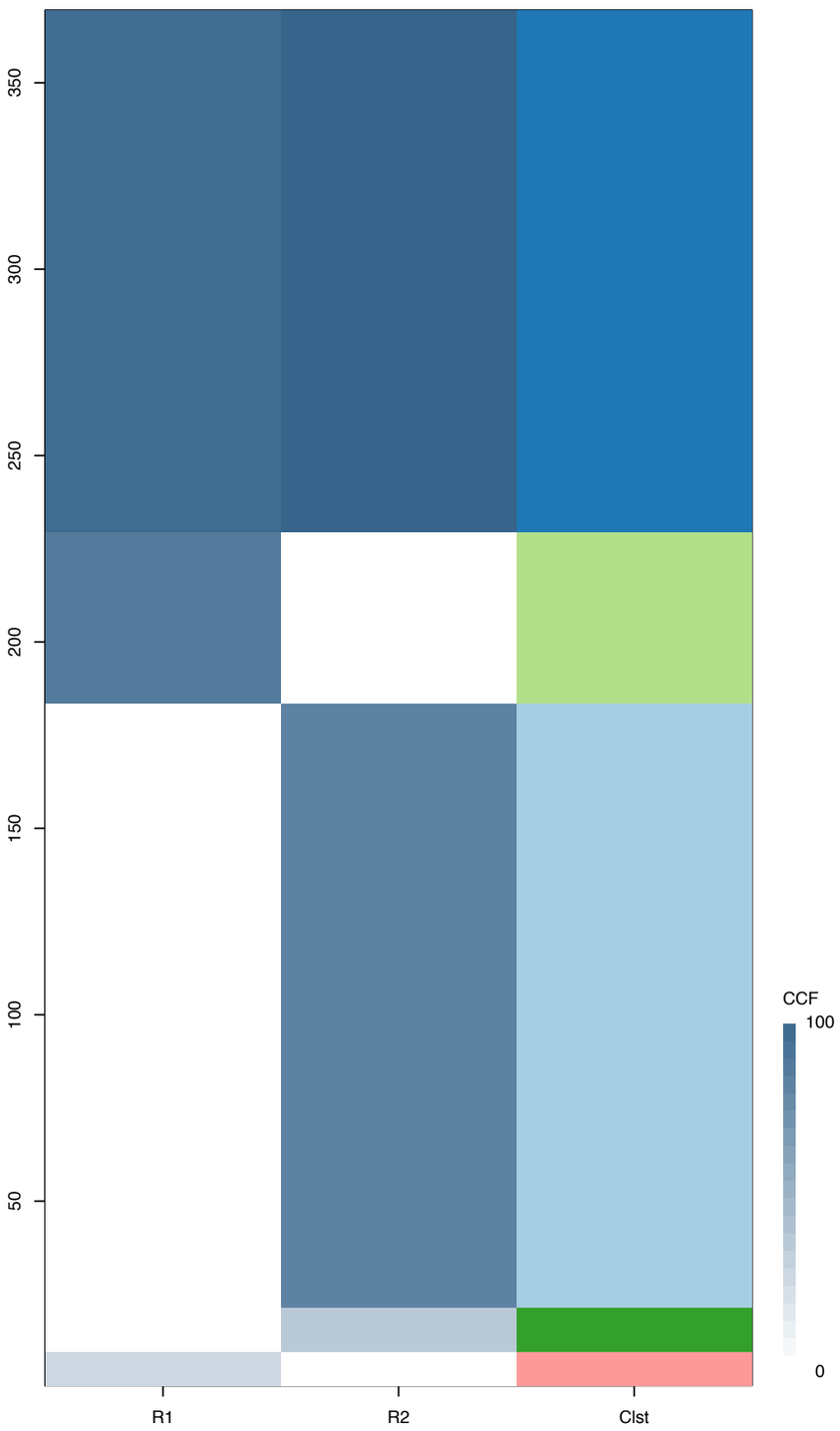
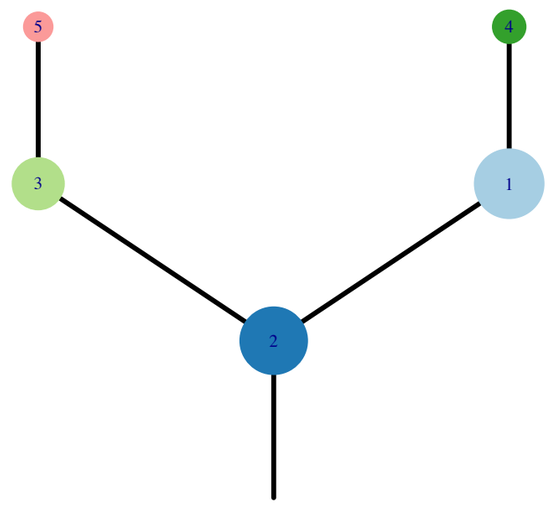


Fig.S12H



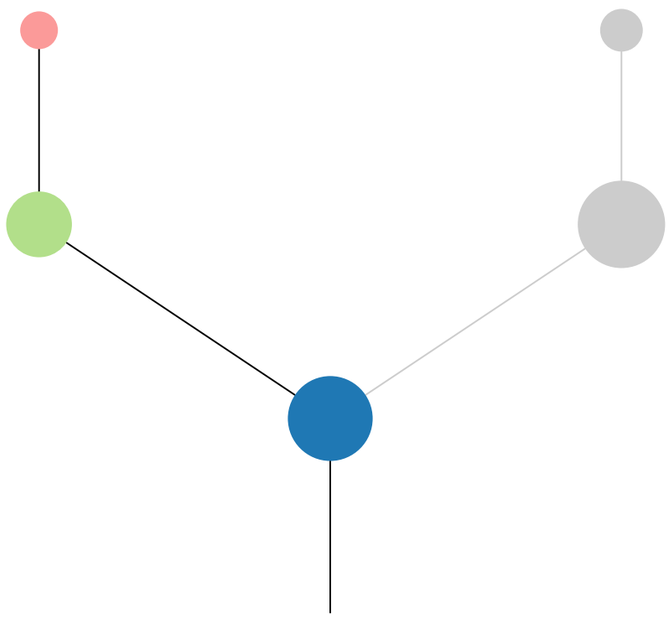
CRUK0008



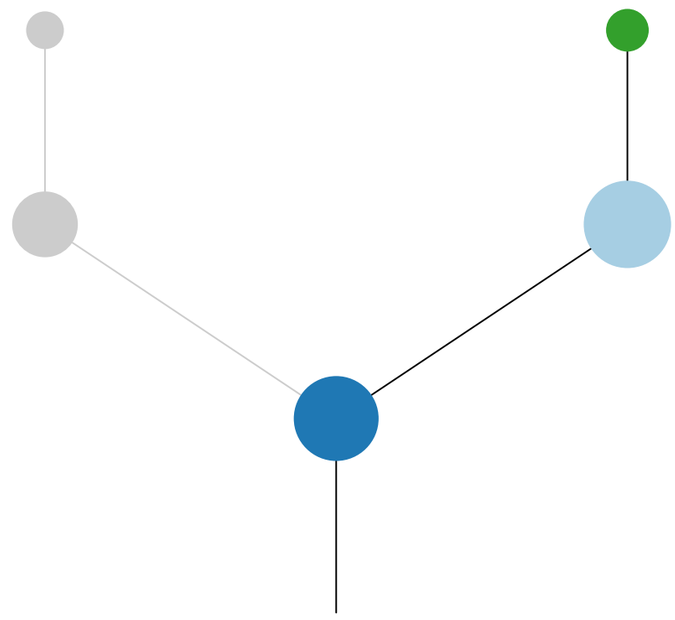
Histology:Adeno, Age:73, PackYears:74, Size:25
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

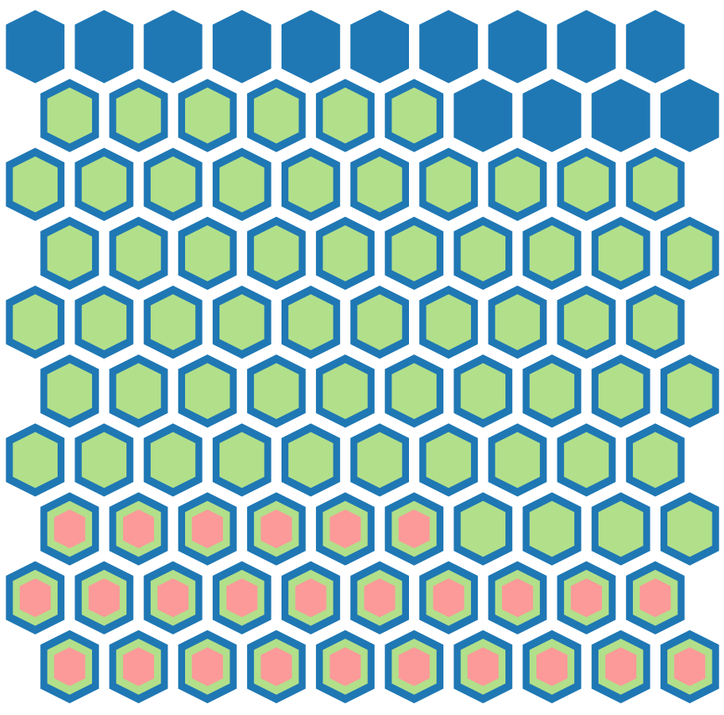
Gene	Cluster	Cytoband	Type
RIT1	2	1q22	SNV
PRDM1	2	6q21	SNV
TCEA1	2	8q11.23	Amp
PLAG1	2	8q12.1	Amp
CHCHD7	2	8q12.1	Amp
NCOA2	2	8q13.3	Amp
HEY1	2	8q21.13	Amp
CNBD1	2	8q21.3	Amp
RUNX1T1	2	8q21.3	Amp
COX6C	2	8q22.2	Amp
RSPO2	2	8q23.1	Amp
EIF3E	2	8q23.1	Amp
MYC	2	8q24.21	Amp
NDRG1	2	8q24.22	Amp
STK11	2	19p13.3	SNV
KEAP1	2	19p13.2	SNV
U2AF1	2	21q22.3	SNV
ARID2	3	12q12	SNV

R1



R2





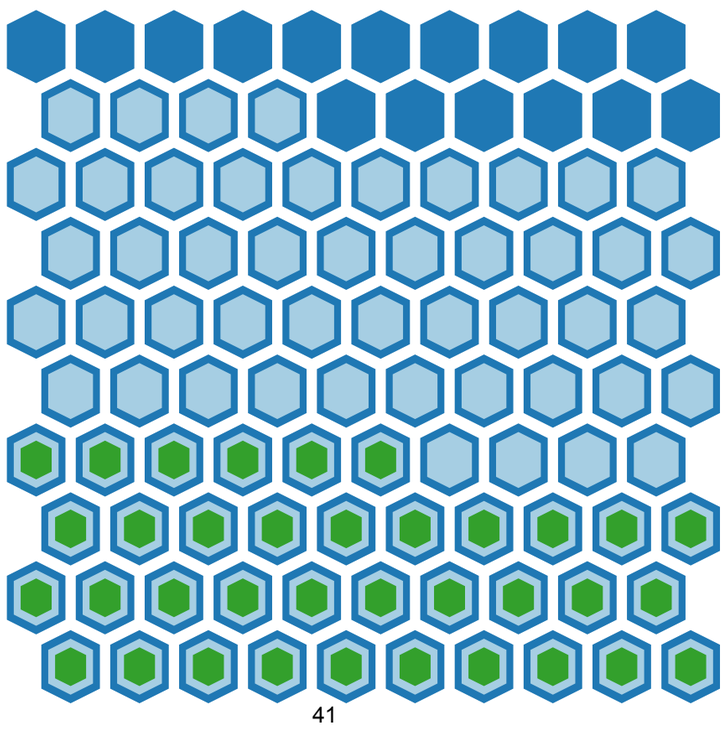
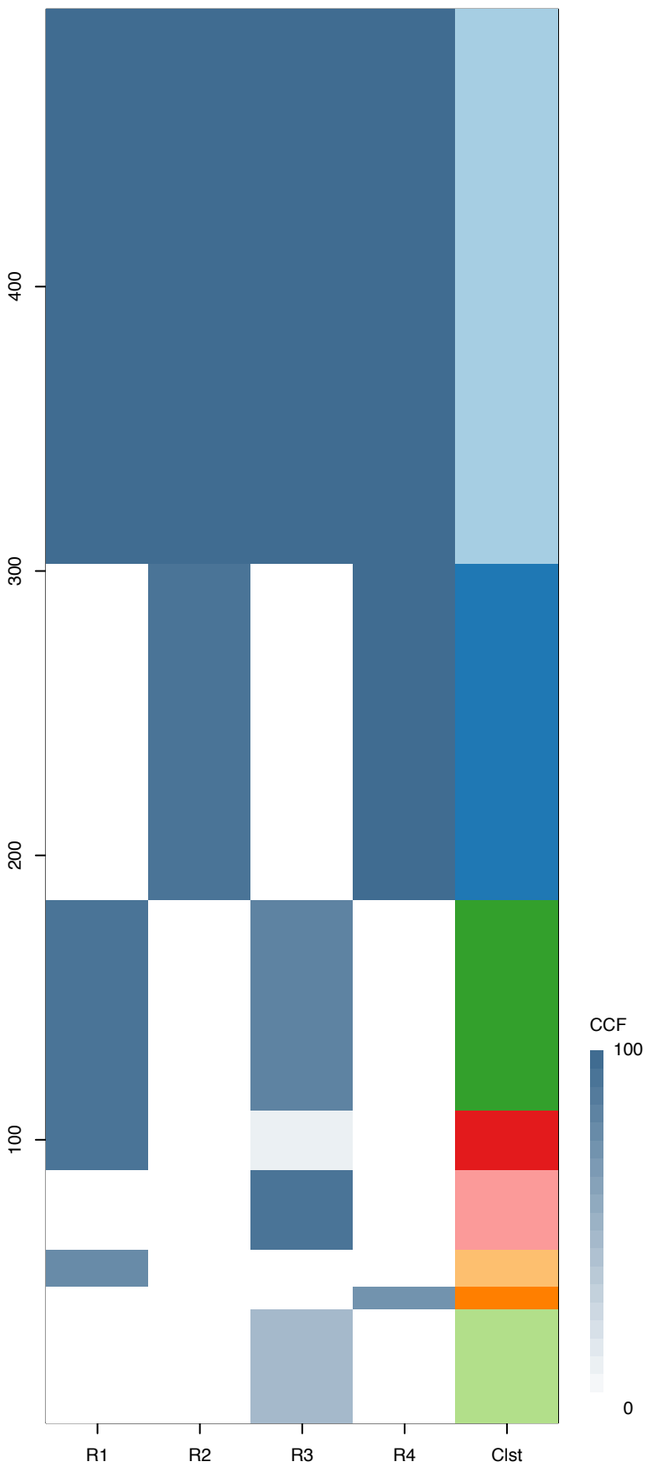
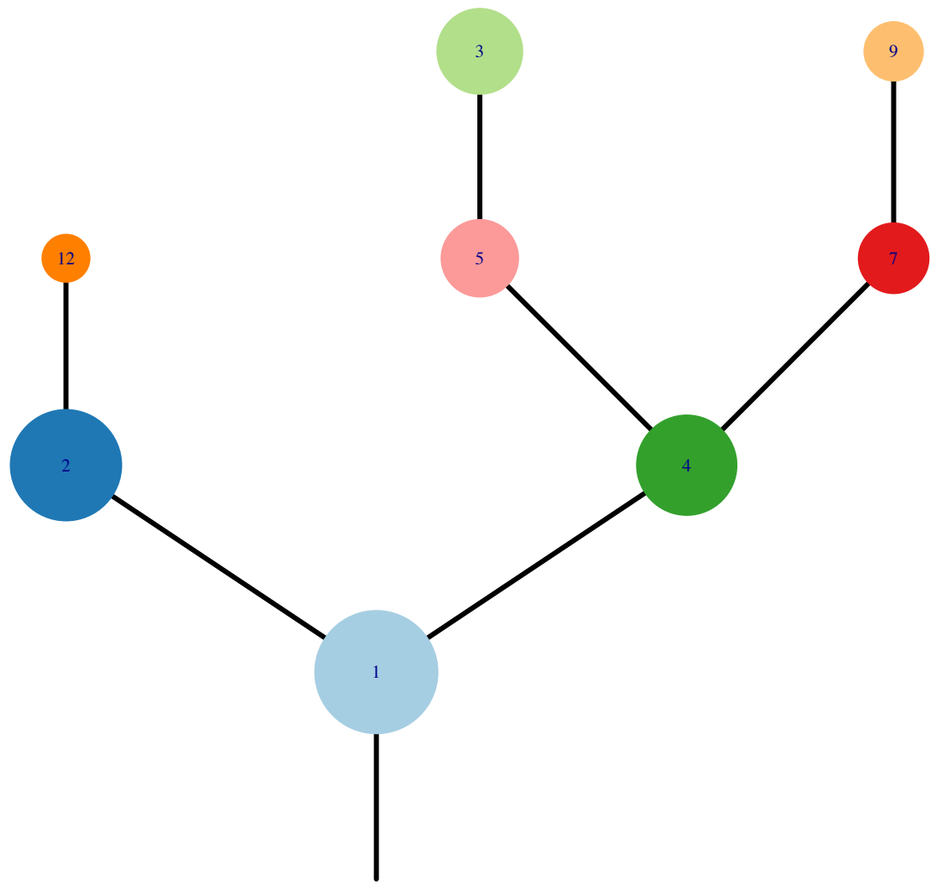


Fig.S12l



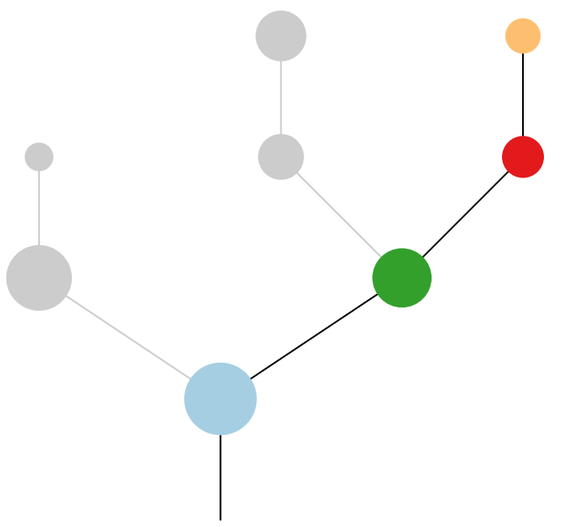
CRUK0009



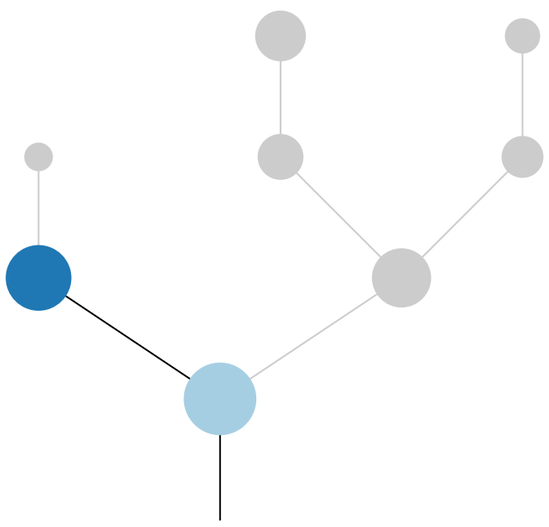
Histology:Adeno, Age:56, PackYears:10.25, Size:65
Stage:2b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
NFE2L2	1	2q31.2	Amp
ELN	1	7q11.23	Amp
HIP1	1	7q11.23	Amp
AKAP9	1	7q21.2	Amp
CDK6	1	7q21.2	Amp
TRRAP	1	7q22.1	Amp
MET	1	7q31.2	Amp
SND1	1	7q32.1	Amp
SMO	1	7q32.1	Amp
CREB3L2	1	7q33	Amp
TRIM24	1	7q33	Amp
KIAA1549	1	7q34	Amp
BRAF	1	7q34	Amp
BRAF	1	7q34	SNV
KMT2C	1	7q36.1	SNV
TP53	1	17p13.1	SNV
ARHGAP35	1	19q13.32	SNV
RHOH	2	4p14	Amp
TERT	2	5p15.33	Amp
RHOH	4	4p14	Amp

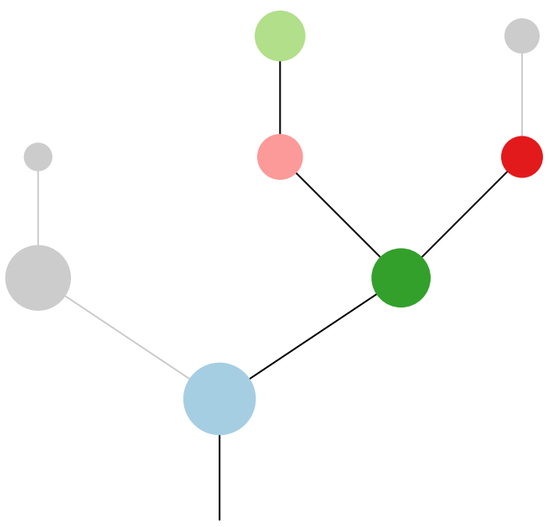
R1



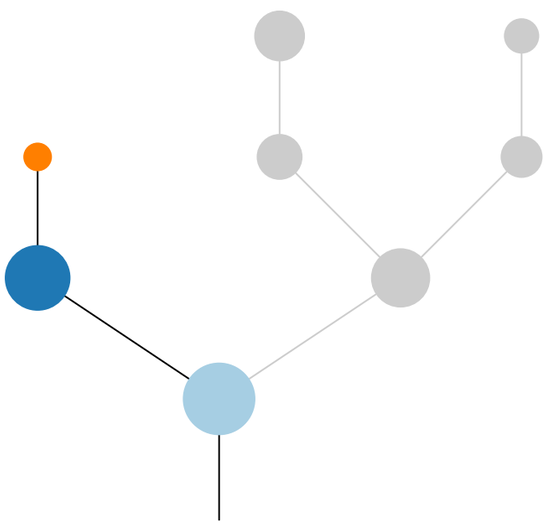
R2

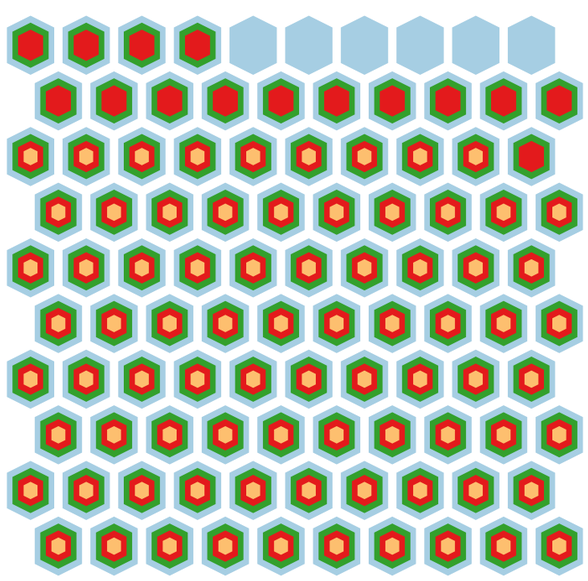


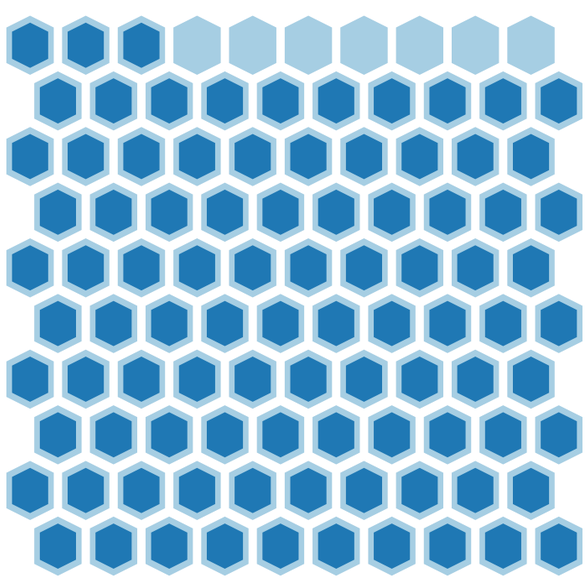
R3

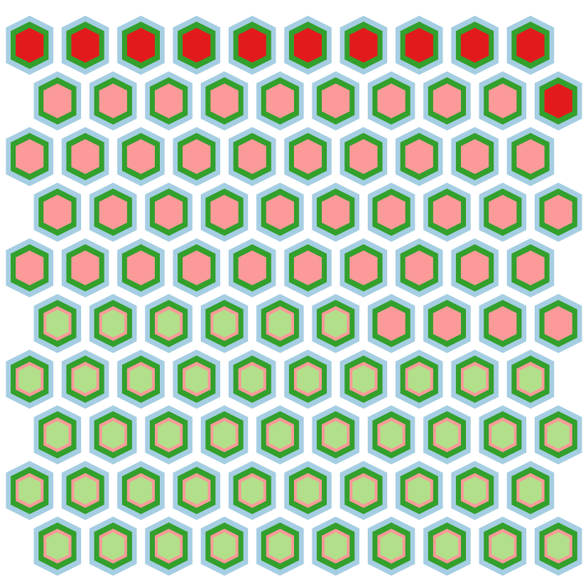


R4









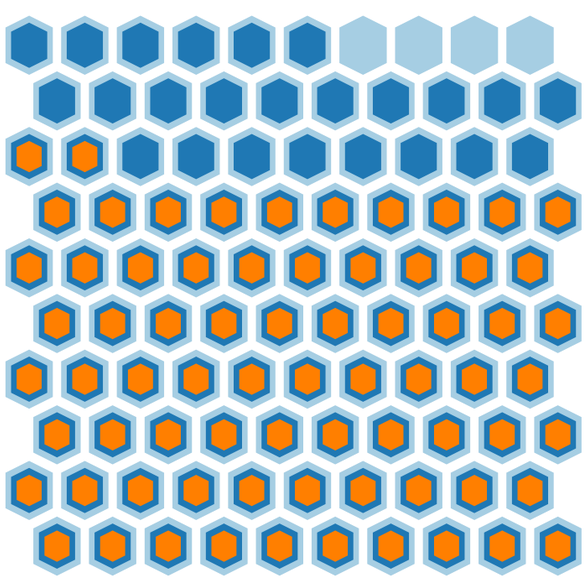
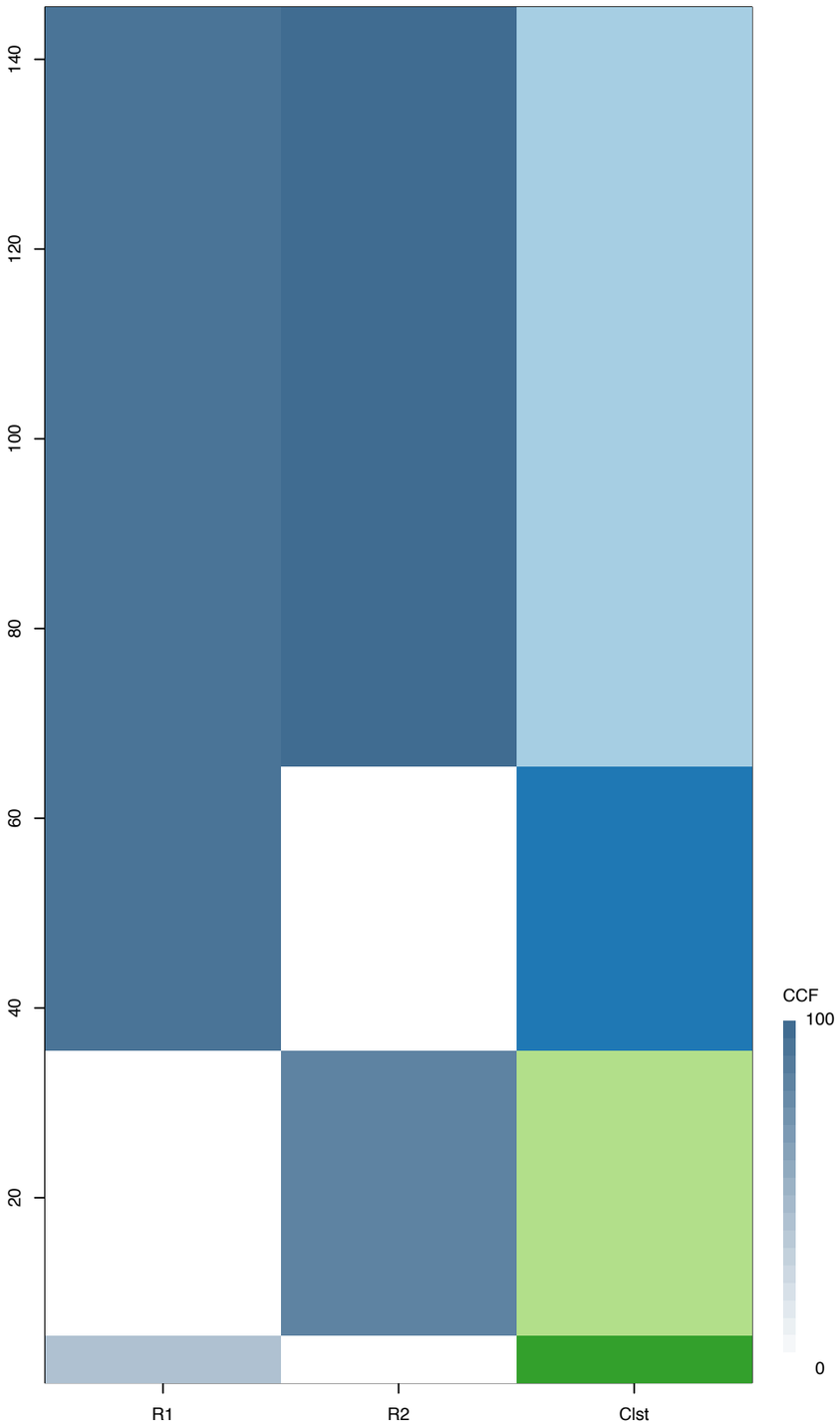
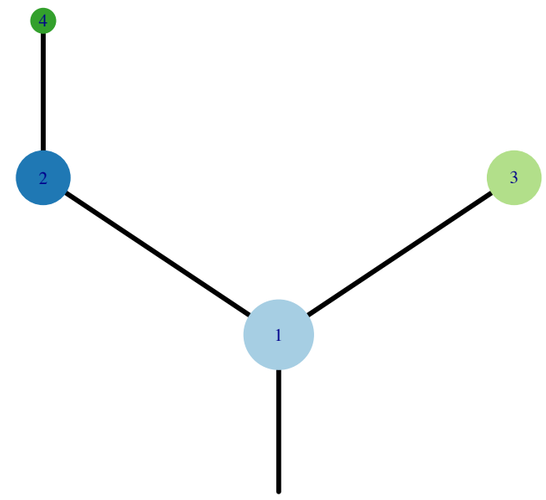


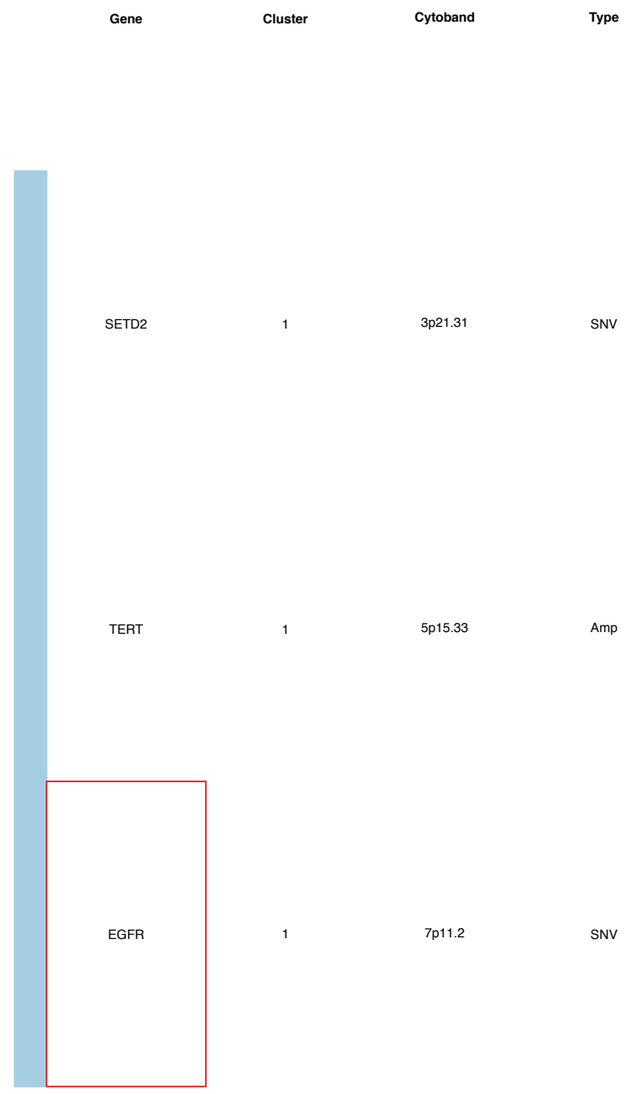
Fig.S12J



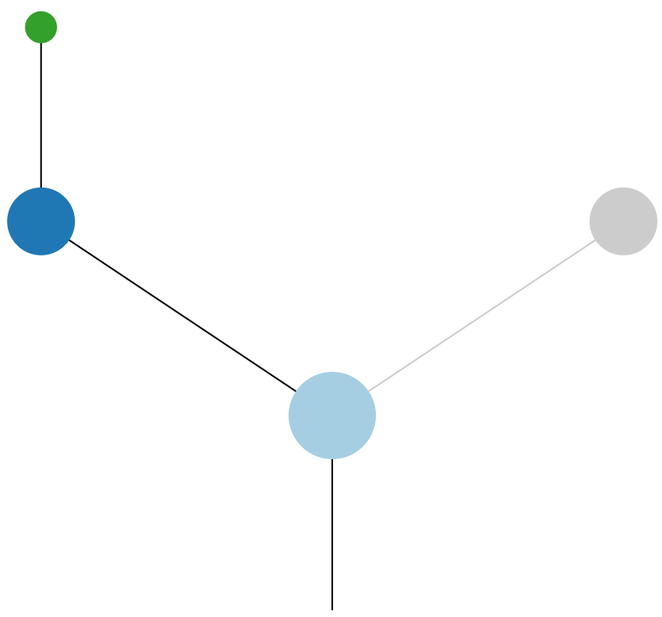
CRUK0010



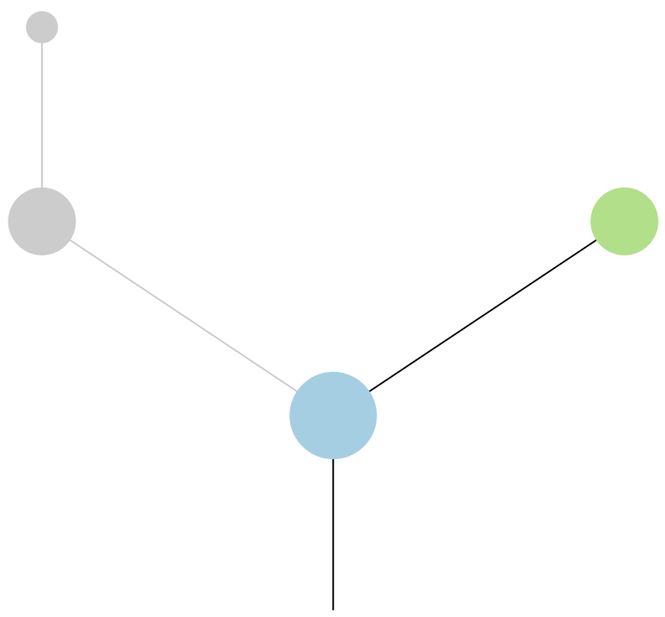
Histology:Adeno, Age:69, PackYears:NA, Size:32
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

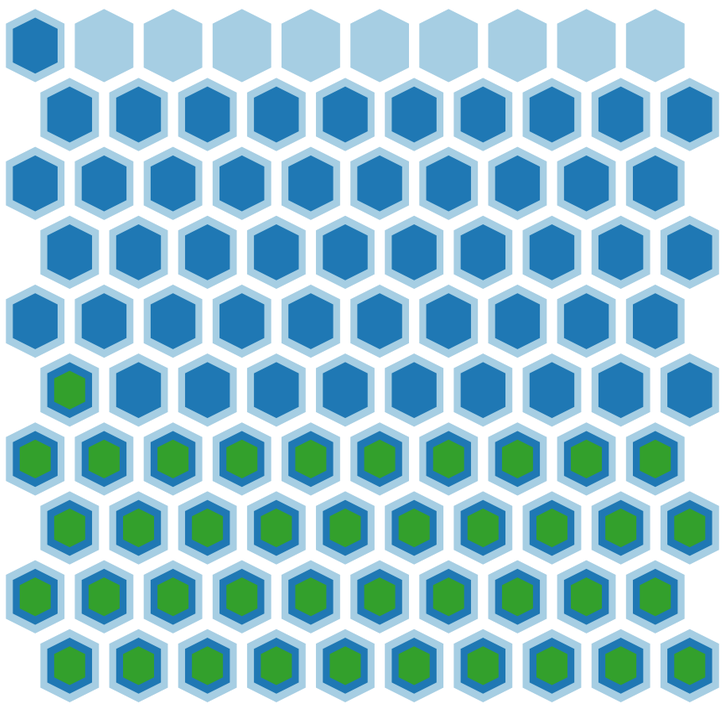


R1



R2





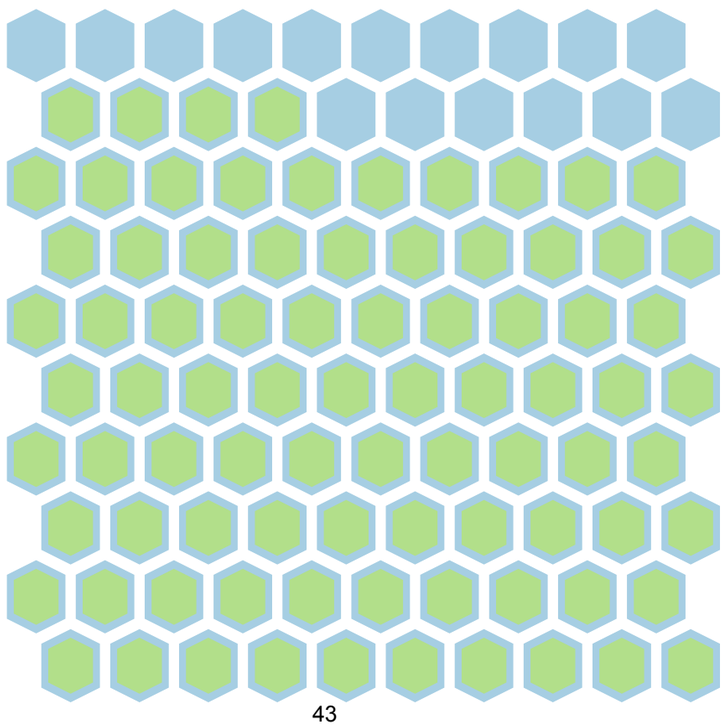
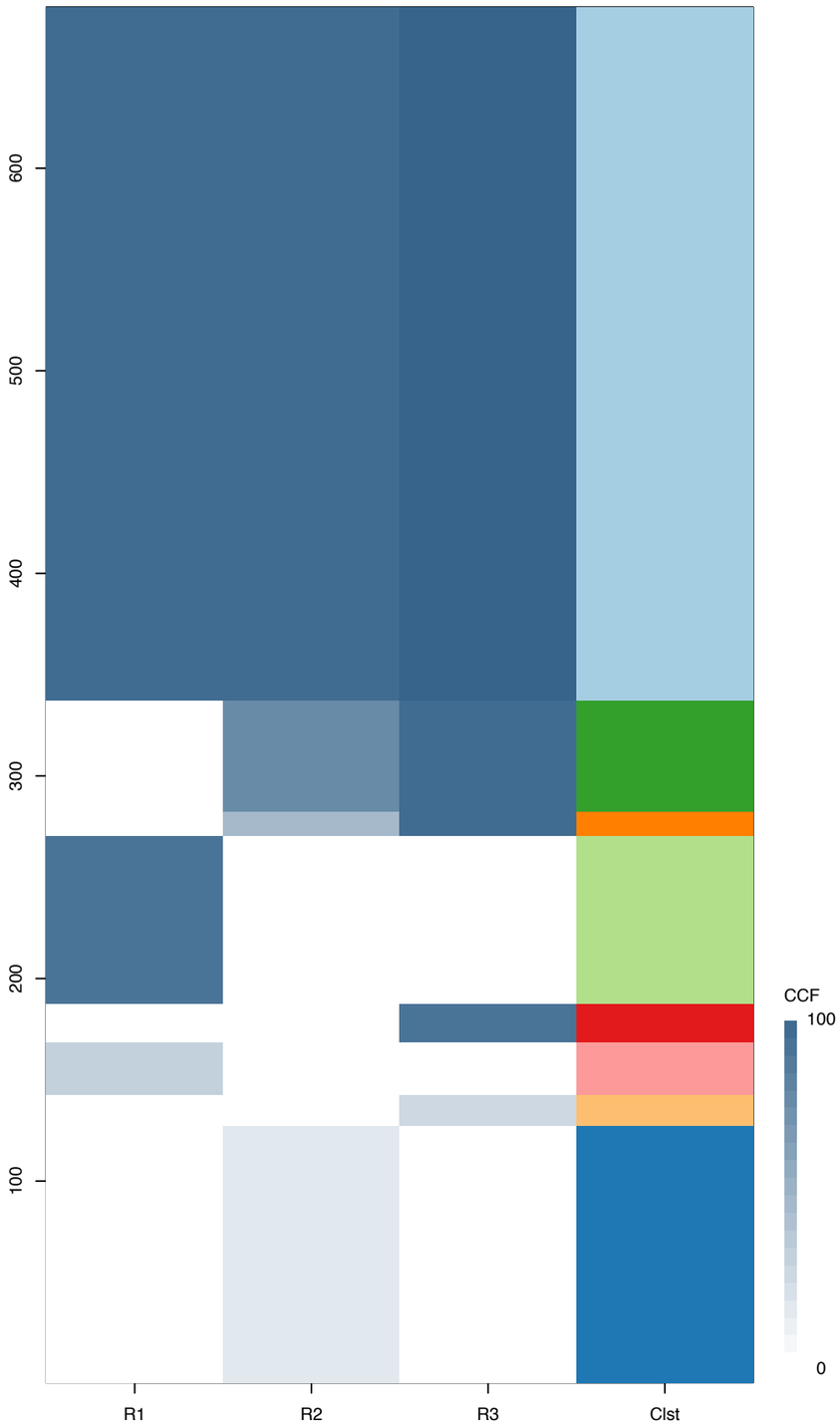
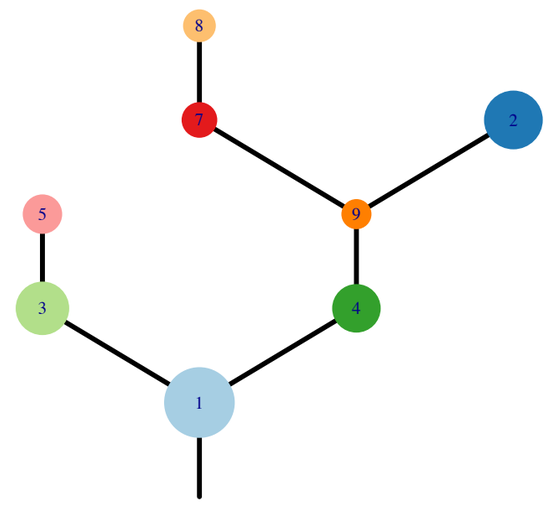


Fig.S12K



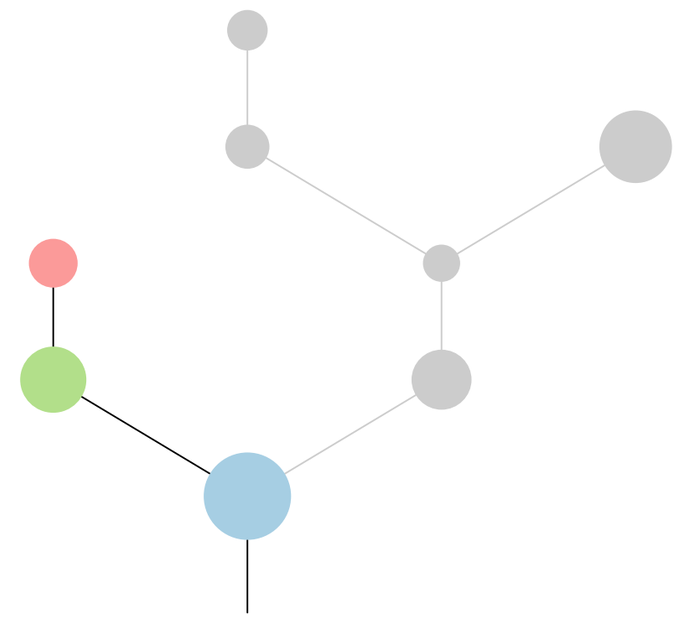
CRUK0011_A



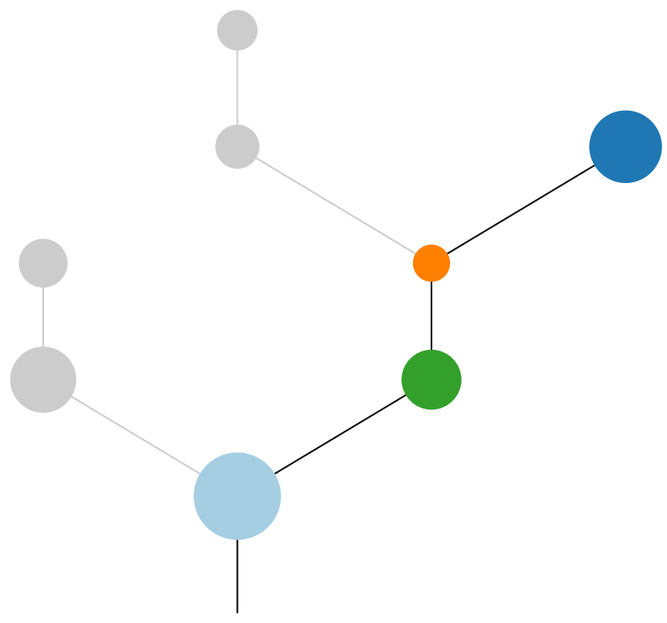
Gene	Cluster	Cytoband	Type
NBN	1	8q21.3	SNV
KLF6	1	10p15.1	SNV
KRAS	1	12p12.1	SNV
FLT4	3	5q35.3	SNV

Histology:Adeno, Age:72, PackYears:35, Size:40
Stage:1b, Gender:Female, GD:Subclonal GD, Recur:no

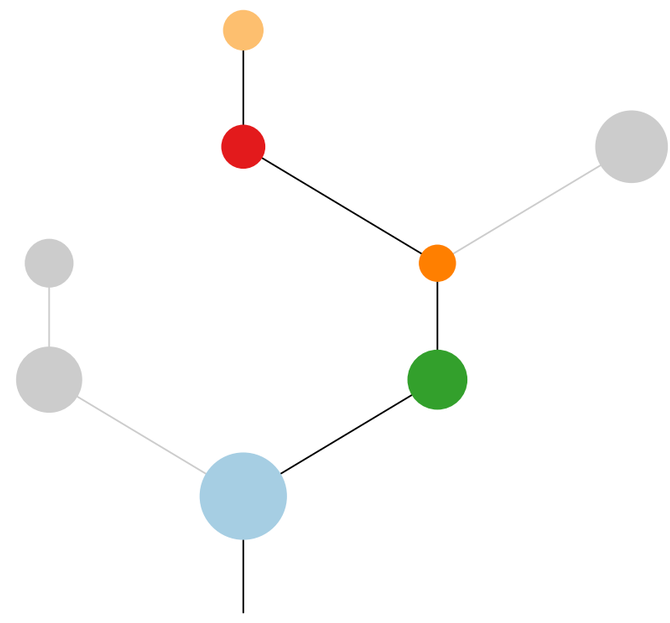
R1

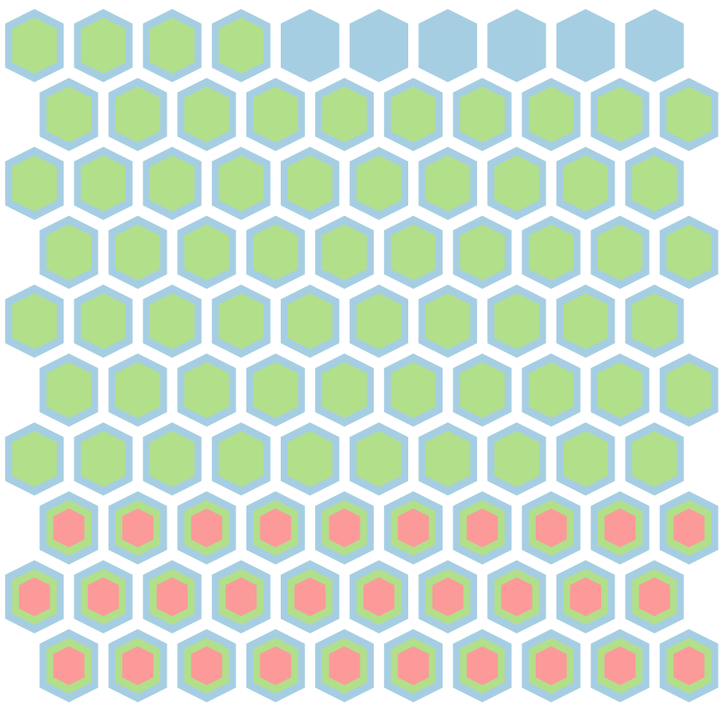


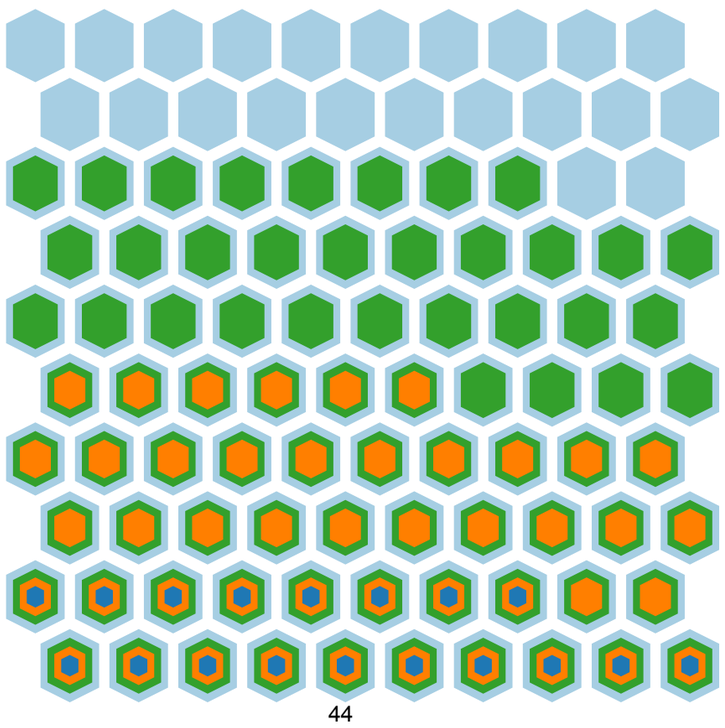
R2



R3







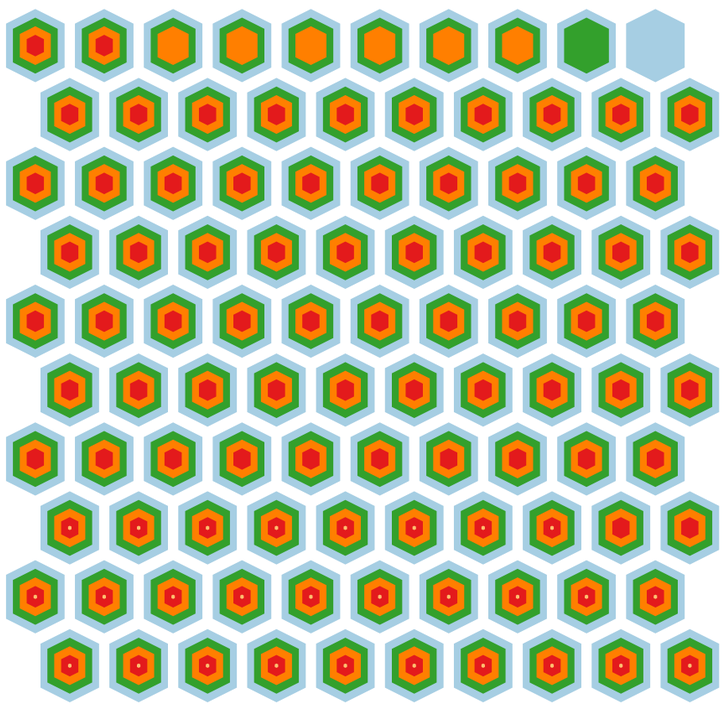
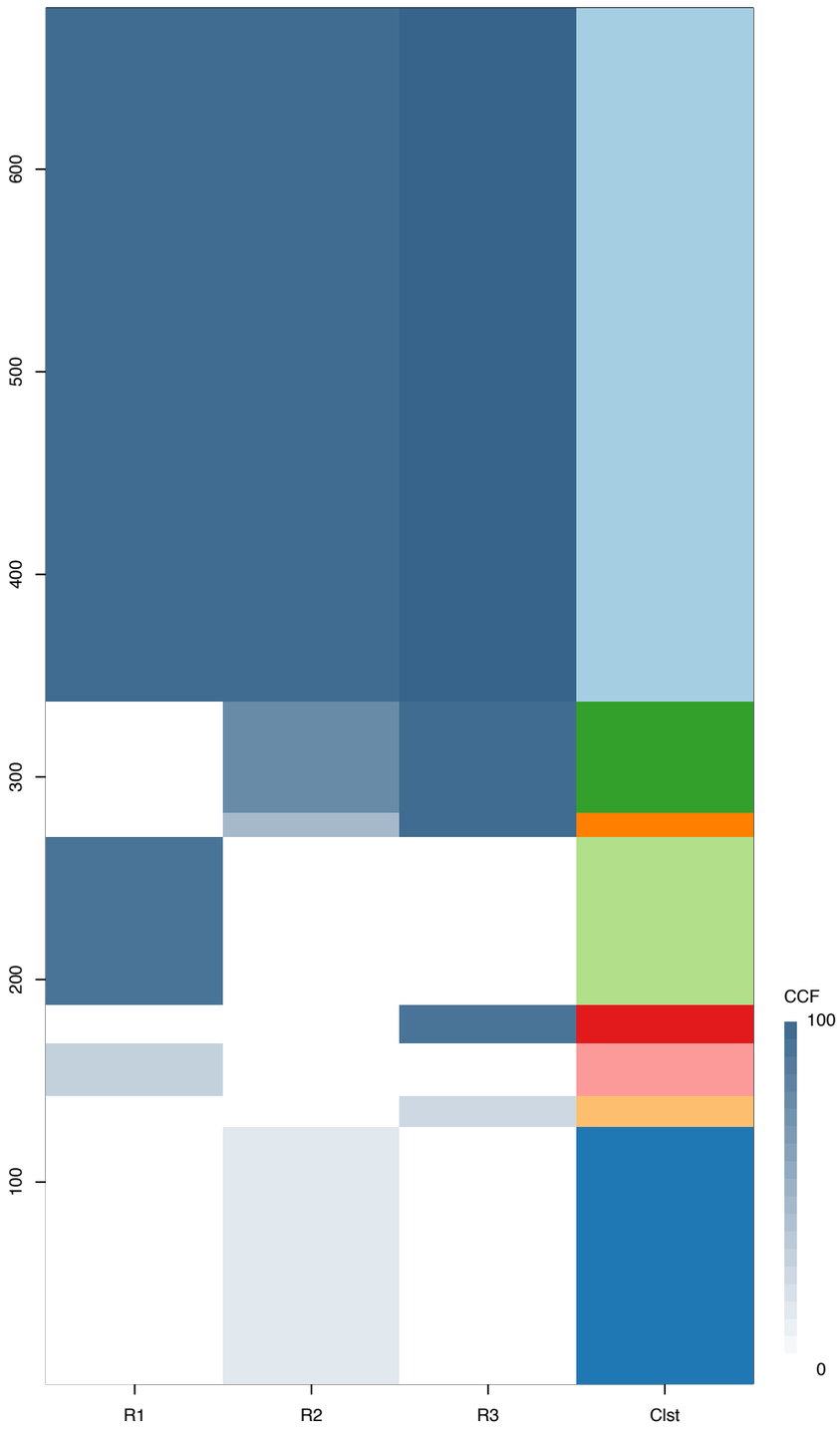
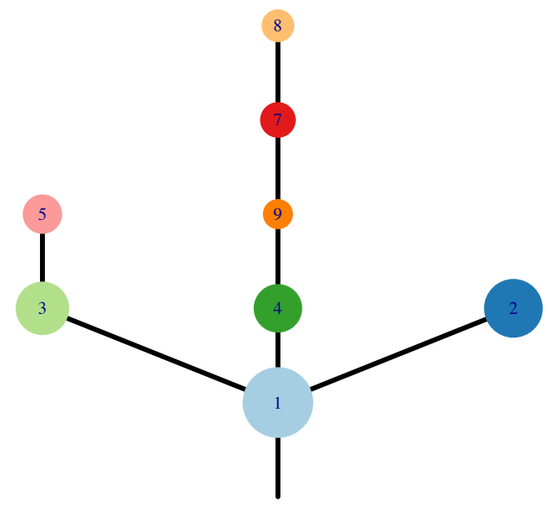


Fig.S12K



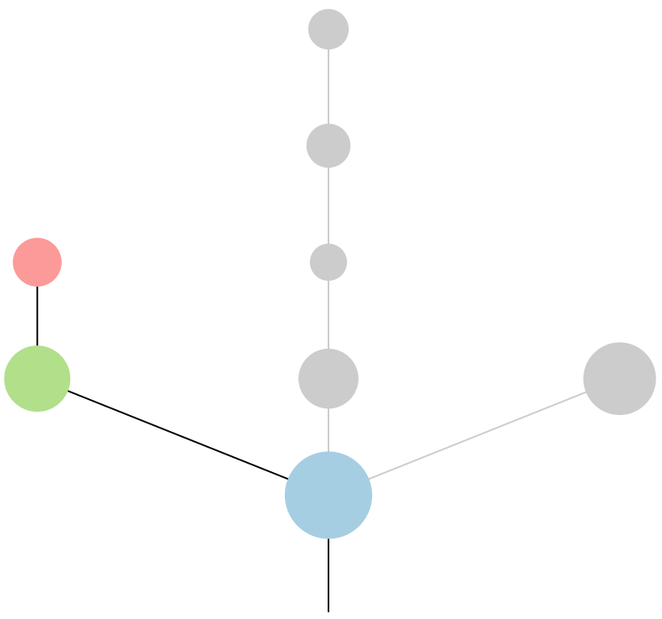
CRUK0011_B



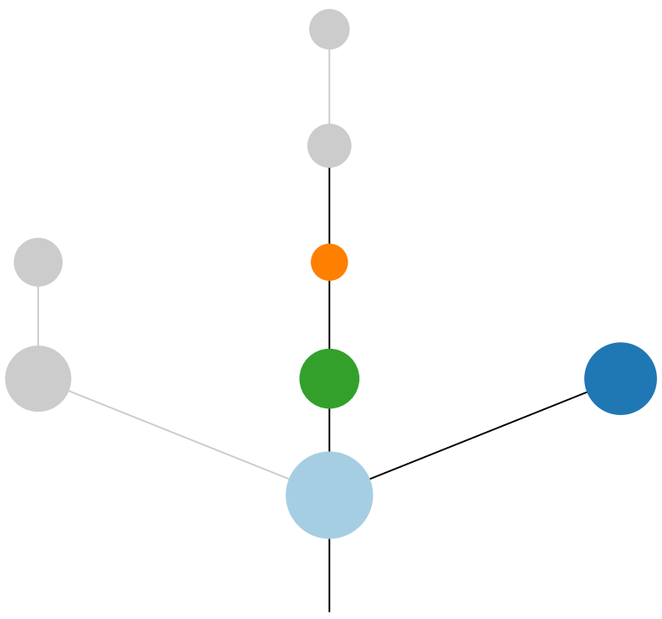
Gene	Cluster	Cytoband	Type
NBN	1	8q21.3	SNV
KLF6	1	10p15.1	SNV
KRAS	1	12p12.1	SNV
FLT4	3	5q35.3	SNV

Histology:Adeno, Age:72, PackYears:35, Size:40
Stage:1b, Gender:Female, GD:Subclonal GD, Recur:no

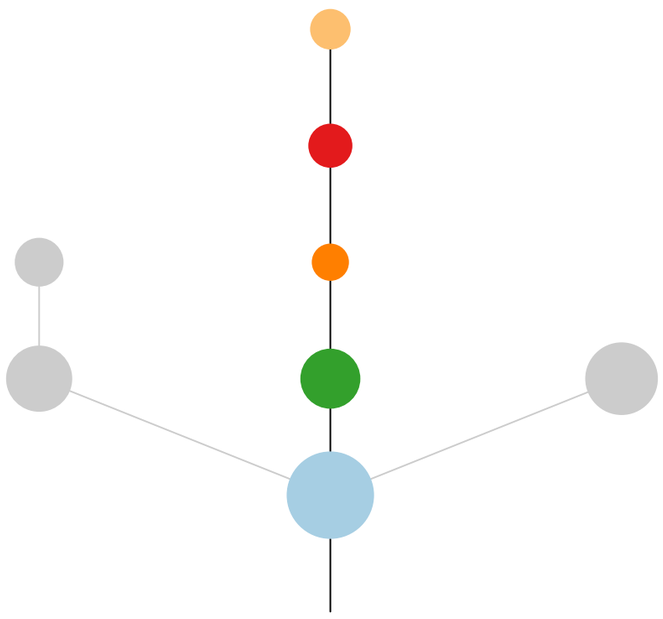
R1

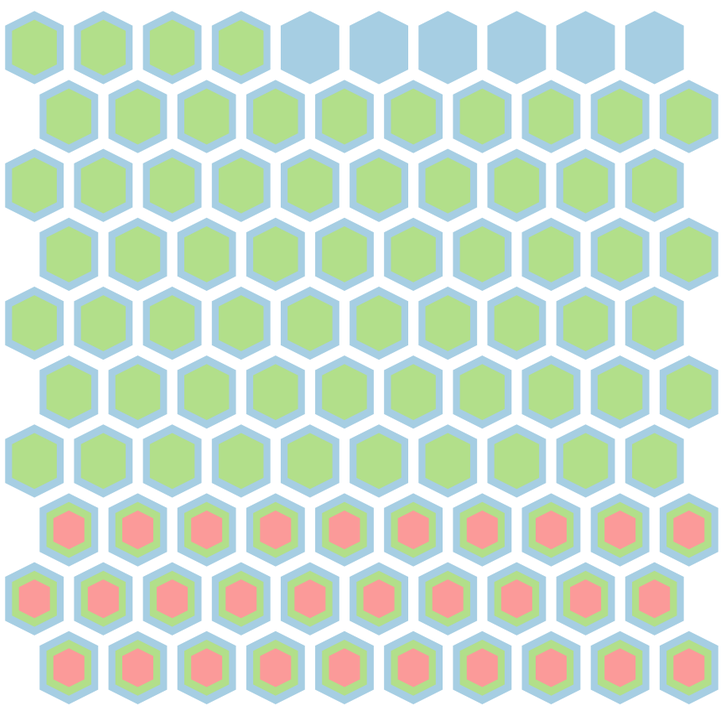


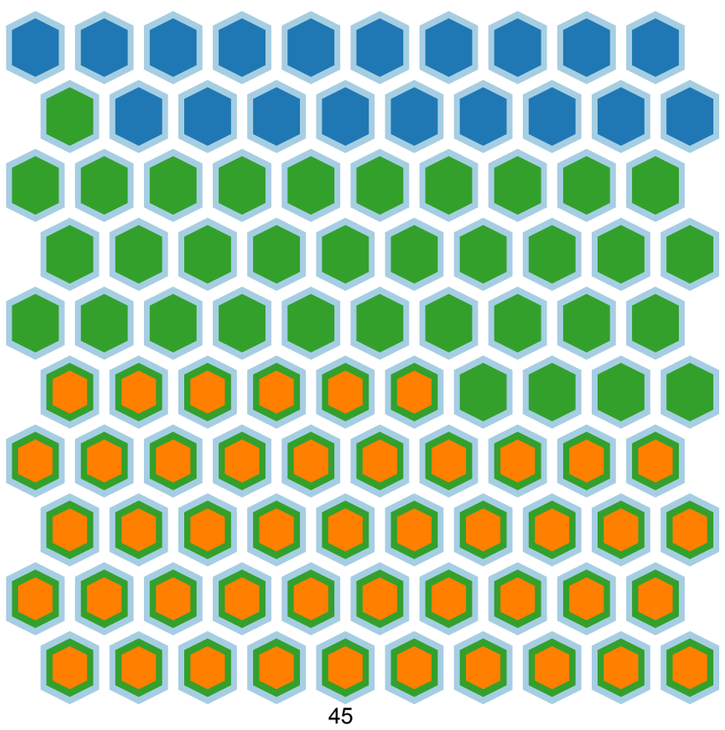
R2



R3







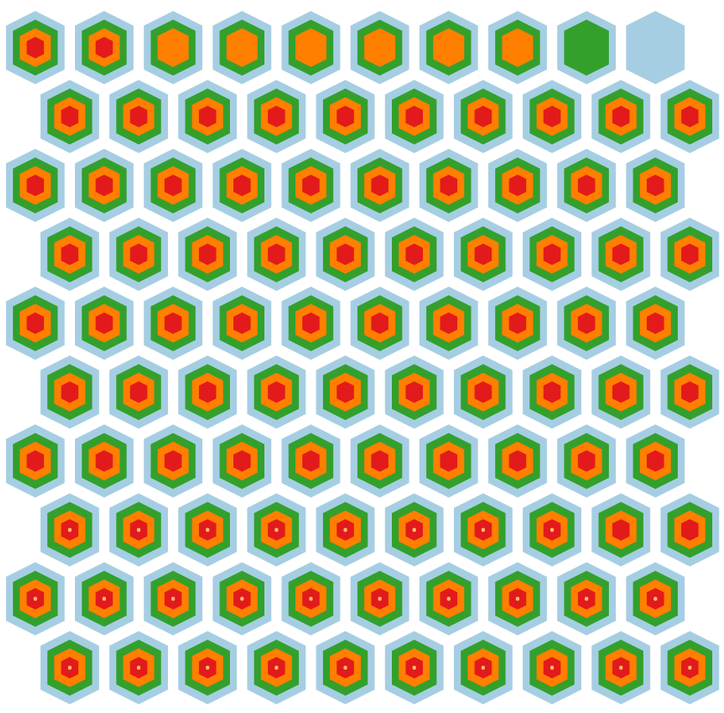
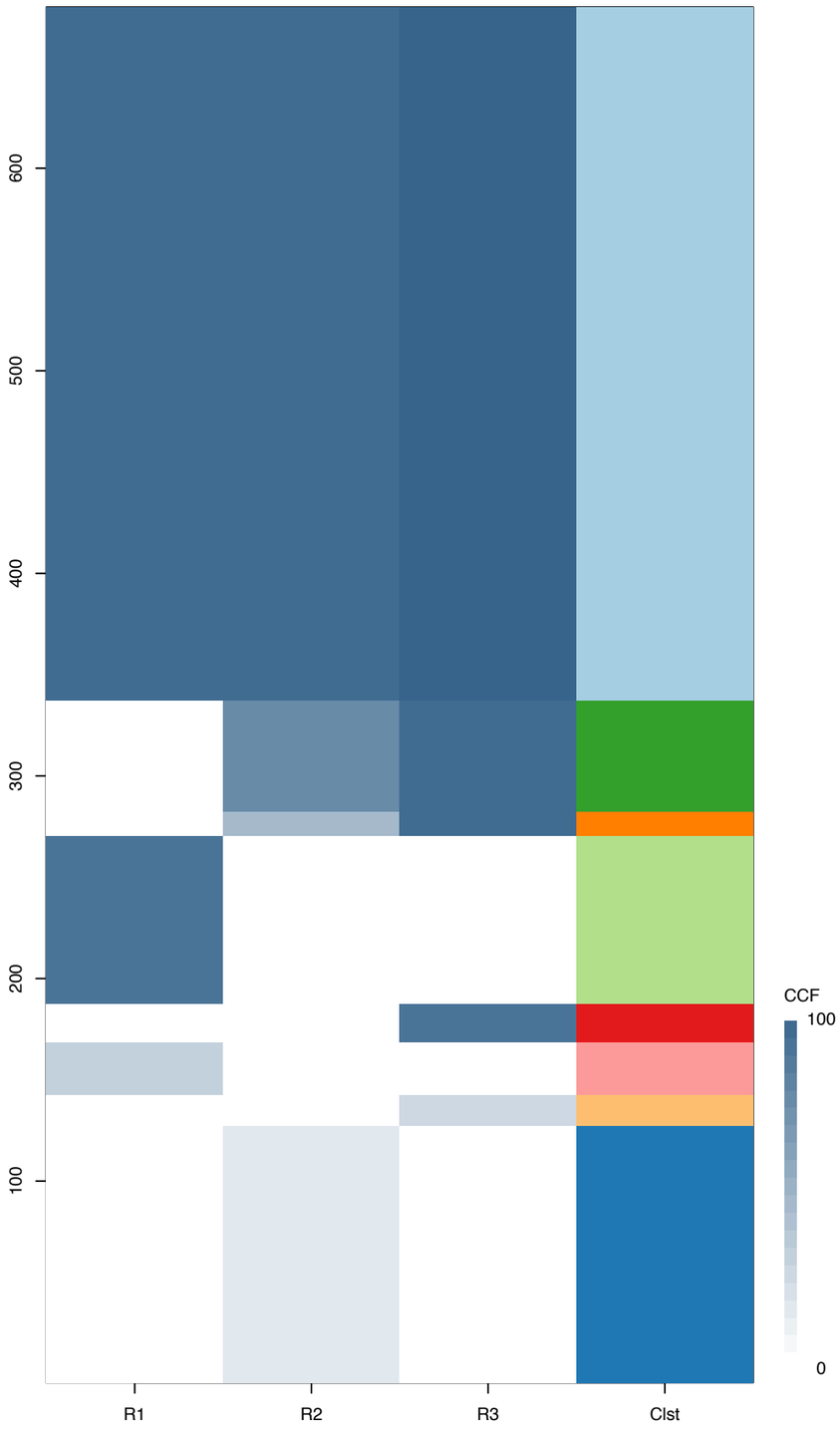
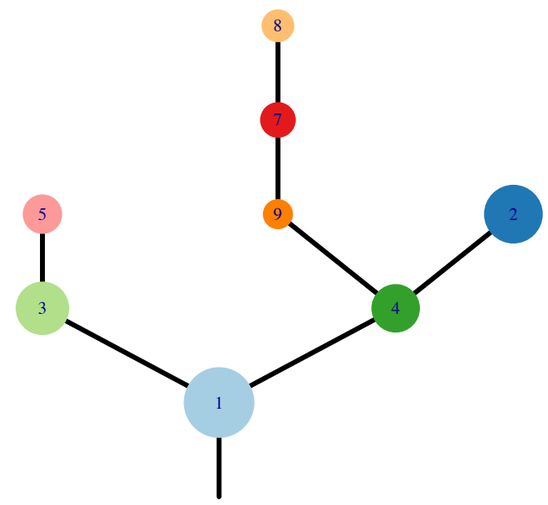


Fig.S12K



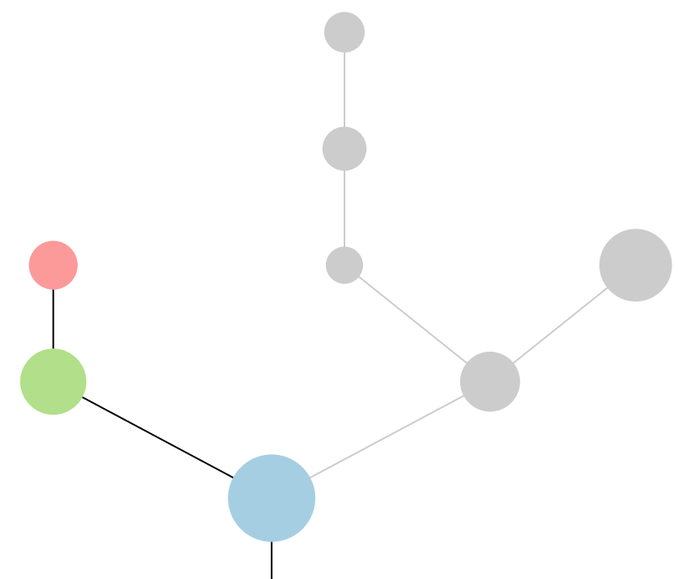
CRUK0011_C



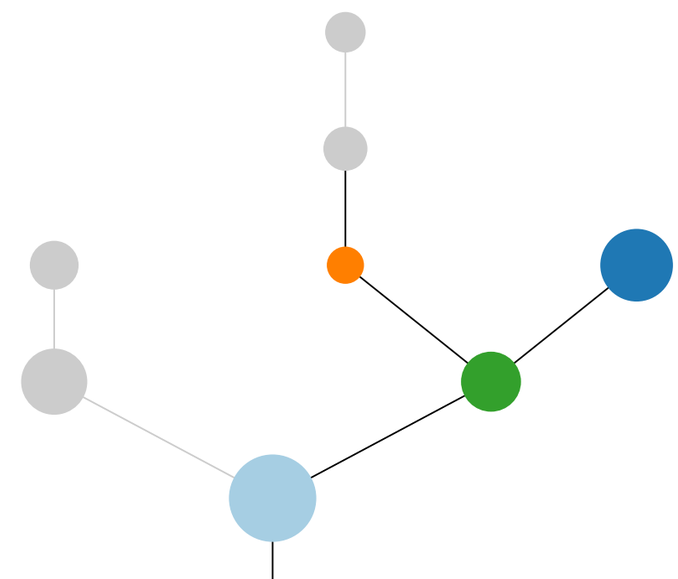
Gene	Cluster	Cytoband	Type
NBN	1	8q21.3	SNV
KLF6	1	10p15.1	SNV
KRAS	1	12p12.1	SNV
FLT4	3	5q35.3	SNV

Histology:Adeno, Age:72, PackYears:35, Size:40
Stage:1b, Gender:Female, GD:Subclonal GD, Recur:no

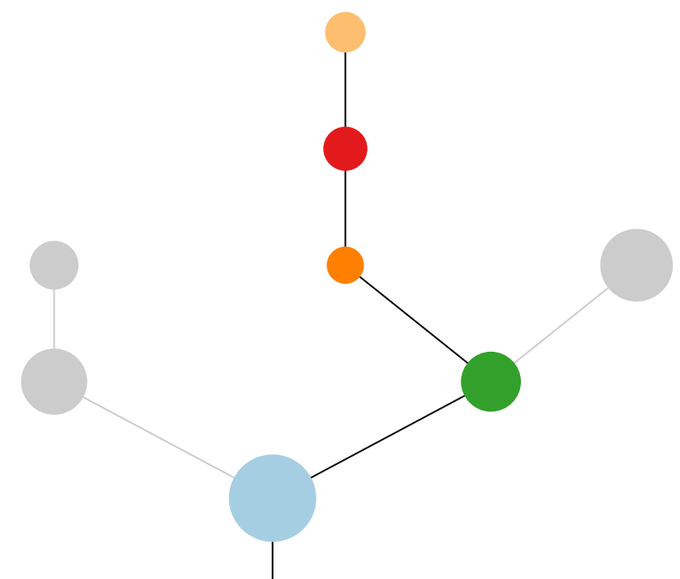
R1

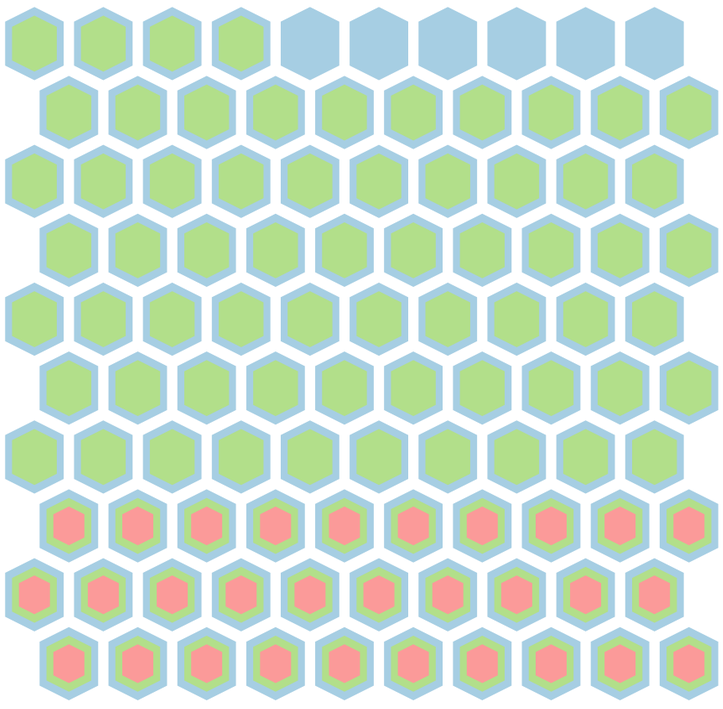


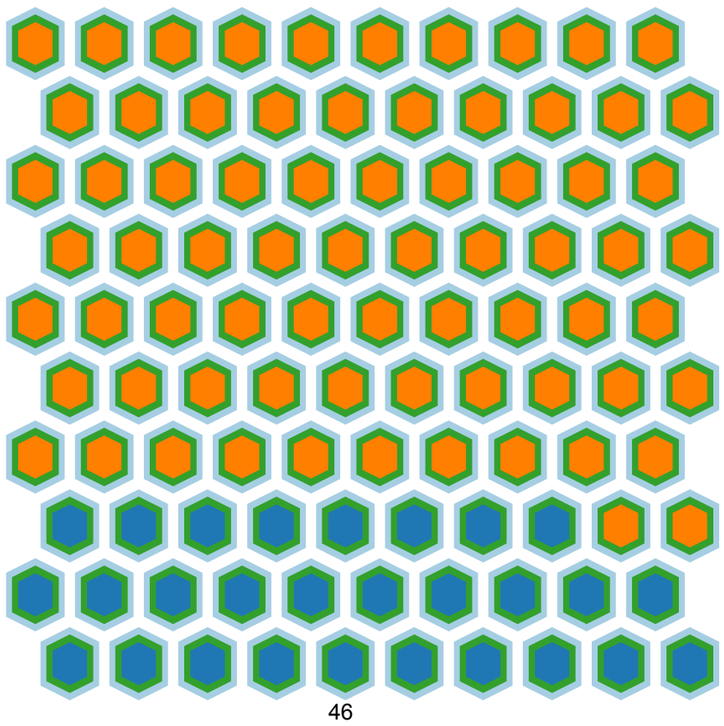
R2



R3







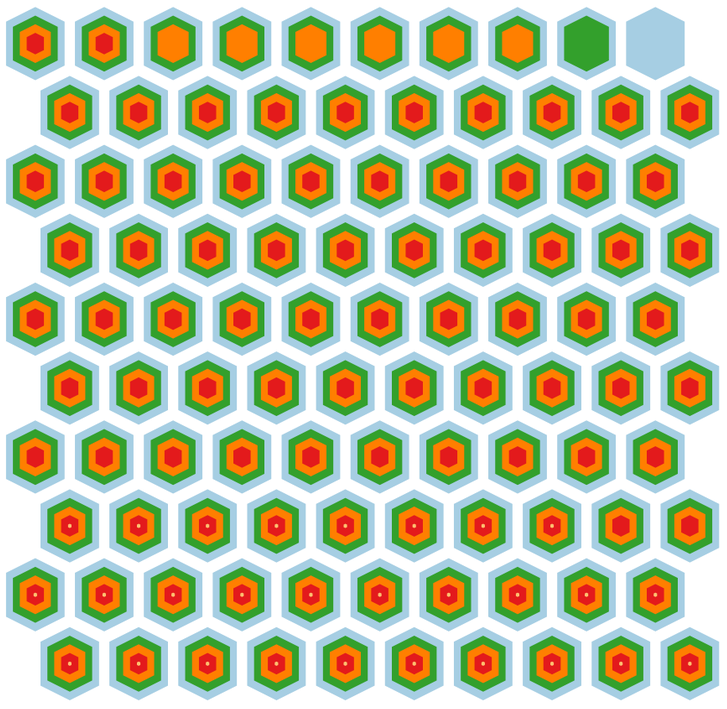
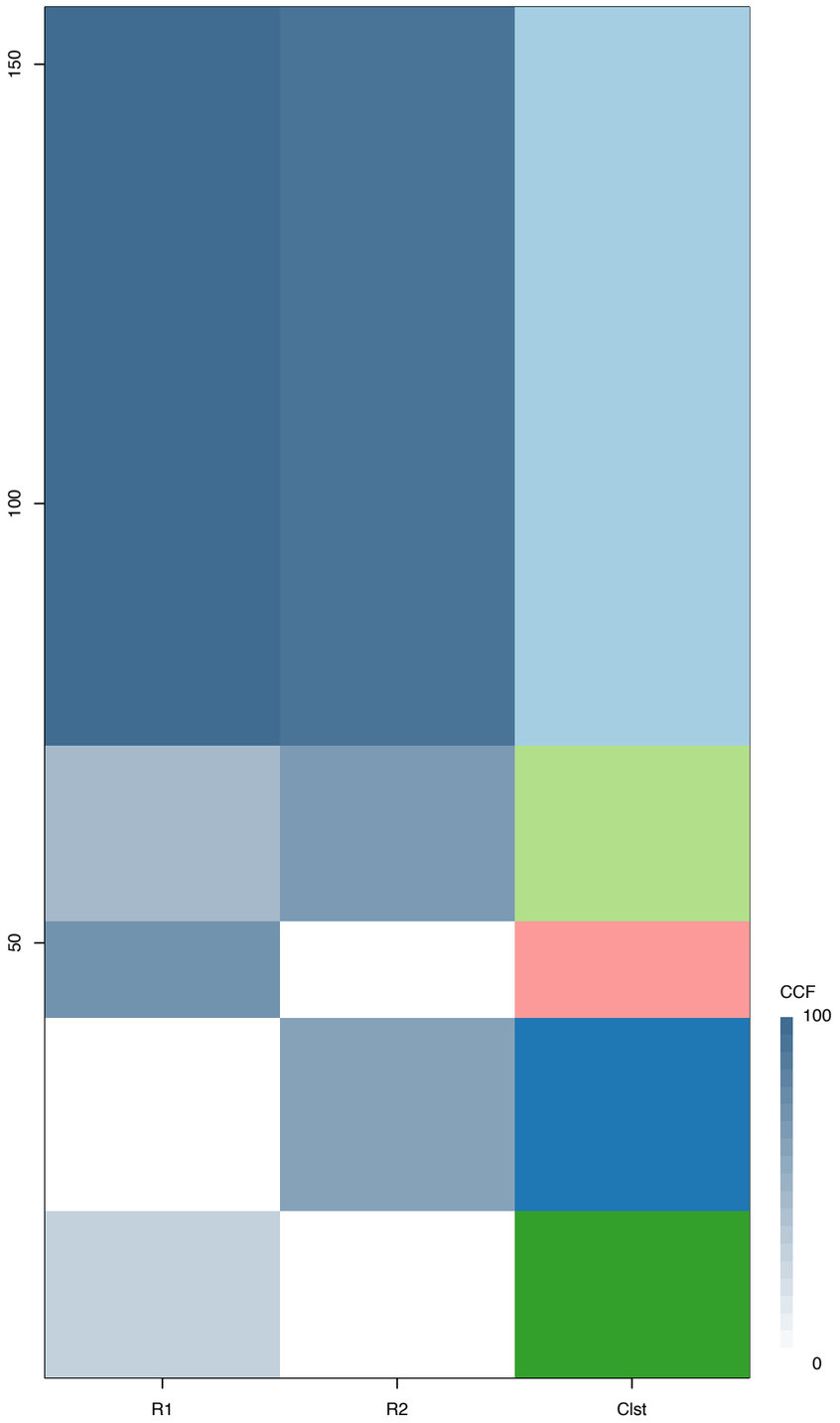
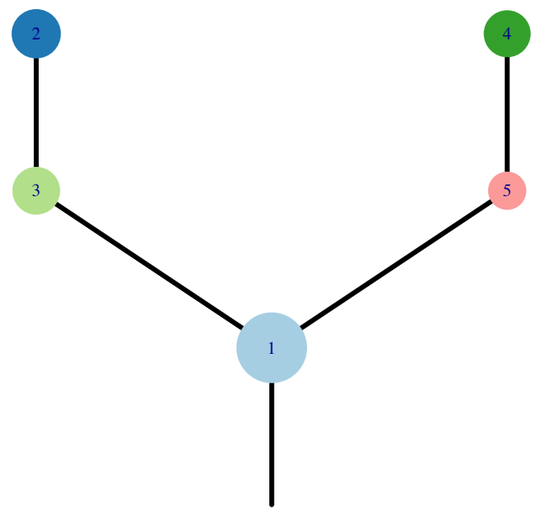


Fig.S12L



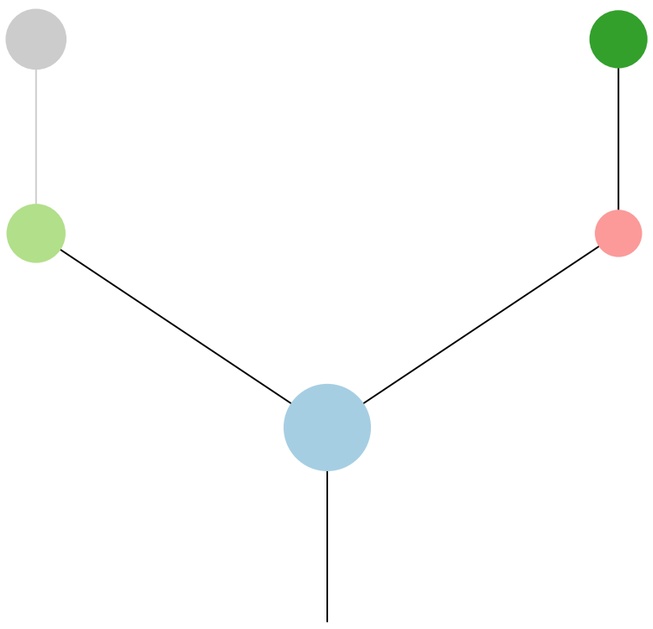
CRUK0012_A



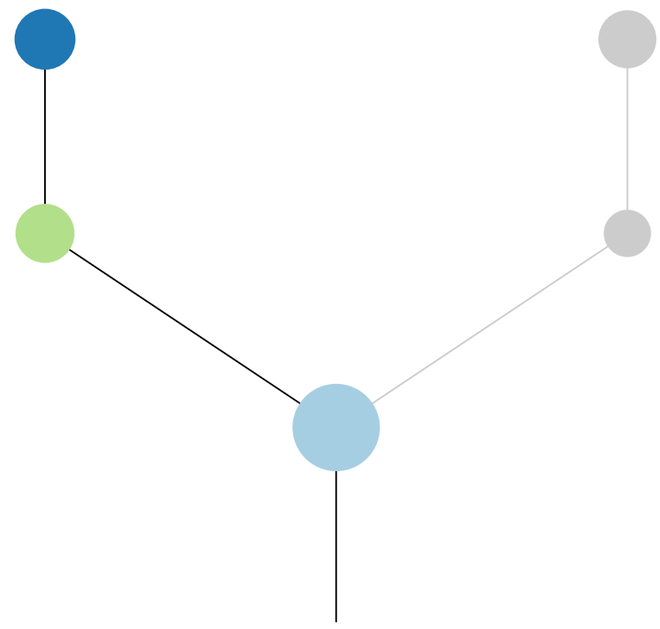
Histology:Adeno, Age:66, PackYears:14.25, Size:28
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

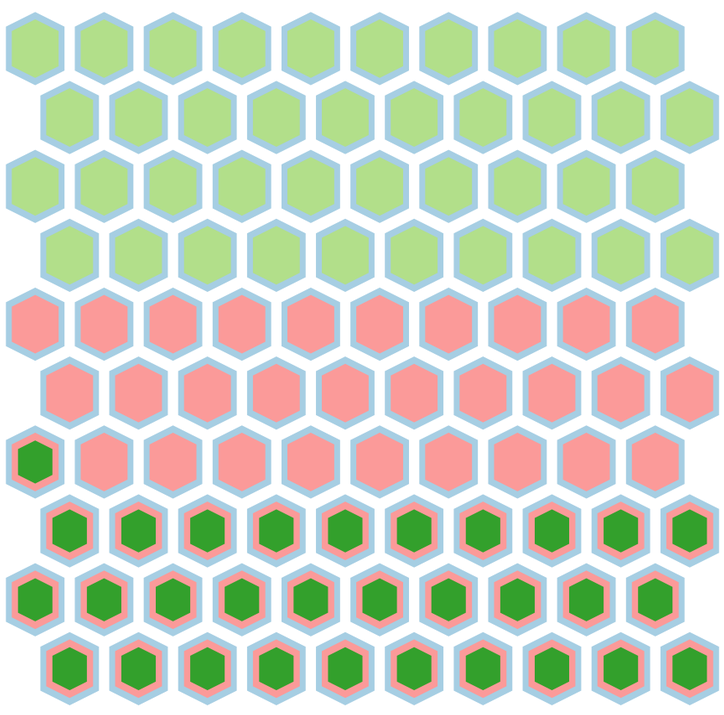
Gene	Cluster	Cytoband	Type
ARNT	1	1q21.3	Amp
SETDB1	1	1q21.3	Amp
TPM3	1	1q21.3	Amp
EGFR	1	7p11.2	SNV
AURKA	1	20q13.2	Amp
GNAS	1	20q13.32	Amp
SS18L1	1	20q13.33	Amp
MUC1	3	1q22	Amp
HNF1A	4	12q24.31	SNV
MUC1	5	1q22	Amp

R1



R2





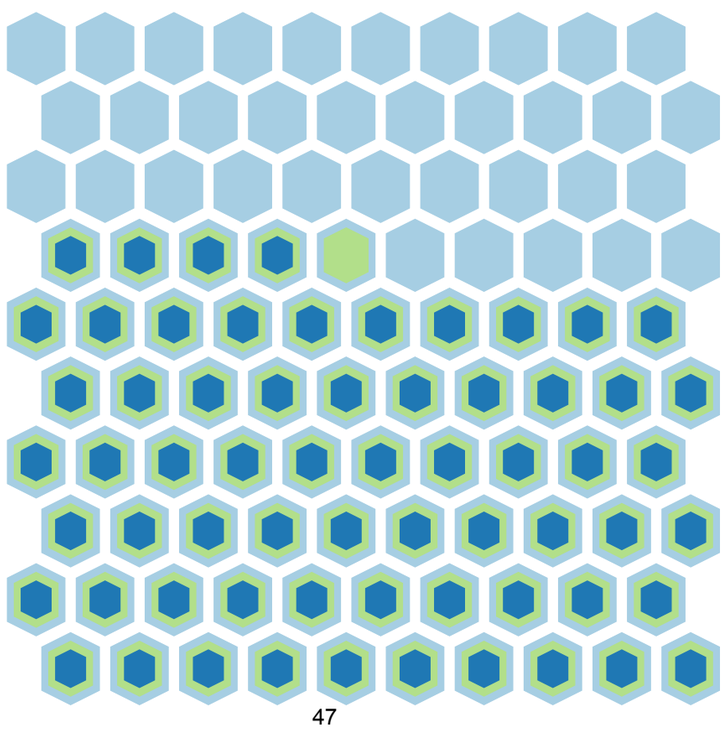
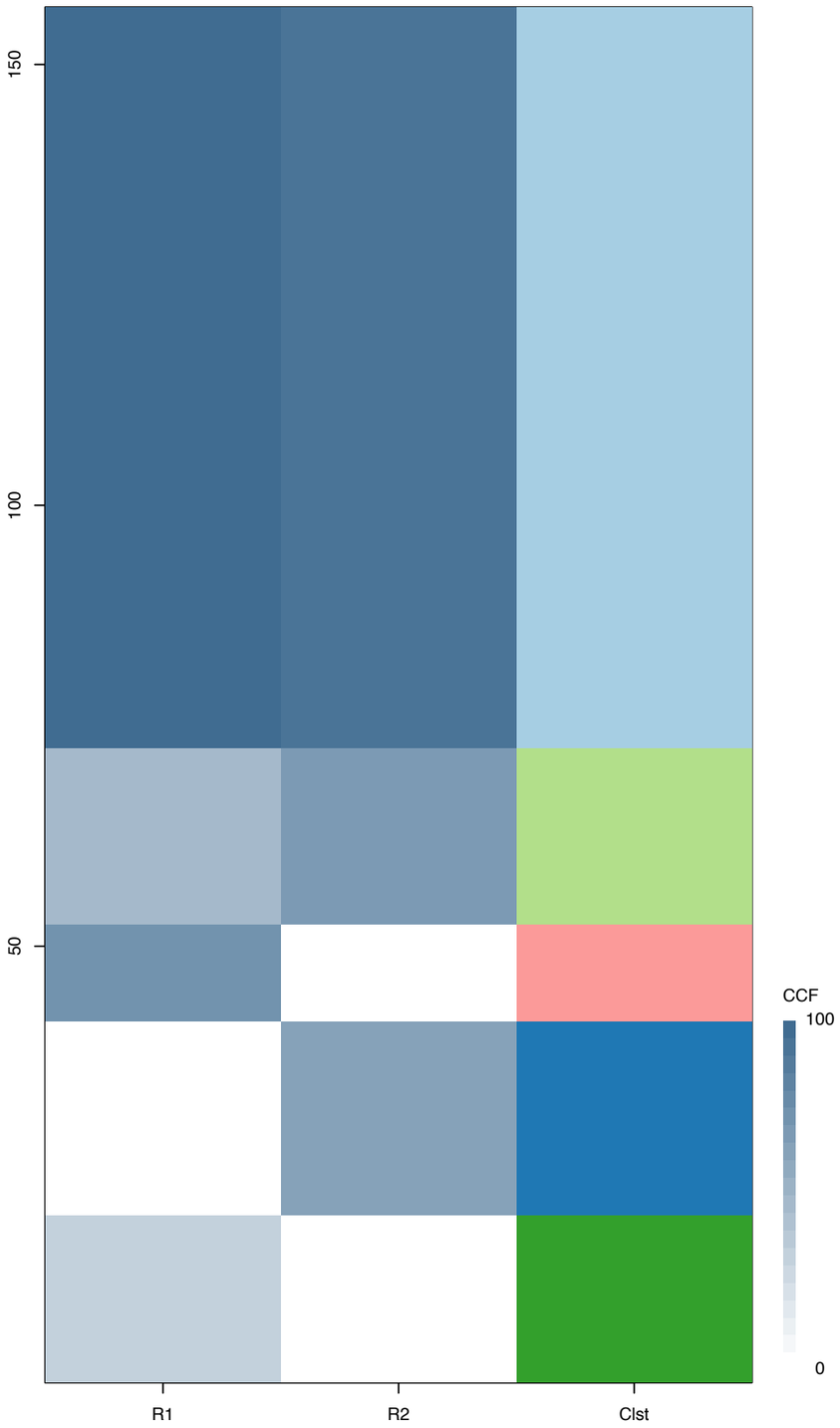
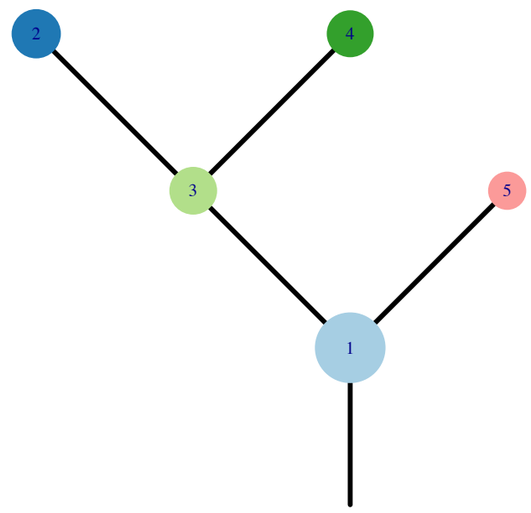


Fig.S12L



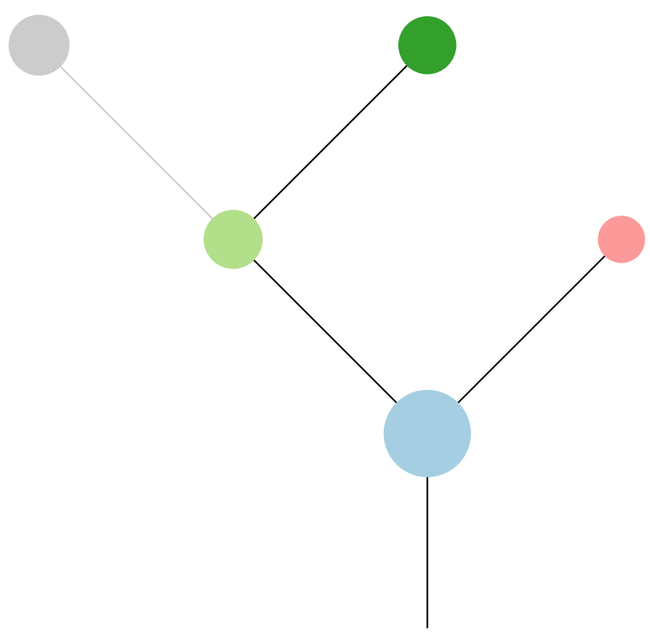
CRUK0012_B



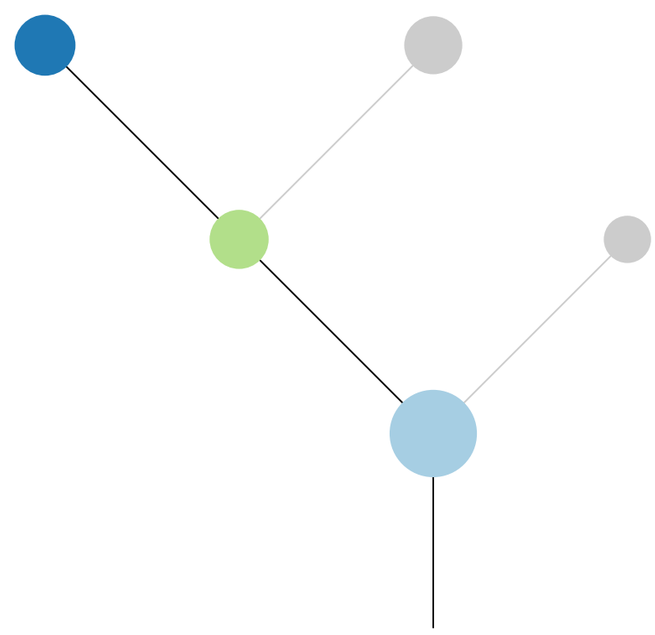
Histology:Adeno, Age:66, PackYears:14.25, Size:28
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

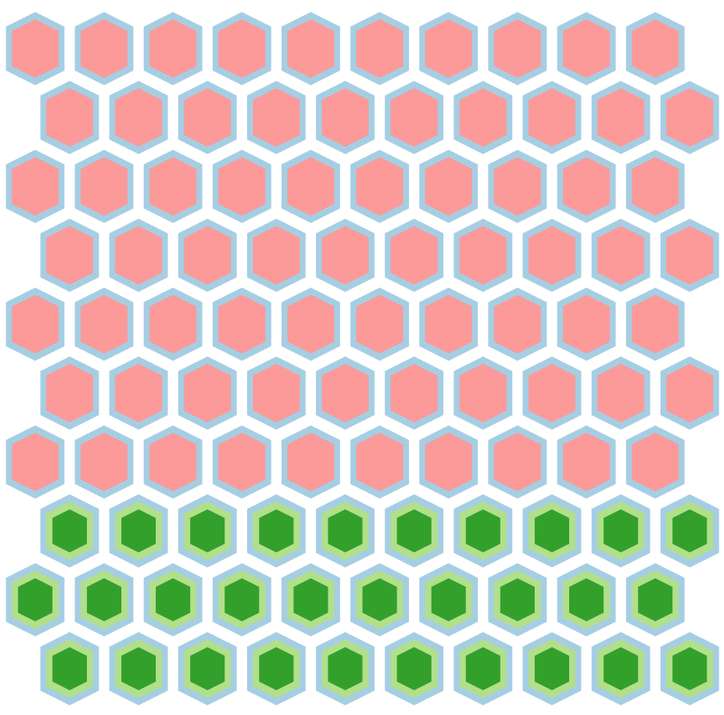
Gene	Cluster	Cytoband	Type
ARNT	1	1q21.3	Amp
SETDB1	1	1q21.3	Amp
TPM3	1	1q21.3	Amp
EGFR	1	7p11.2	SNV
AURKA	1	20q13.2	Amp
GNAS	1	20q13.32	Amp
SS18L1	1	20q13.33	Amp
MUC1	3	1q22	Amp
HNF1A	4	12q24.31	SNV
MUC1	5	1q22	Amp

R1



R2





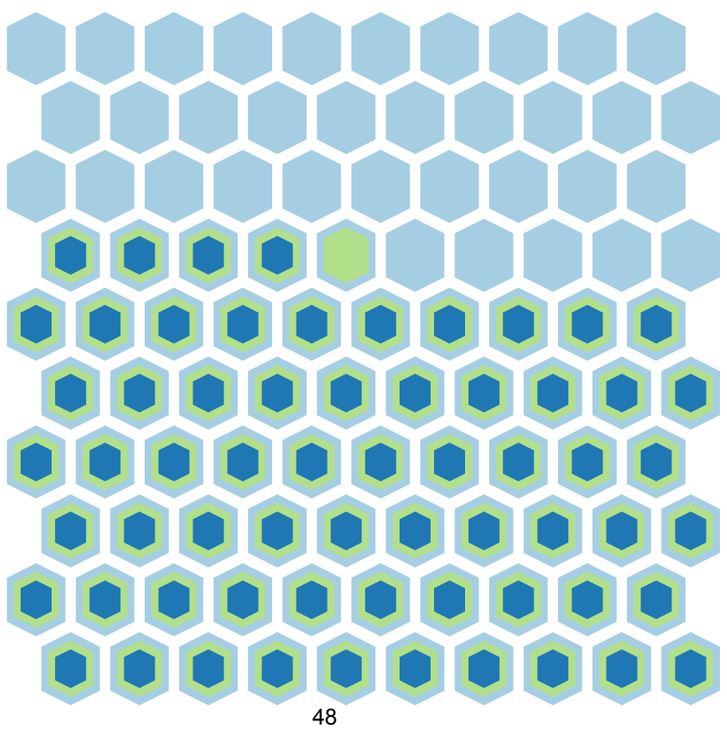
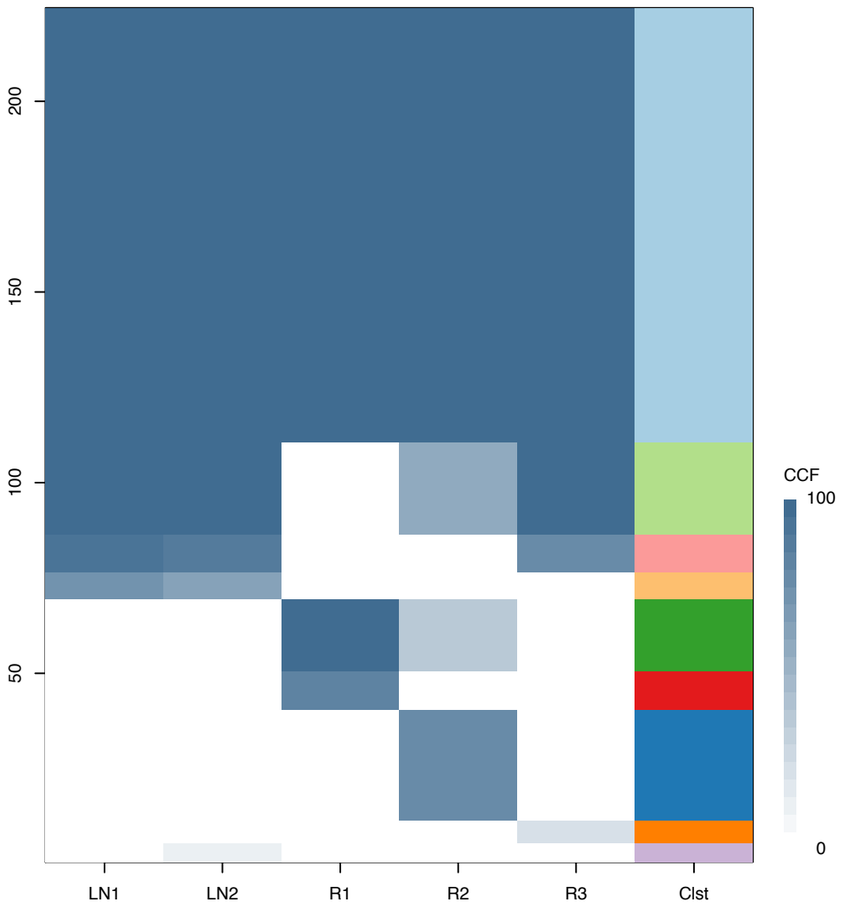
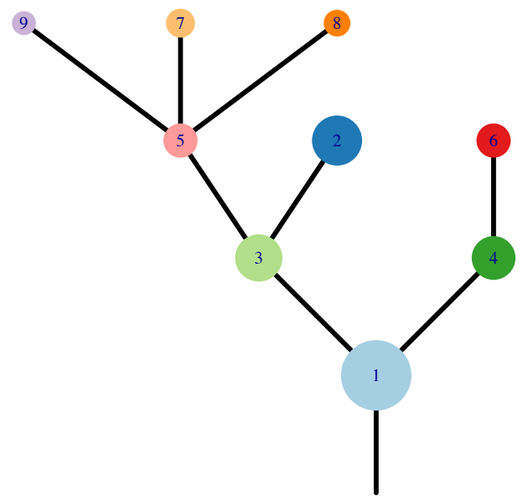


Fig.S12M



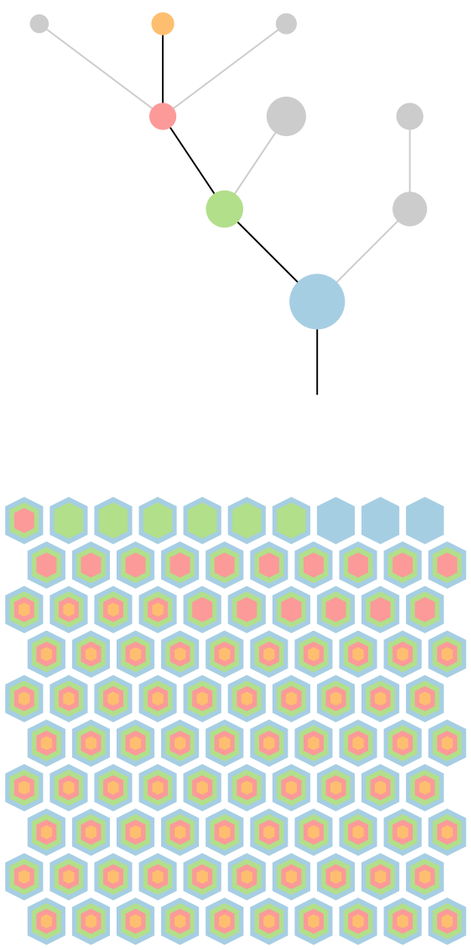
CRUK0013_A



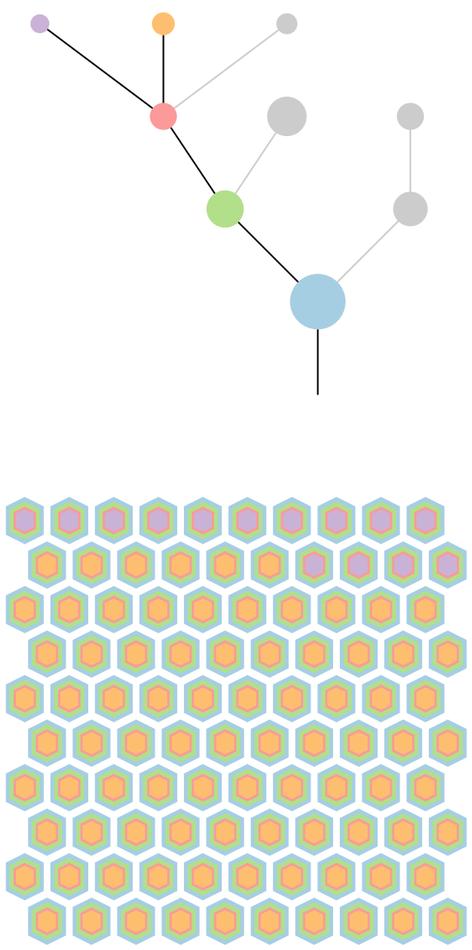
Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

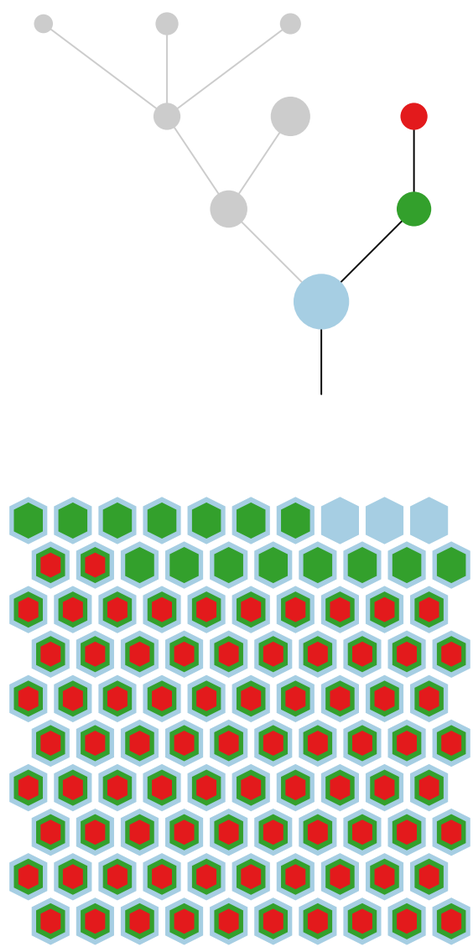
LN1



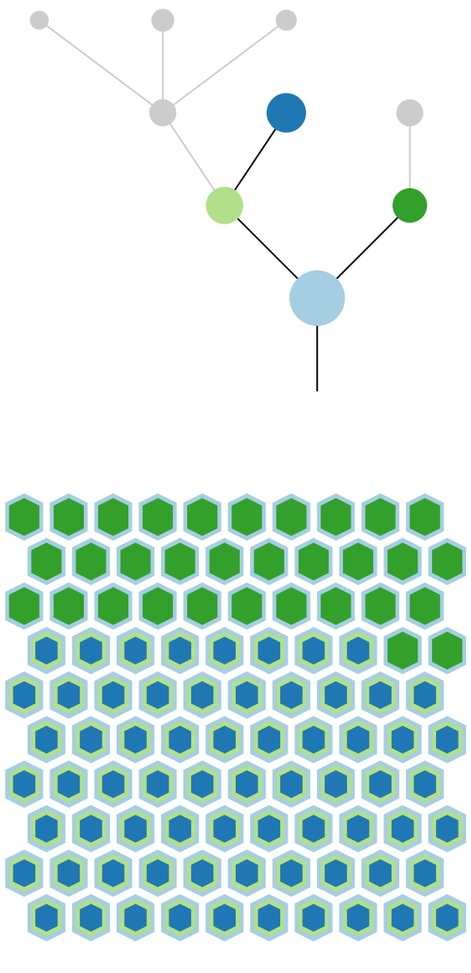
LN2



R1



R2



R3

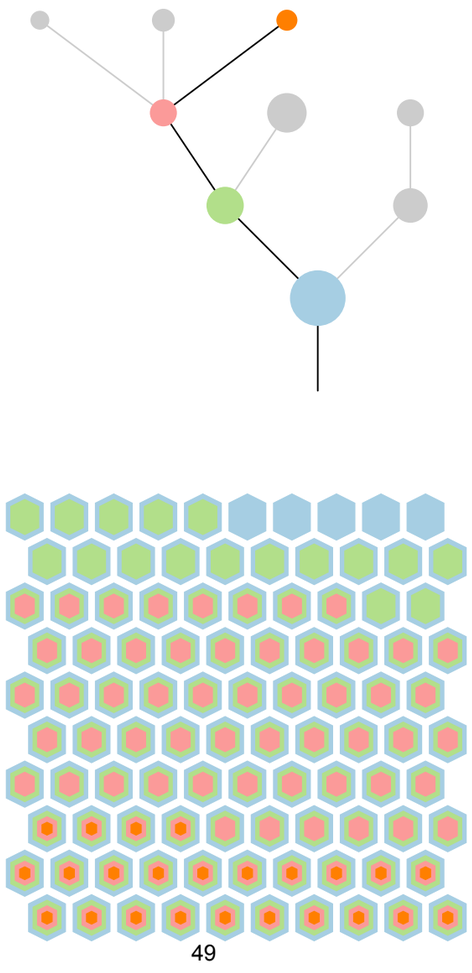
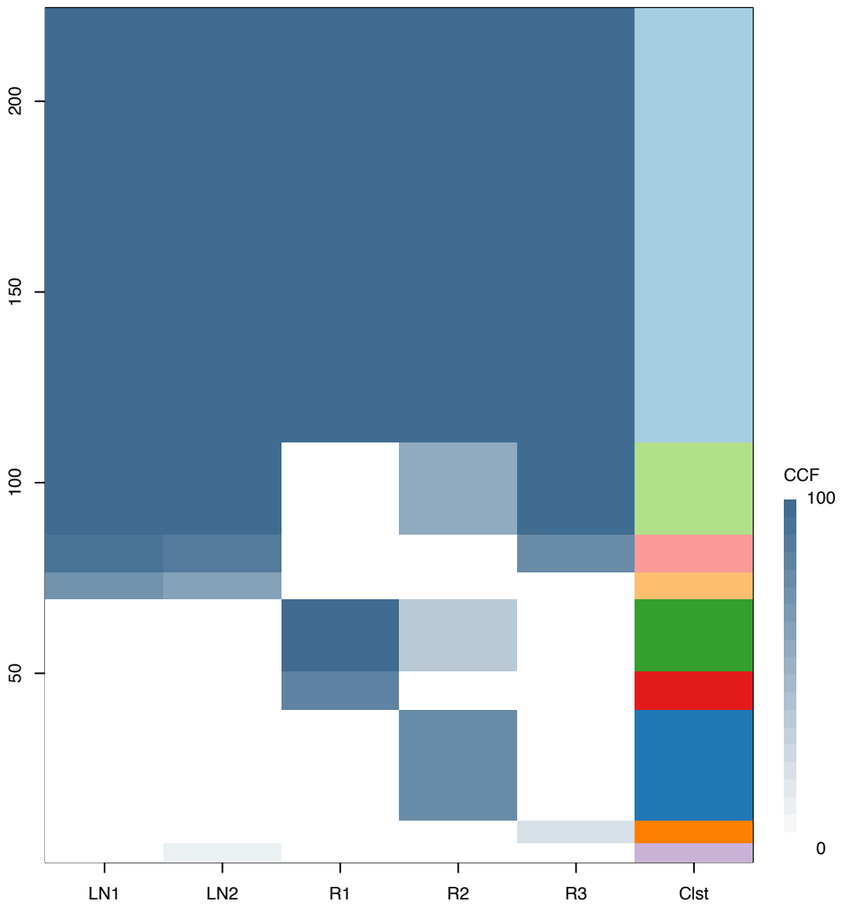
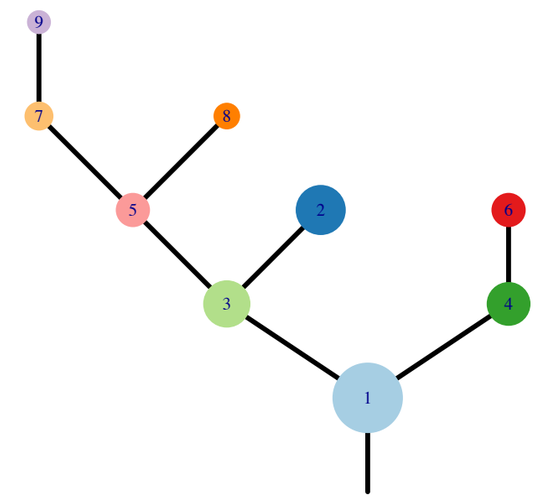


Fig.S12M



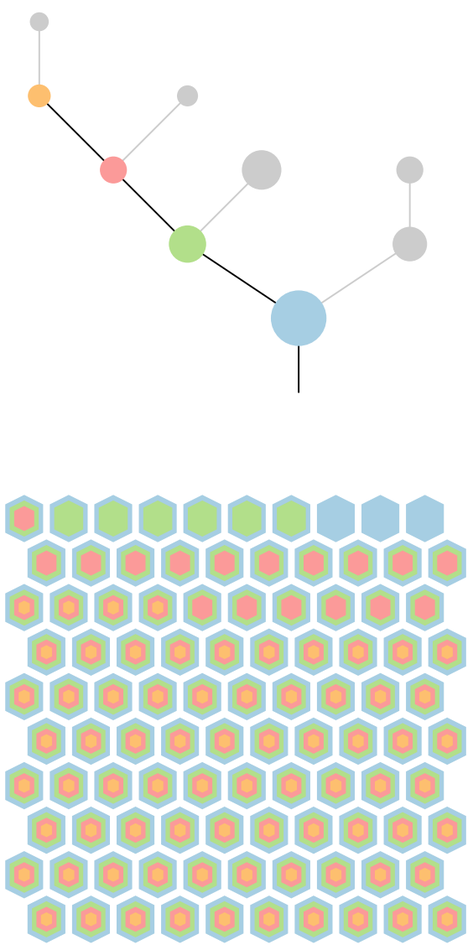
CRUK0013_B



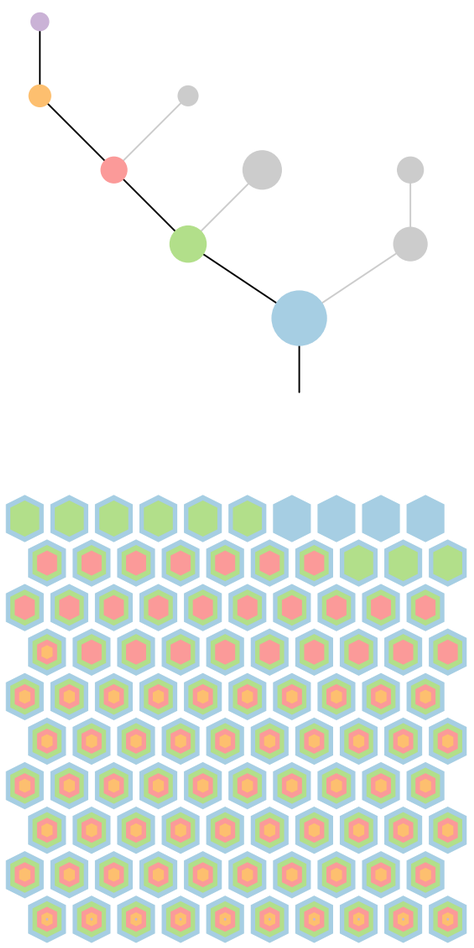
Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

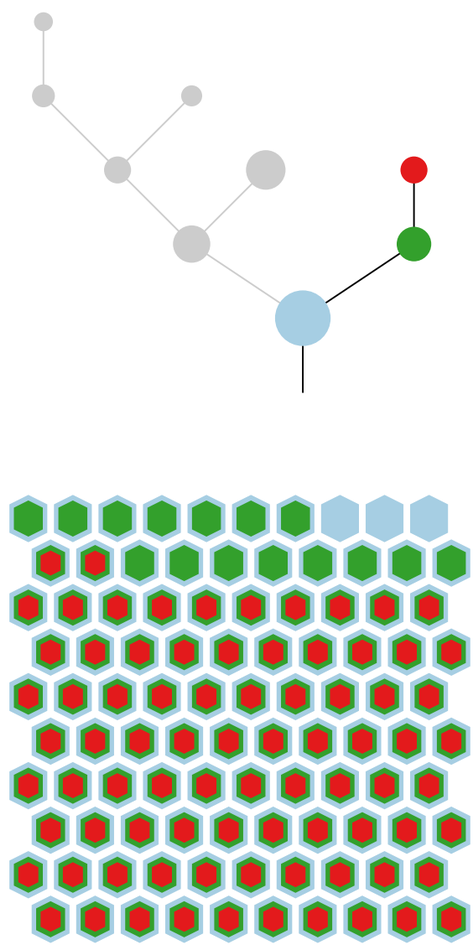
LN1



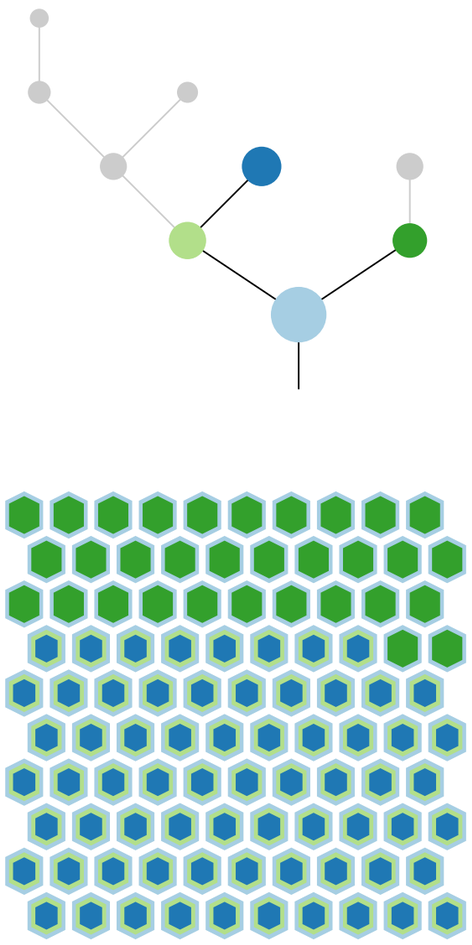
LN2



R1



R2



R3

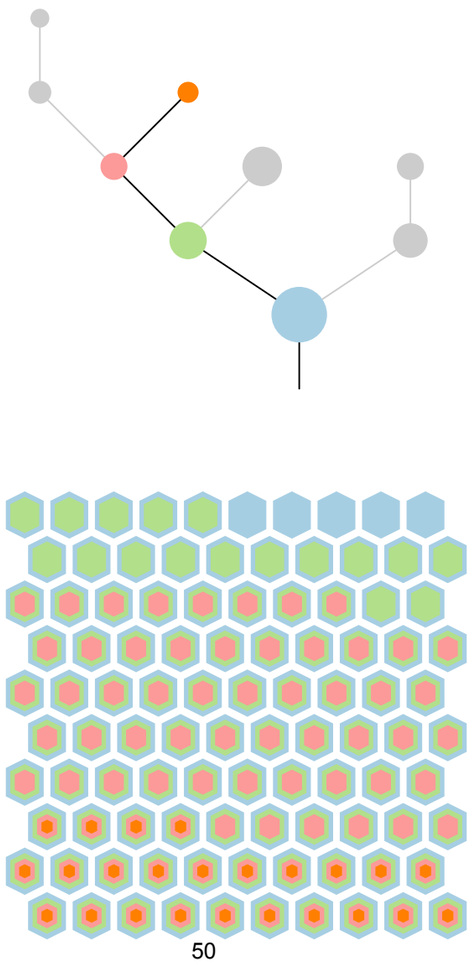
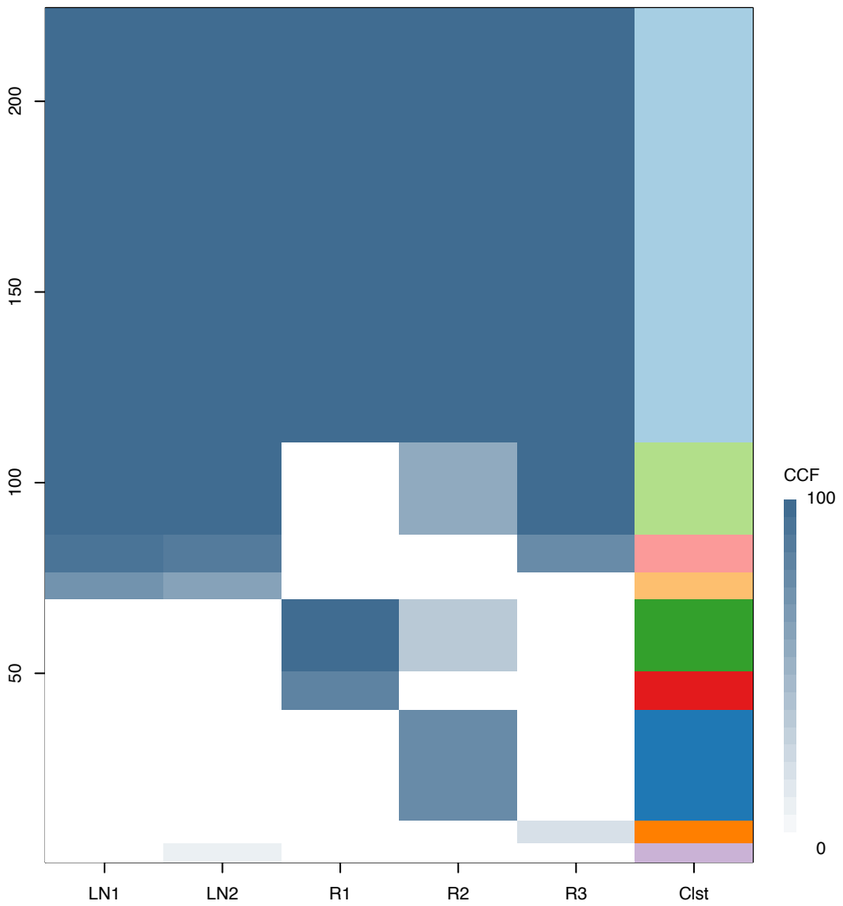
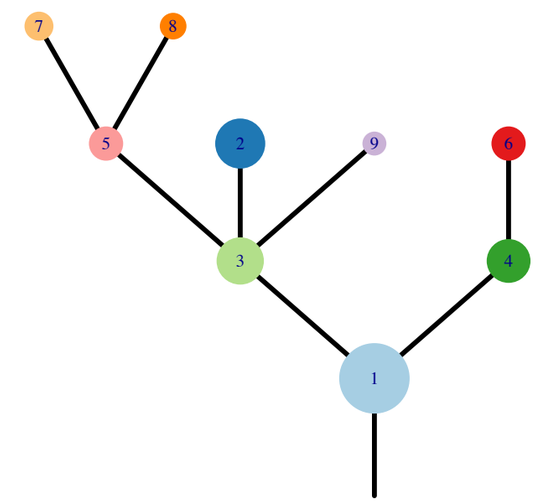


Fig.S12M



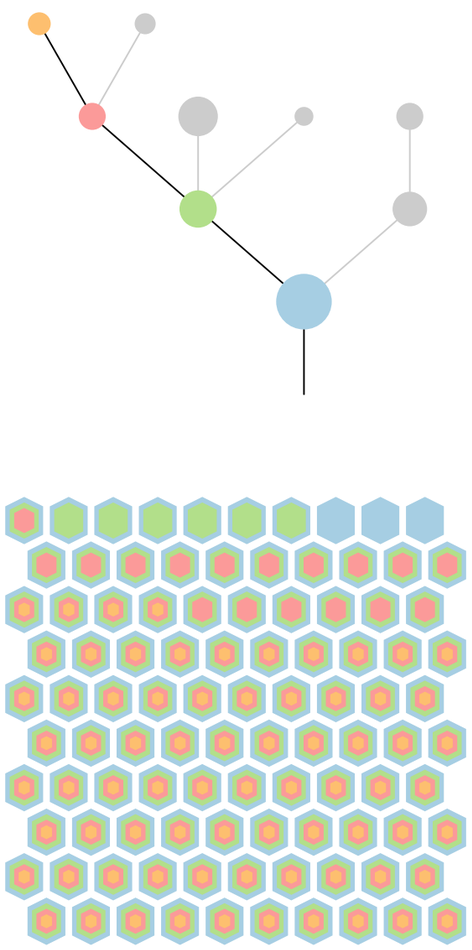
CRUK0013_C



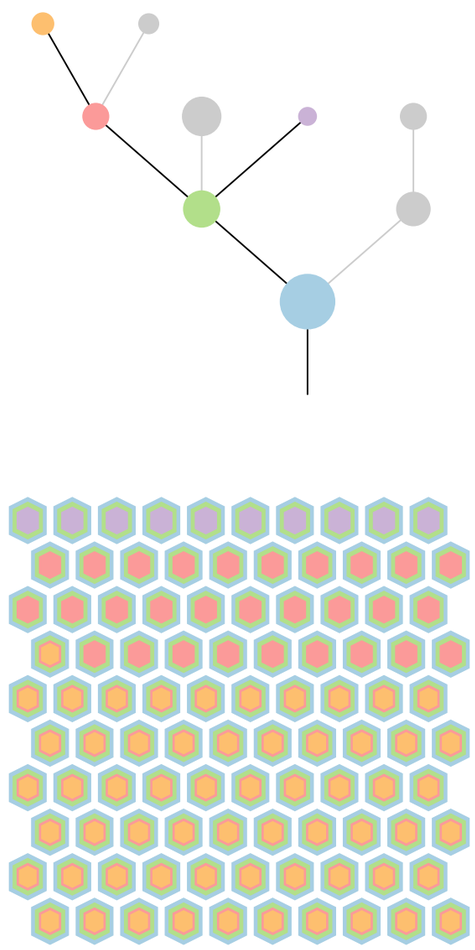
Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

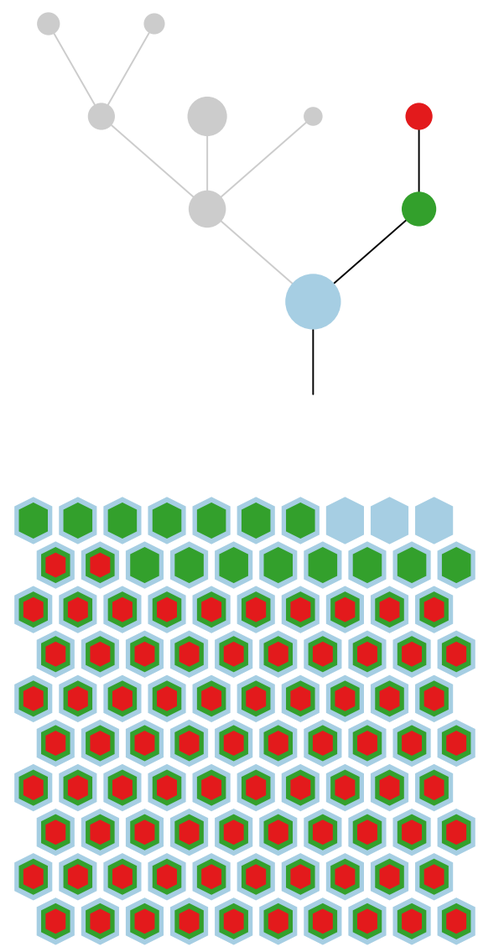
LN1



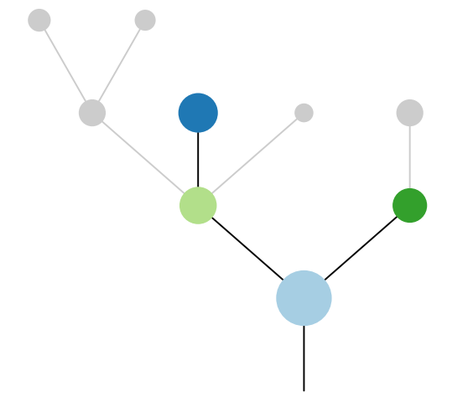
LN2



R1



R2



R3

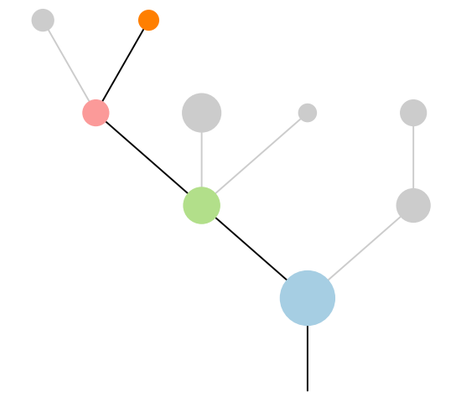
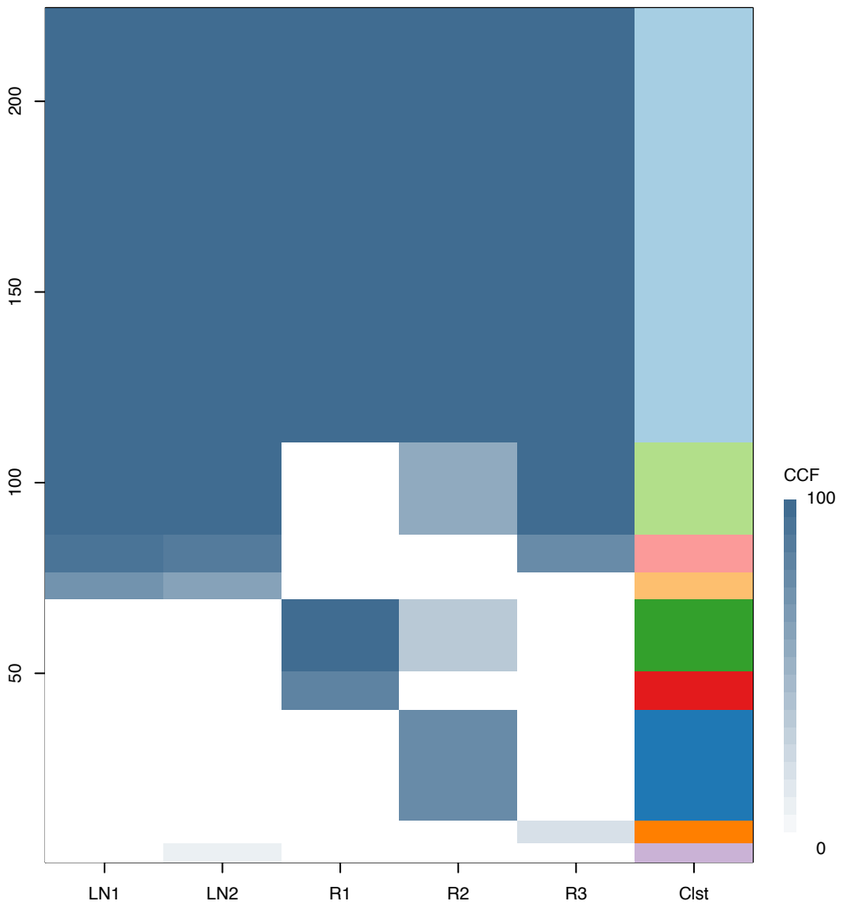
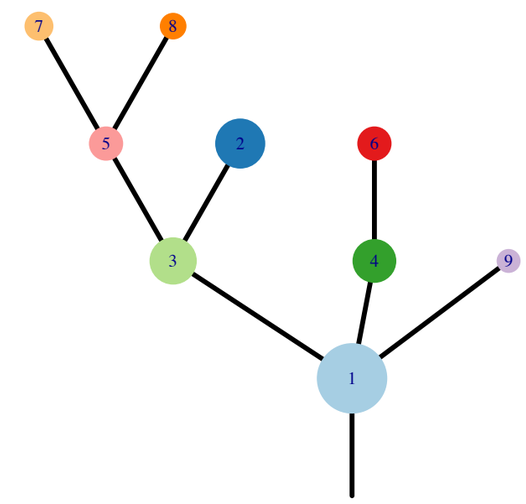


Fig.S12M



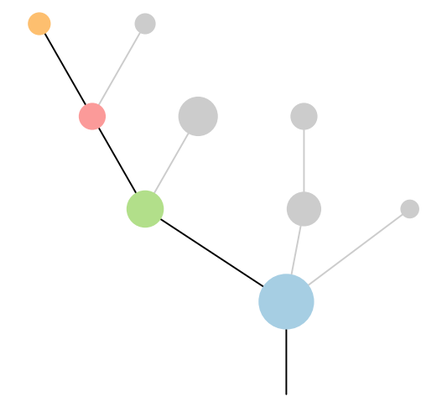
CRUK0013_D

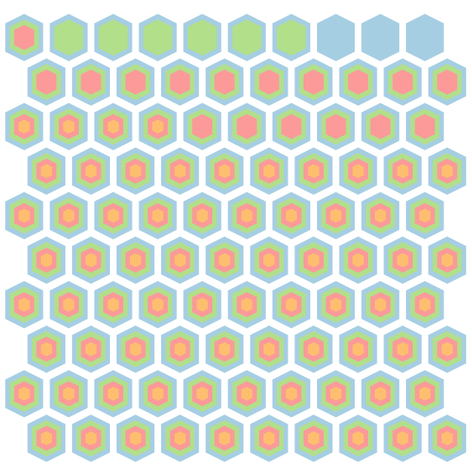


Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

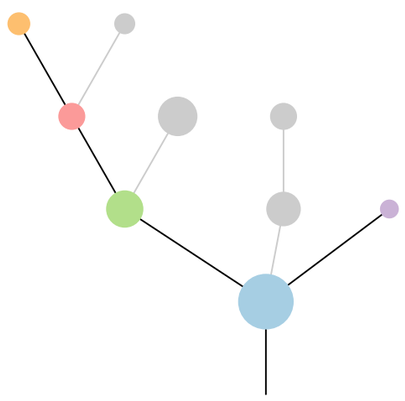
Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

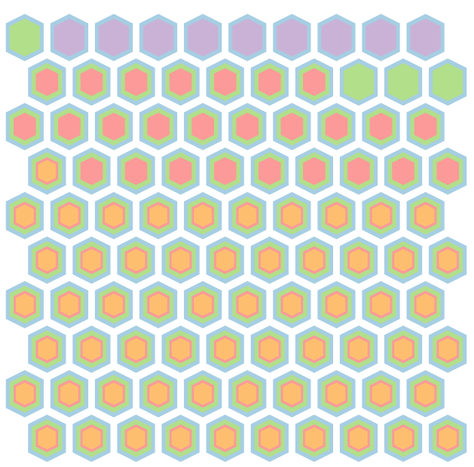
LN1



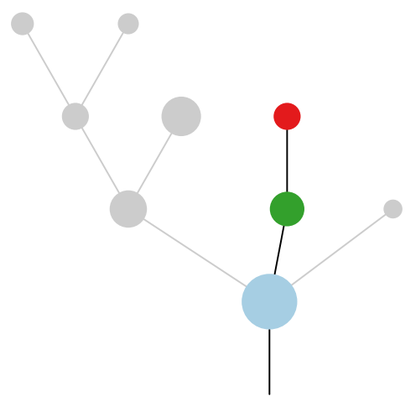


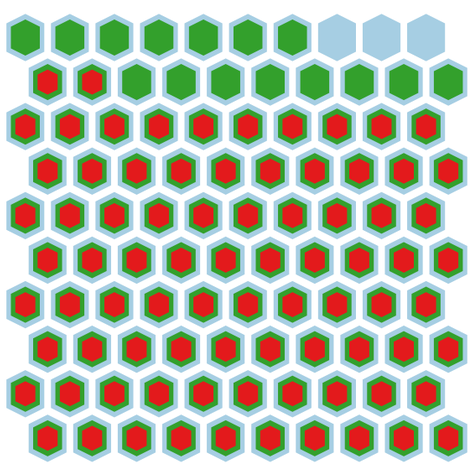
LN2



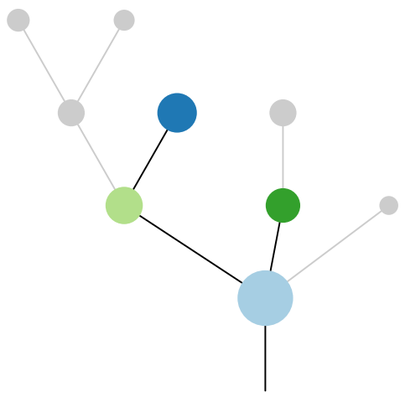


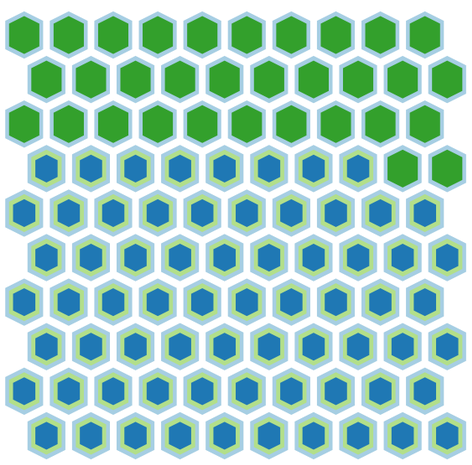
R1



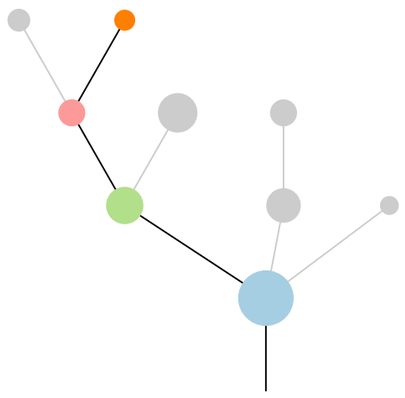


R2





R3



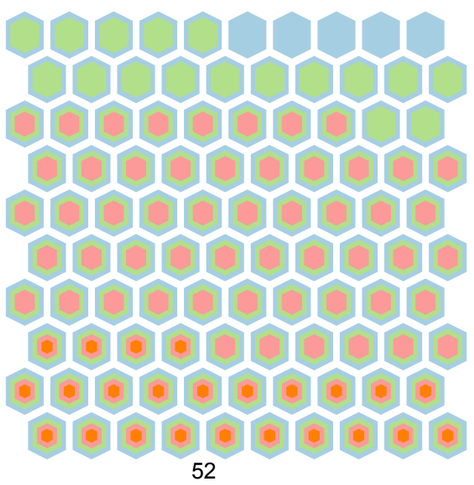
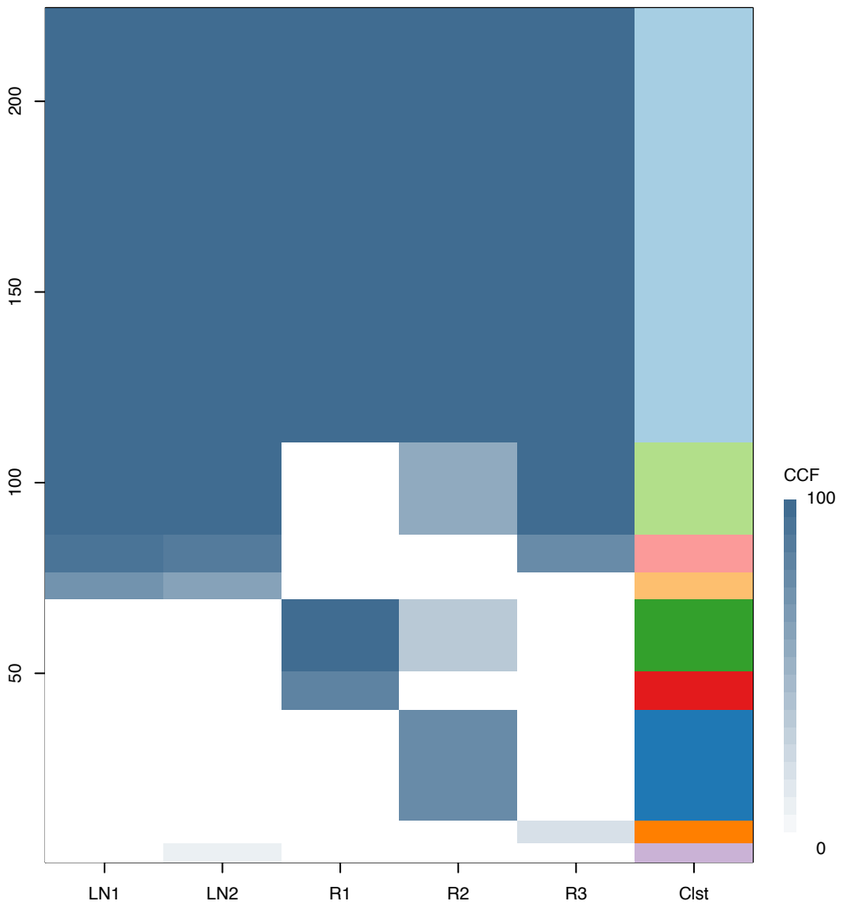
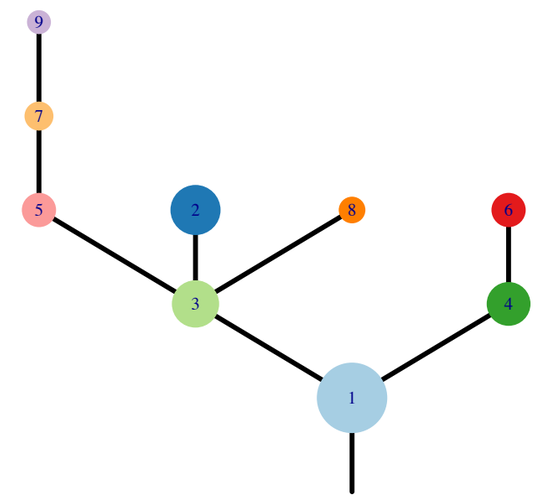


Fig.S12M



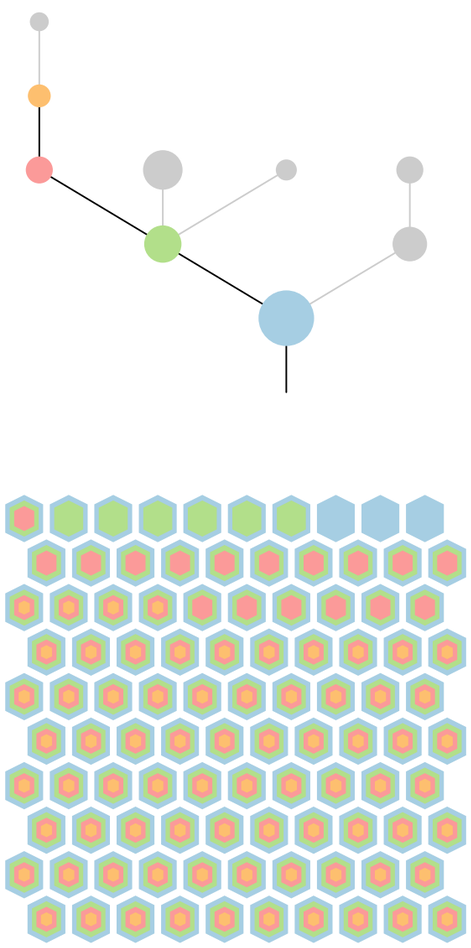
CRUK0013_E



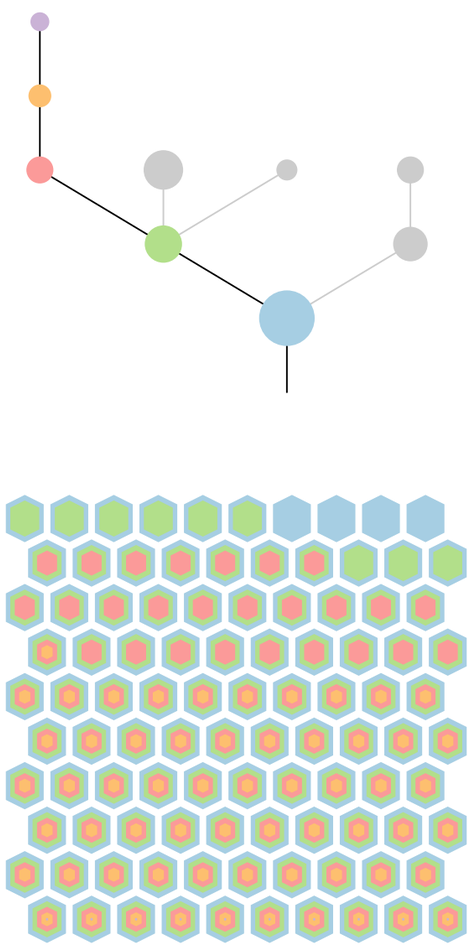
Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

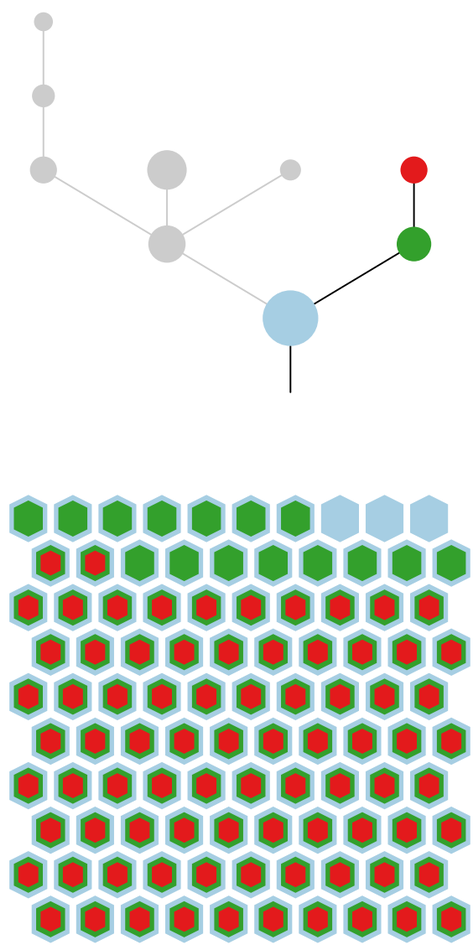
LN1



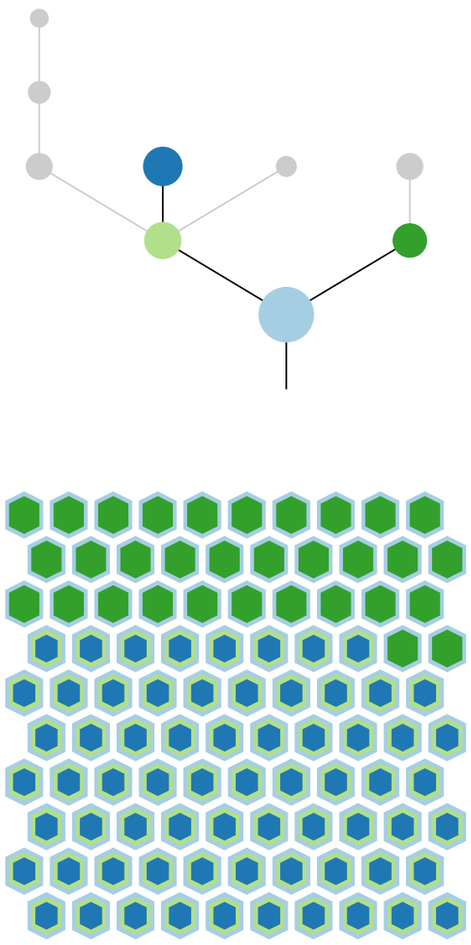
LN2



R1



R2



R3

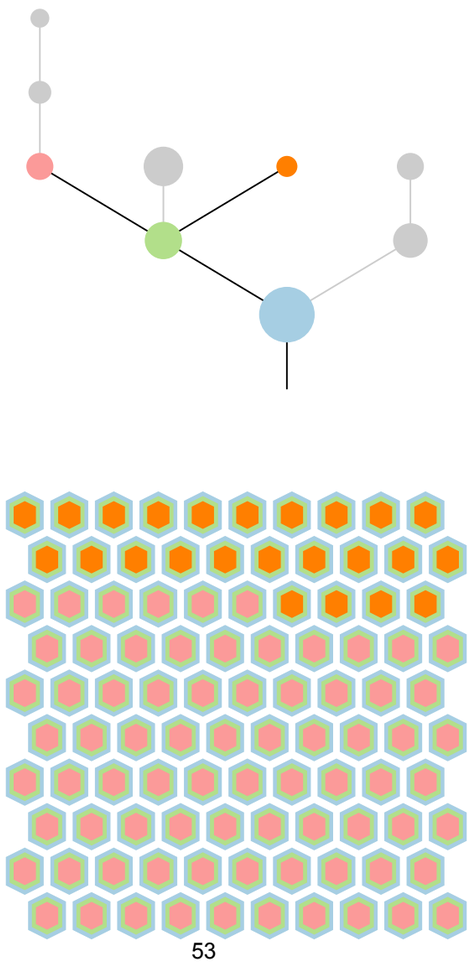
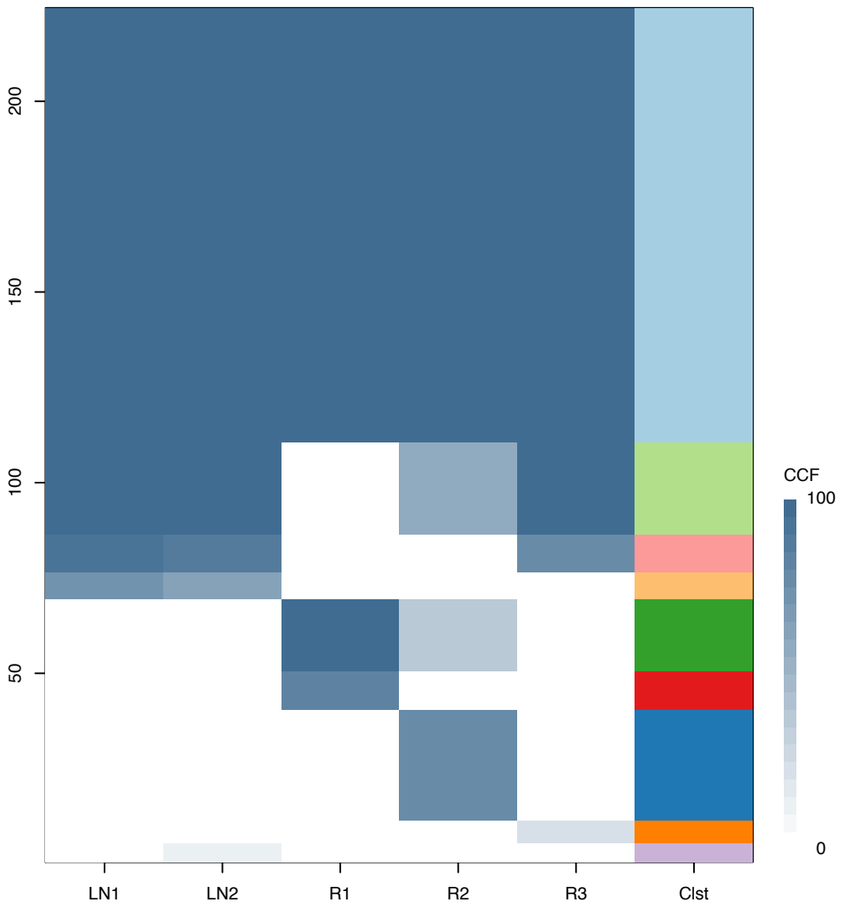
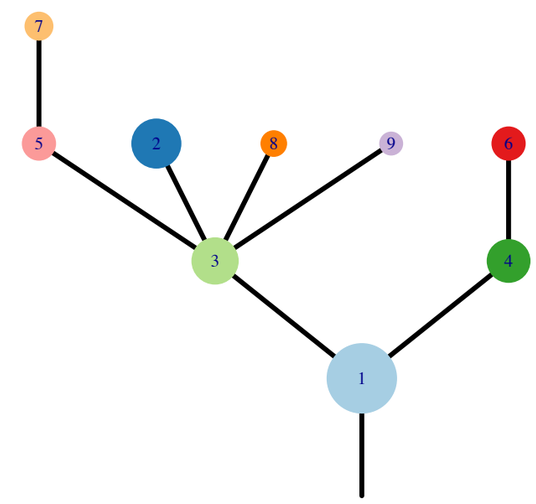


Fig.S12M



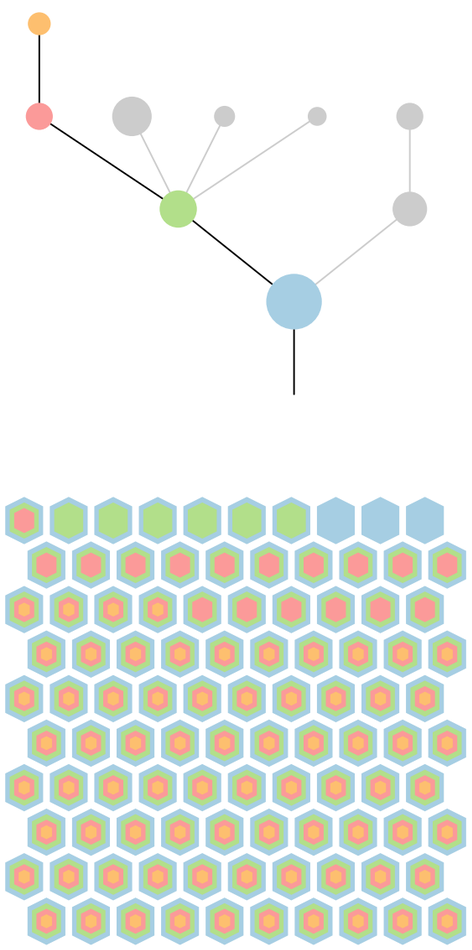
CRUK0013_F



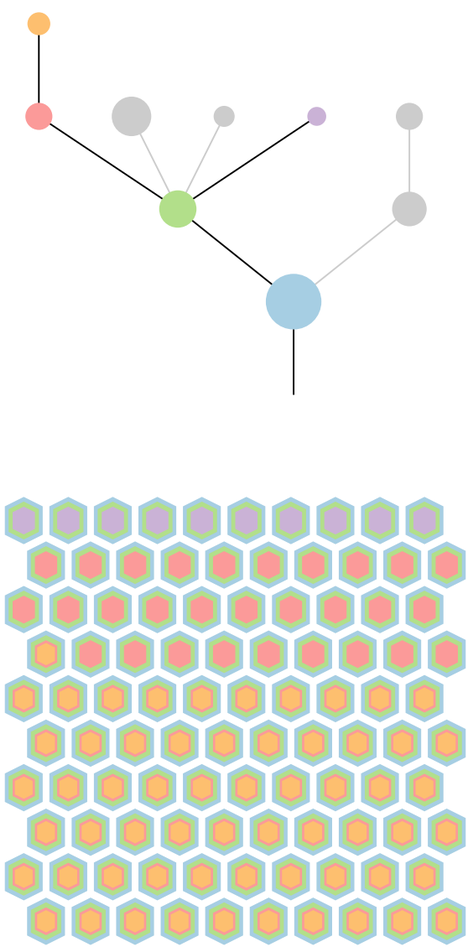
Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

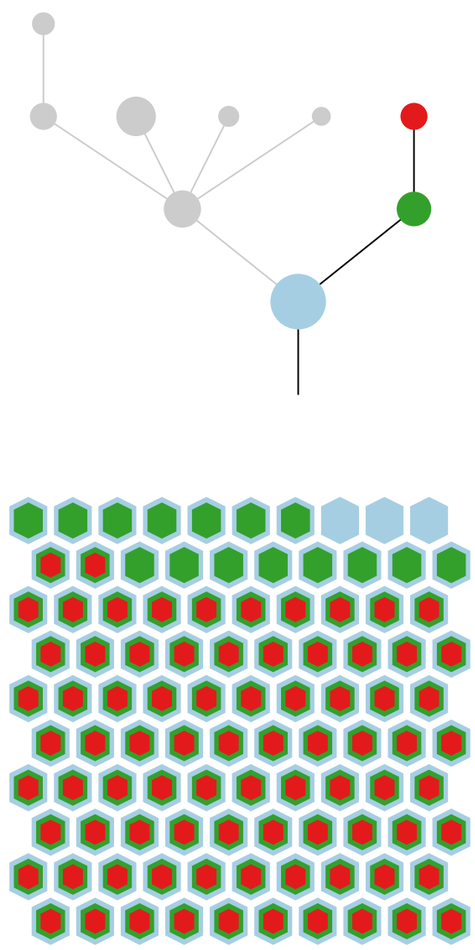
LN1



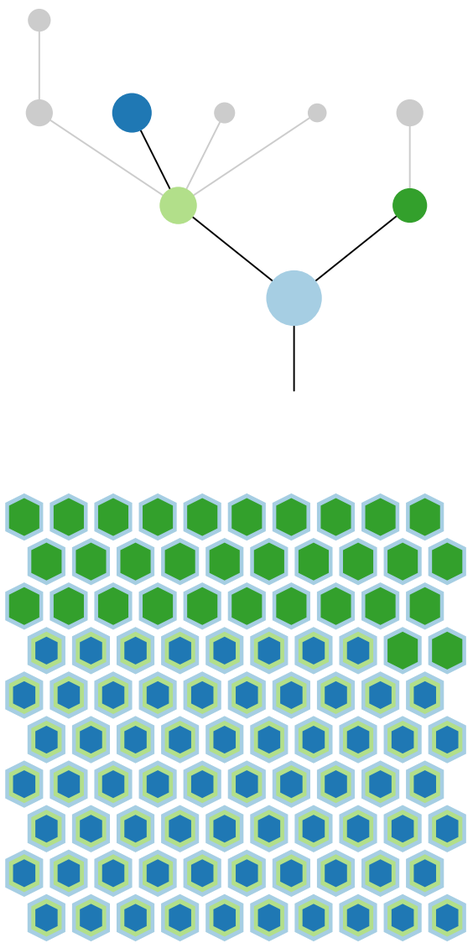
LN2



R1



R2



R3

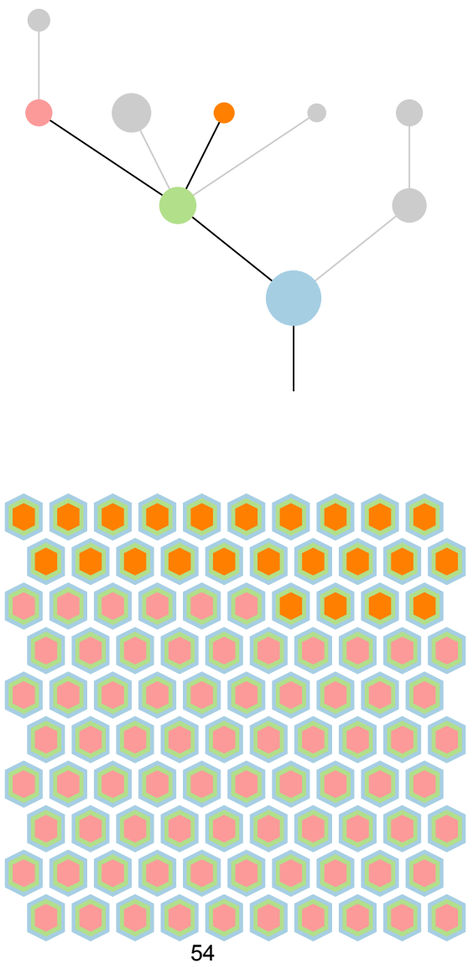
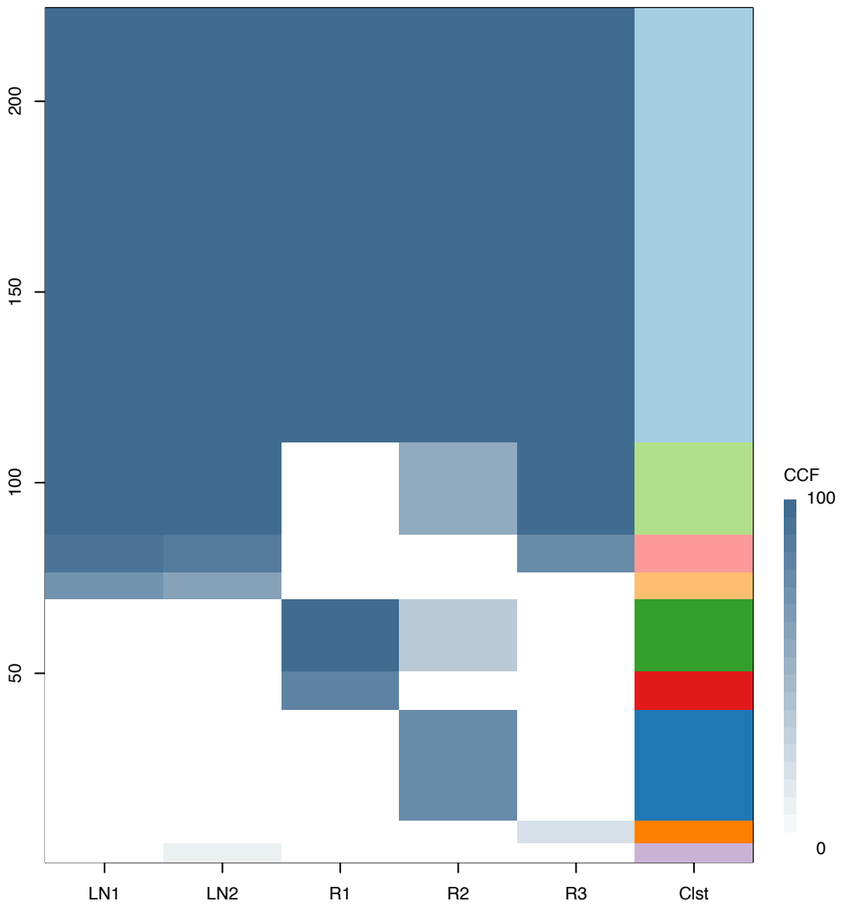
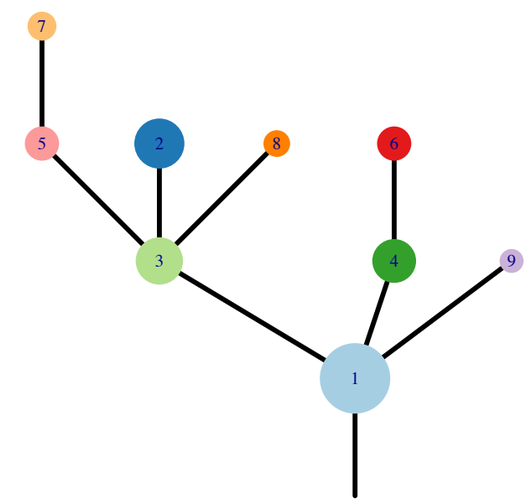


Fig.S12M



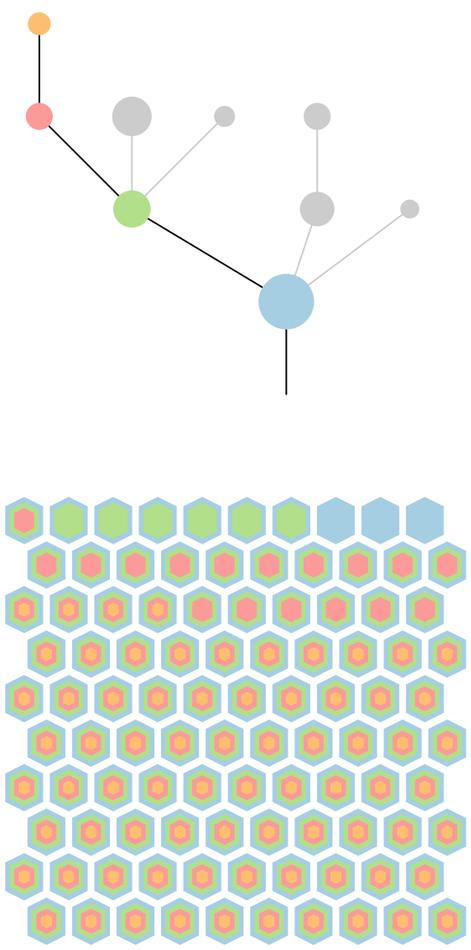
CRUK0013_G



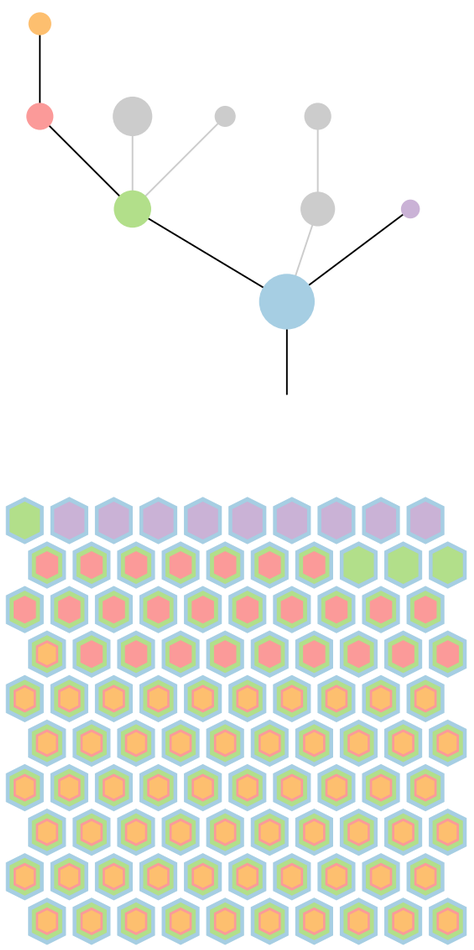
Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

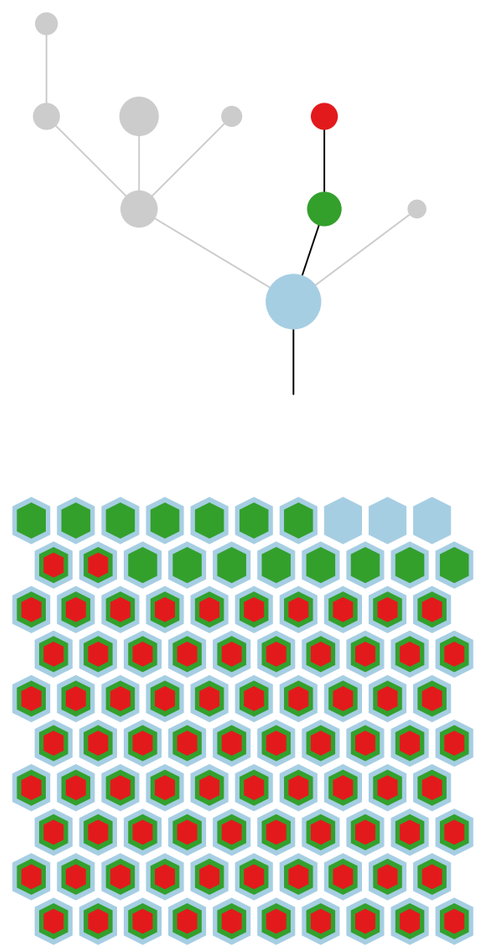
LN1



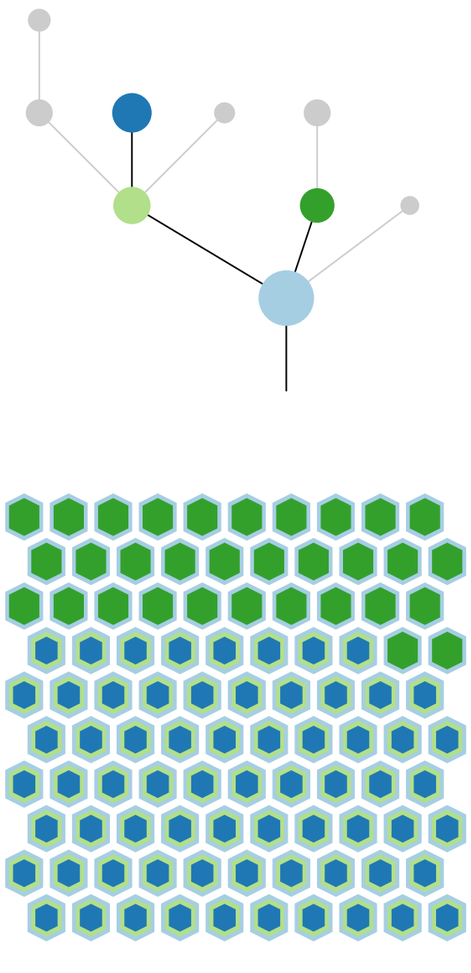
LN2



R1



R2



R3

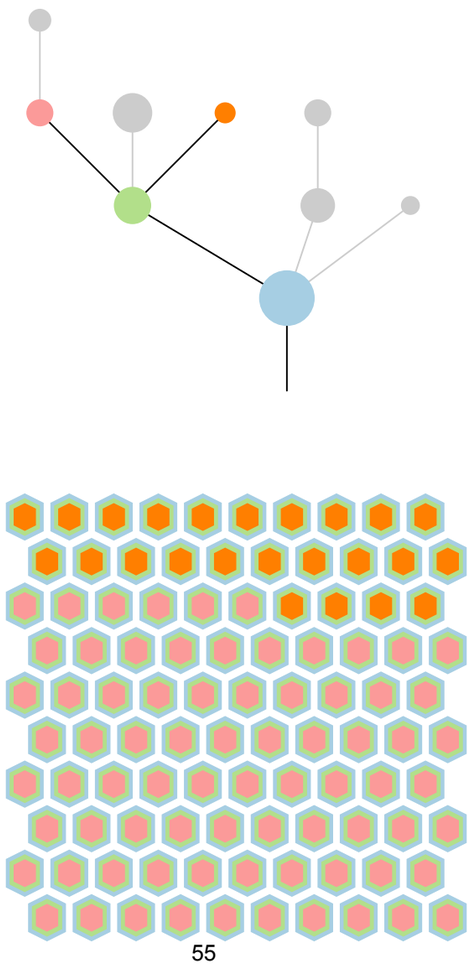
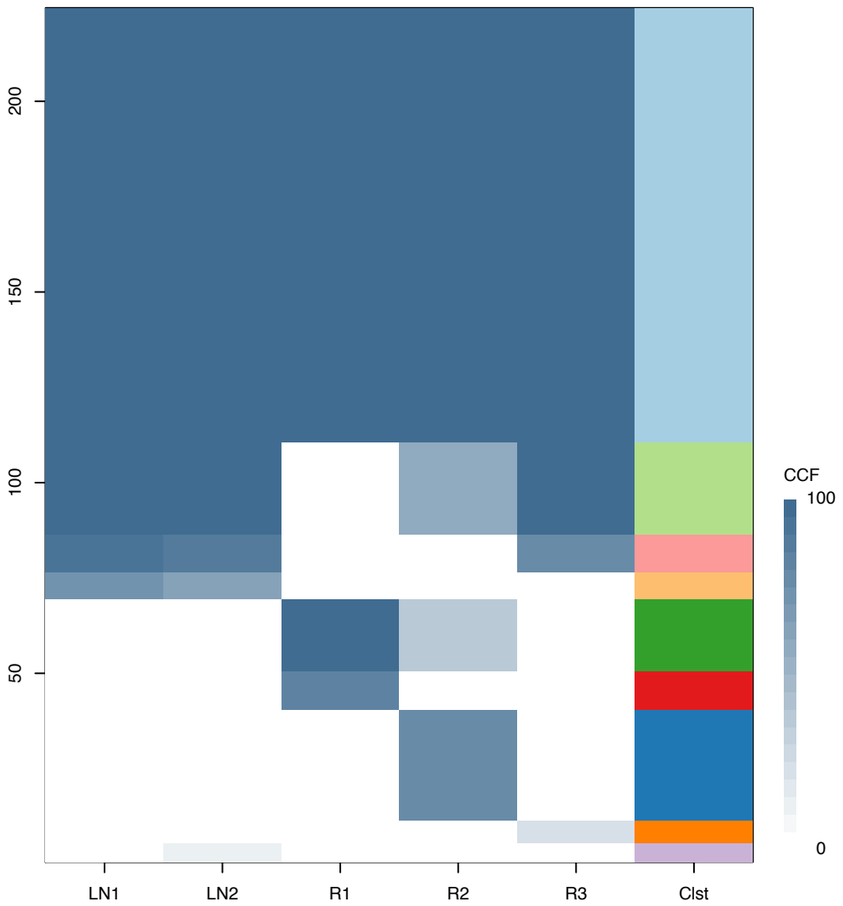
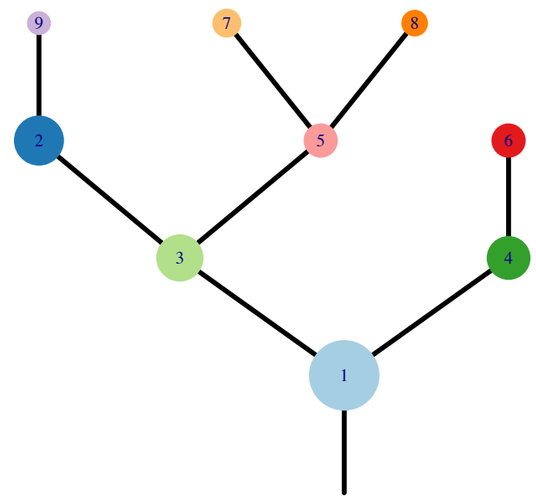


Fig.S12M



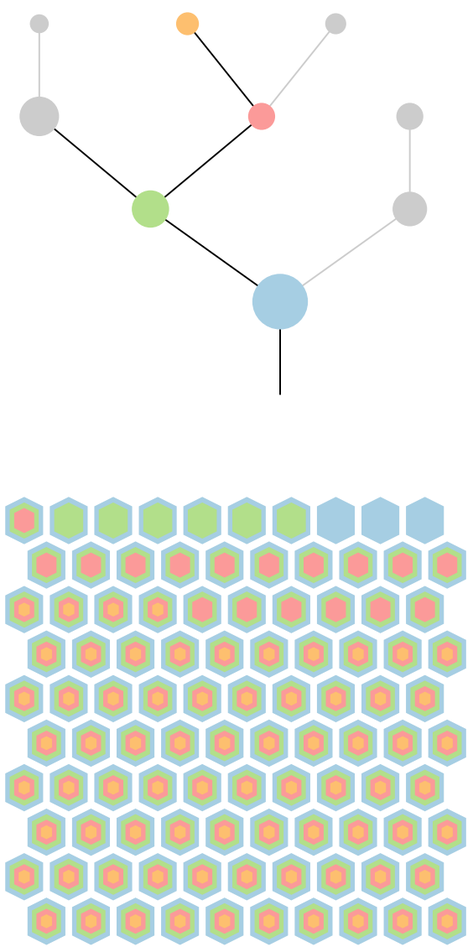
CRUK0013_H



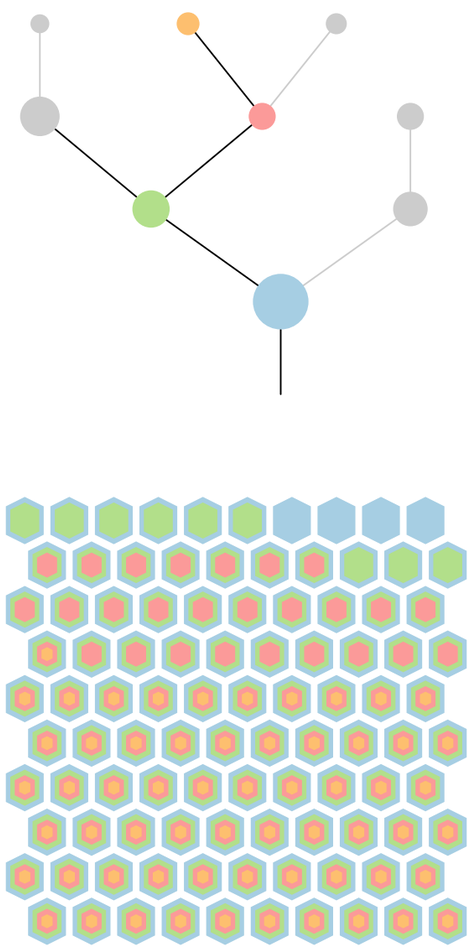
Histology:Adeno, Age:67, PackYears:26.25, Size:35
Stage:3a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
STK11	1	19p13.3	SNV
NOTCH1	2	9q34.3	SNV
KRAS	3	12p12.1	Amp
PPFIBP1	3	12p11.23	Amp
CARD11	4	7p22.2	Amp
RAC1	4	7p22.1	Amp
HNRNPA2B1	4	7p15.2	Amp
HOXA9	4	7p15.2	Amp
HOXA11	4	7p15.2	Amp
HOXA13	4	7p15.2	Amp
JAZF1	4	7p15.2	Amp
EGFR	4	7p11.2	Amp
ETV1	6	7p21.2	Amp

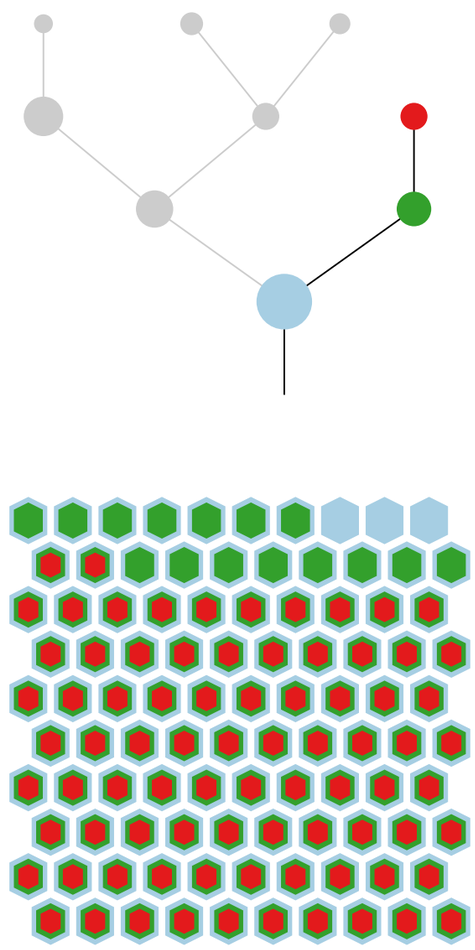
LN1



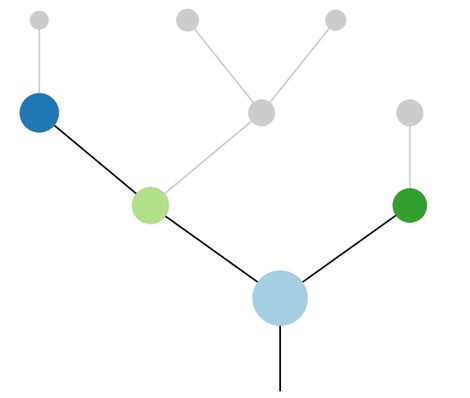
LN2



R1



R2



R3

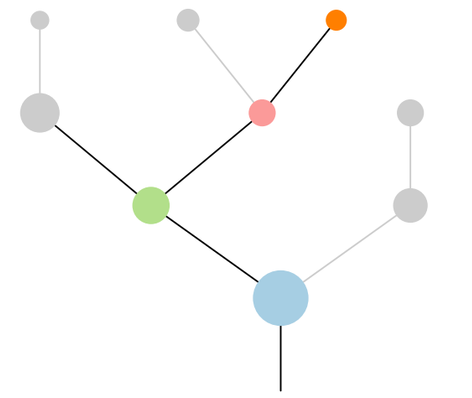
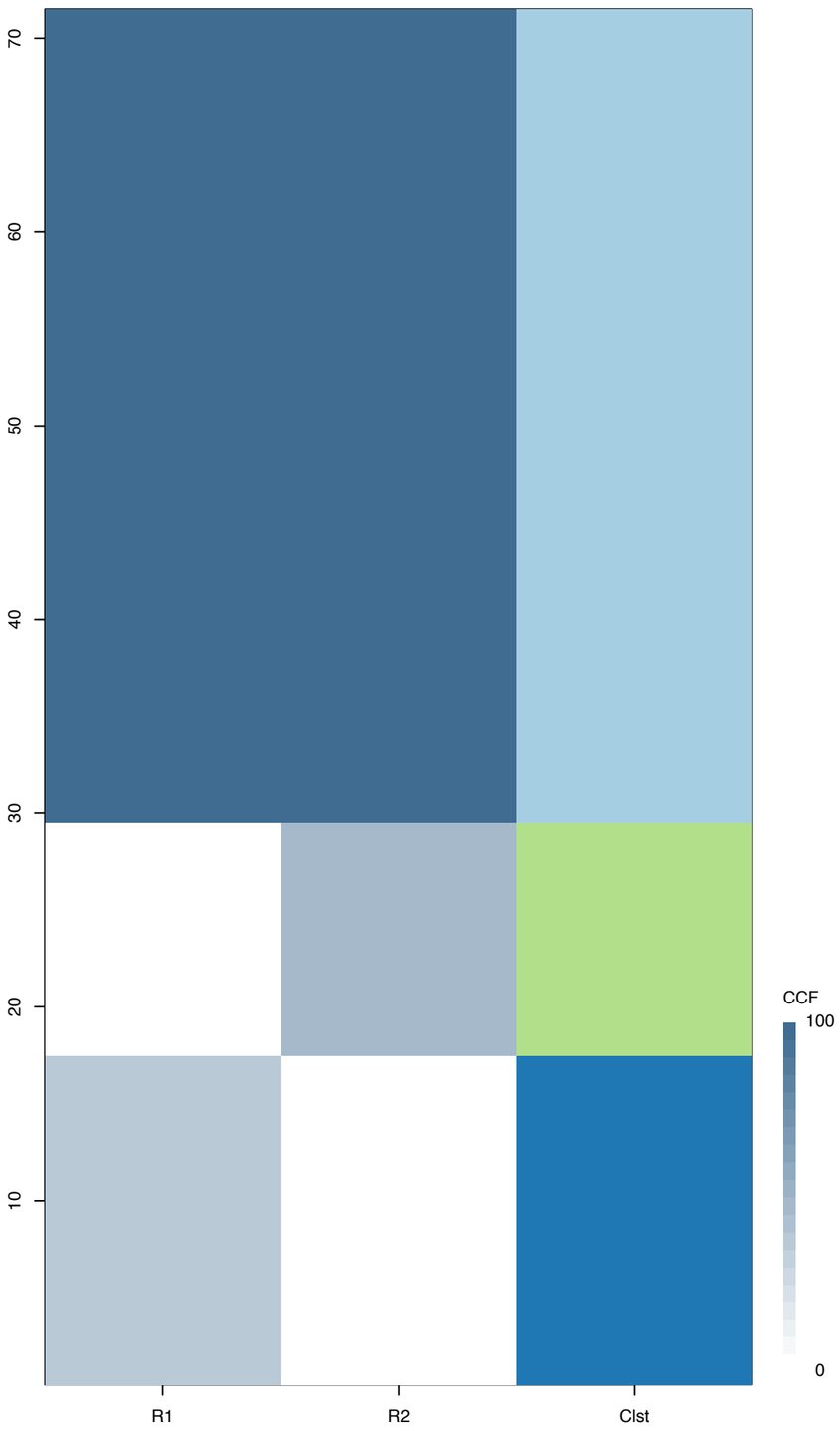
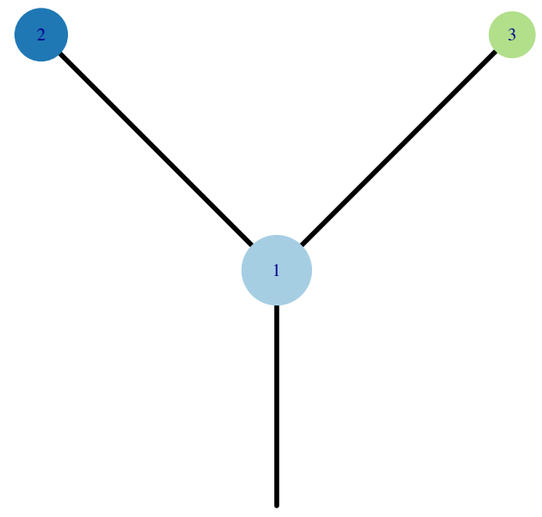


Fig.S12N

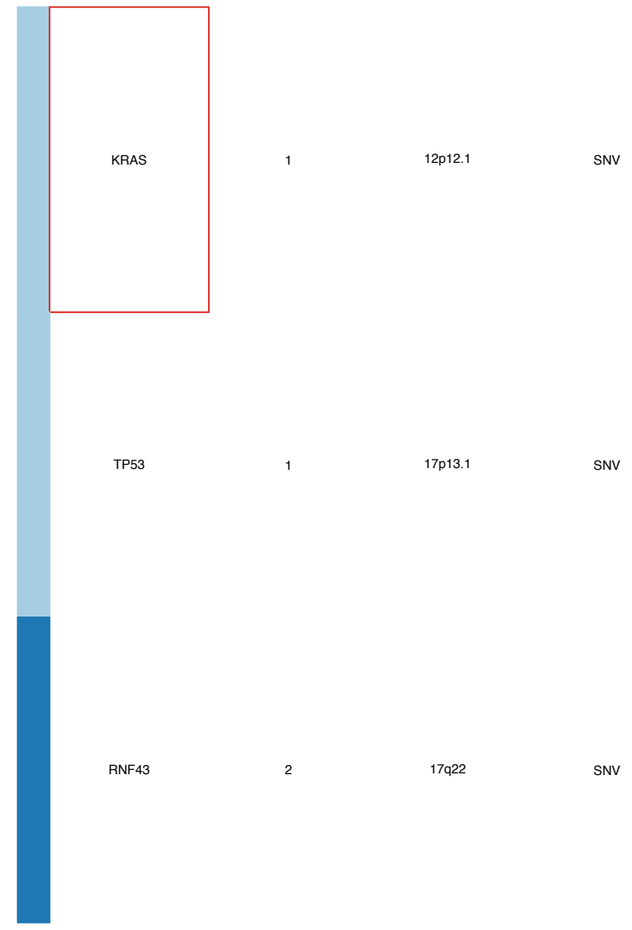


CRUK0014

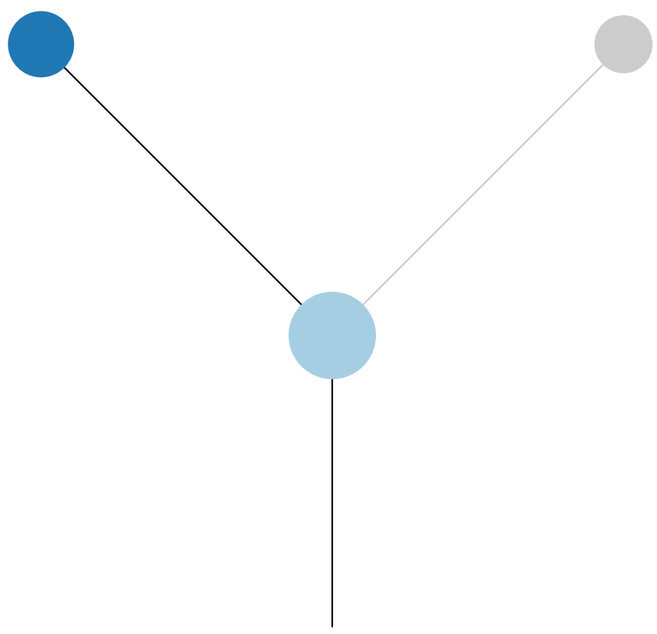


Histology:Adeno, Age:67, PackYears:NA, Size:13
Stage:1a, Gender:Male, GD:Not GD, Recur:no

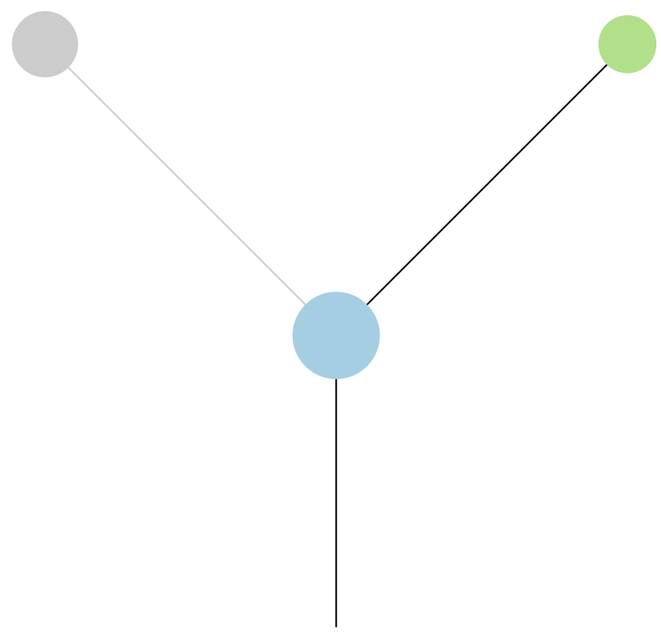
Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
TP53	1	17p13.1	SNV
RNF43	2	17q22	SNV

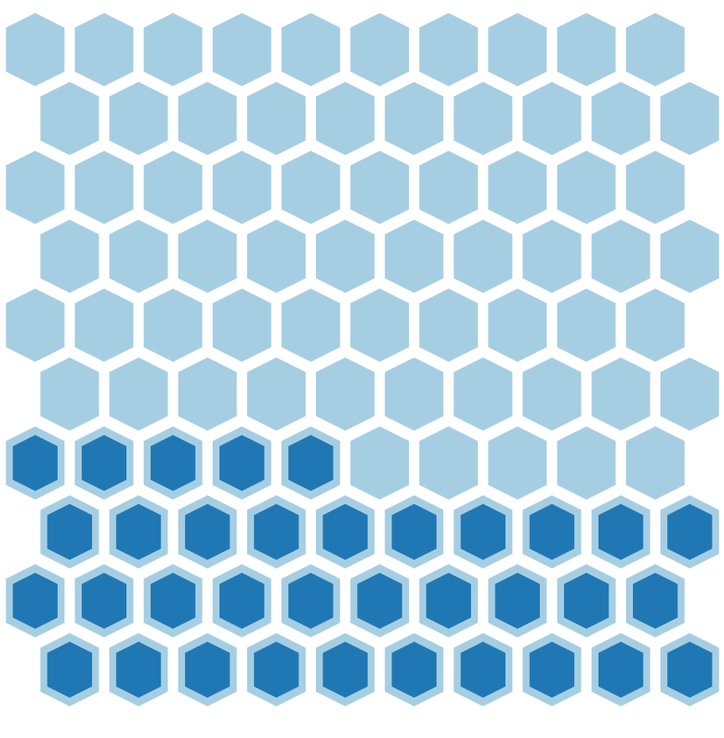


R1



R2





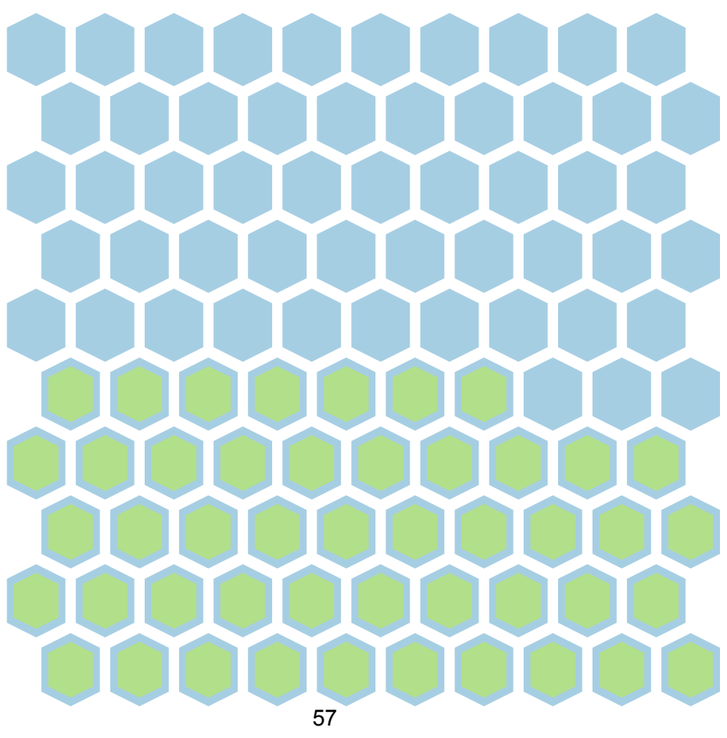
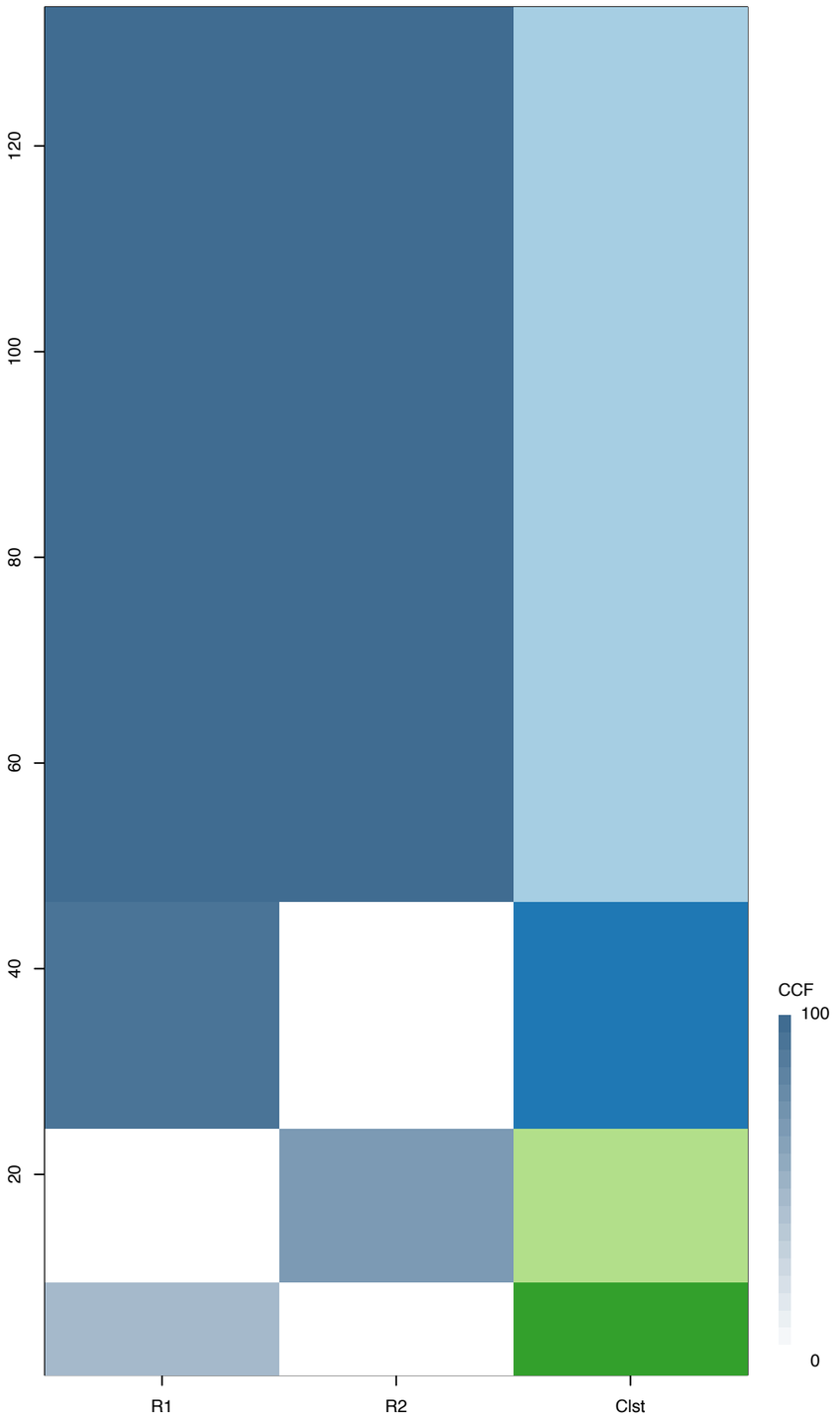
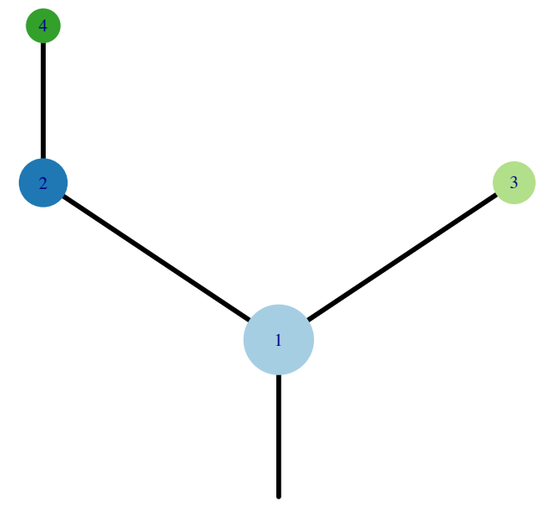


Fig.S12O



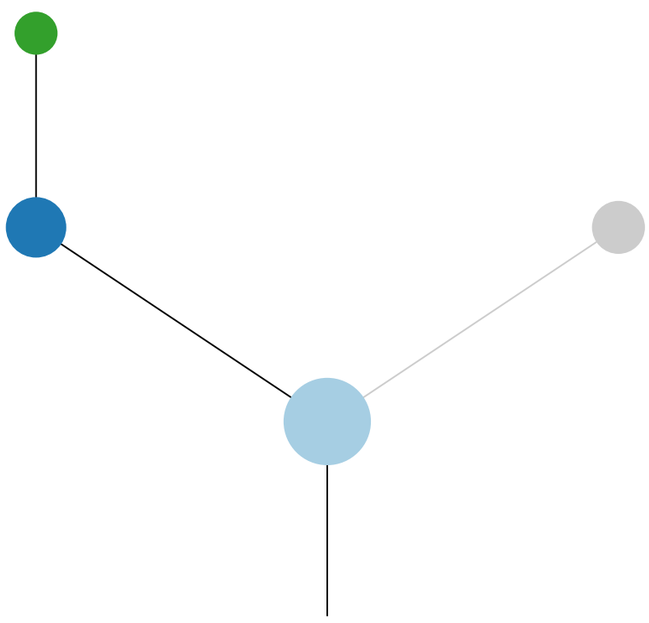
CRUK0015



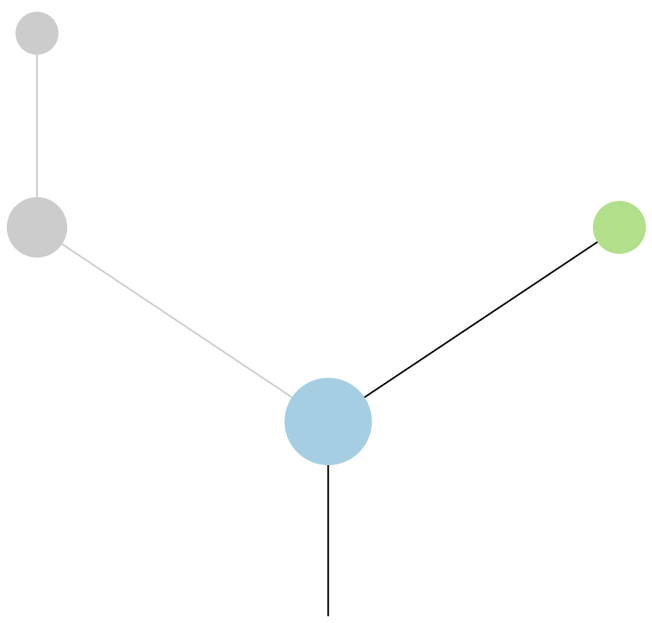
Histology:Adeno, Age:75, PackYears:NA, Size:19
Stage:1a, Gender:Female, GD:Clonal GD, Recur:no

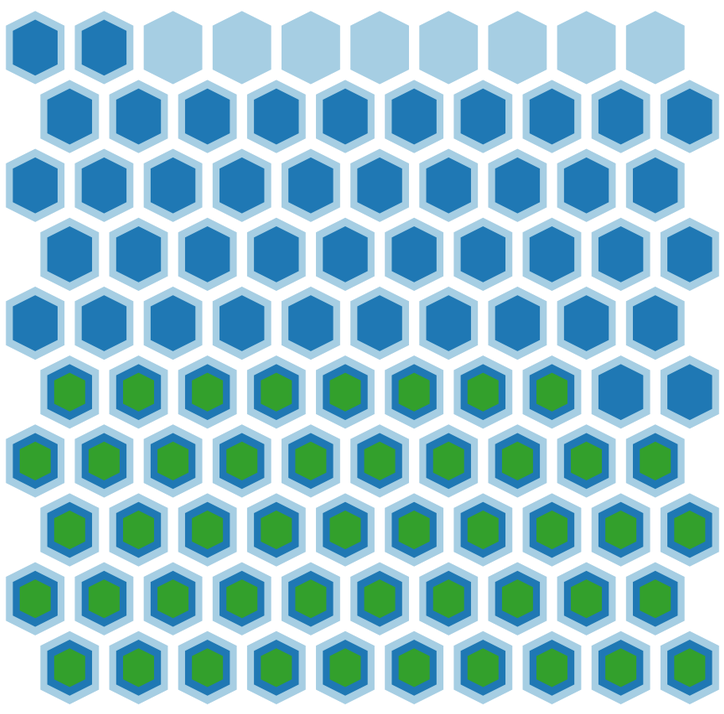
Gene	Cluster	Cytoband	Type
BAP1	1	3p21.1	SNV
EGFR	1	7p11.2	SNV
TP53	2	17p13.1	SNV
RB1	3	13q14.2	SNV

R1



R2





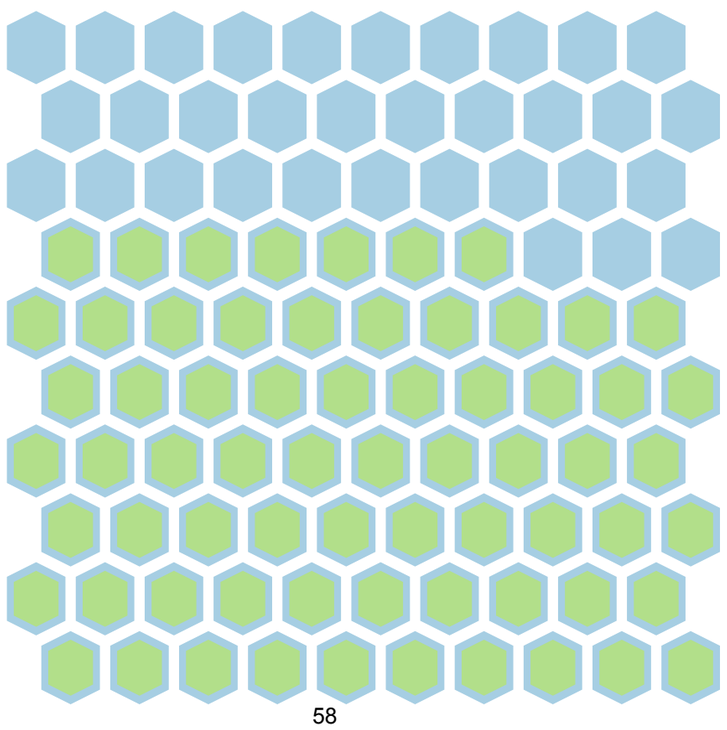
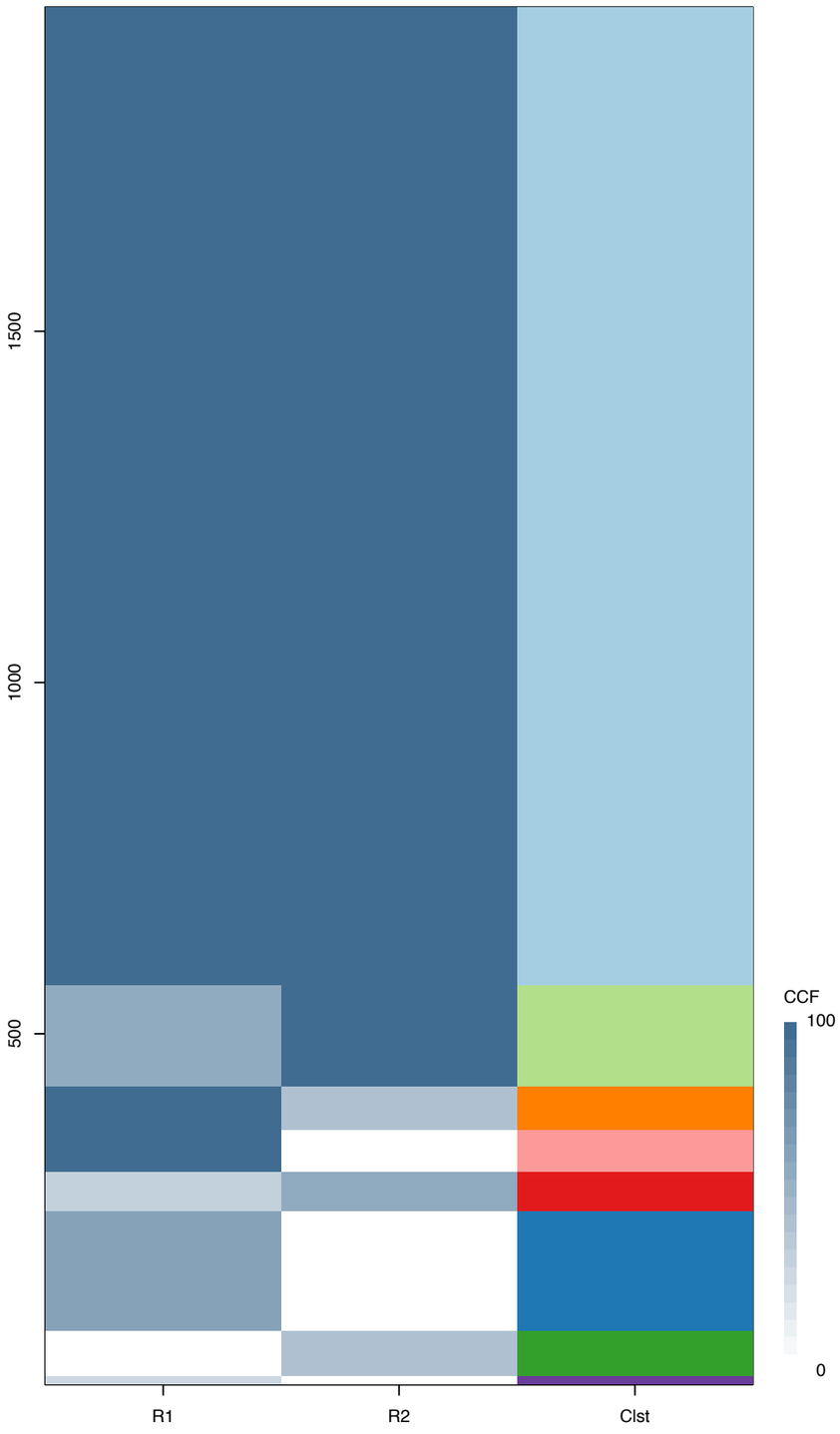
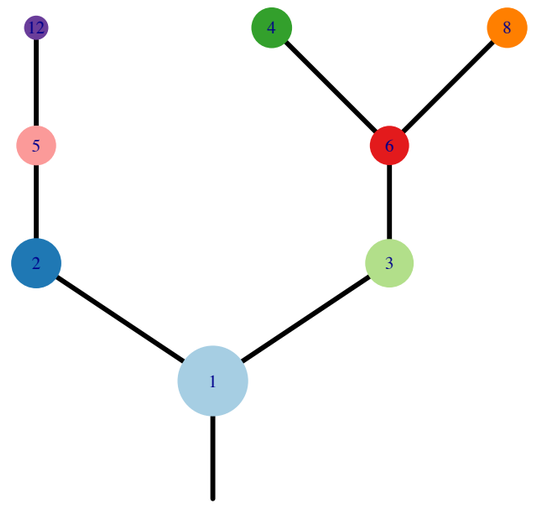


Fig.S12P

CRUK0016

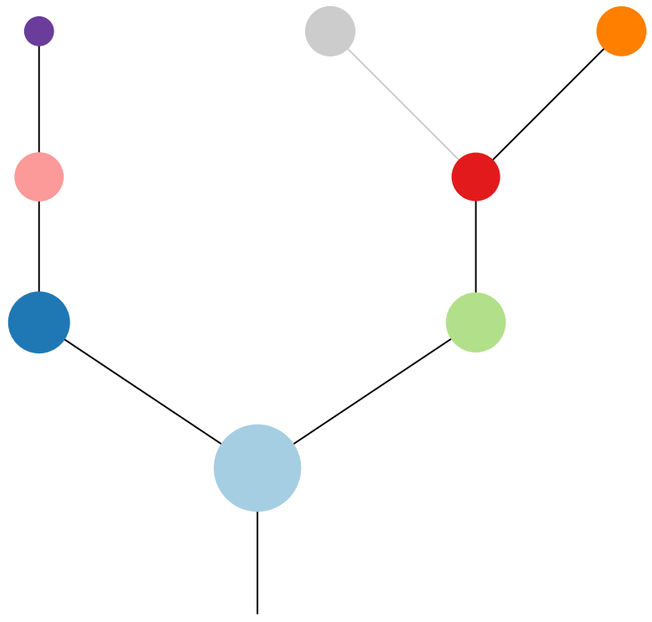




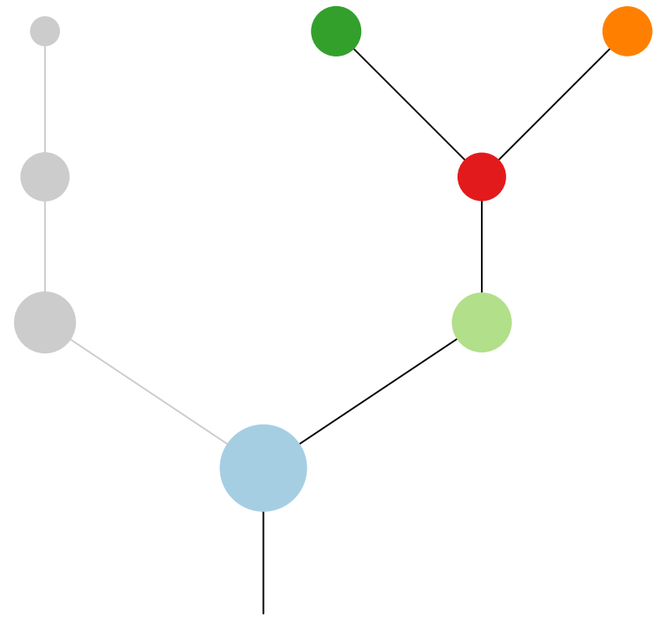
Gene	Cluster	Cytoband	Type
SPEN	1	1p36.13	SNV
CBLB	1	3q13.11	SNV
FAT1	1	4q35.2	SNV
TERT	1	5p15.33	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
CDKN2A	1	9p21.3	Del
CCDC6	1	10q21.2	Amp
KDM5A	1	12p13.33	Amp
ERC1	1	12p13.33	Amp
ZMYM2	1	13q12.11	Del
CIITA	1	16p13.13	Amp
RMI2	1	16p13.13	Amp
TNFRSF17	1	16p13.13	Amp
TP53	1	17p13.1	SNV
ERCC3	2	2q14.3	SNV
ASXL1	2	20q11.21	SNV
PTPRC	6	1q32.1	SNV
LATS1	?	6q25.1	SNV
ARID1B	?	6q25.3	SNV
DNM2	?	19p13.2	SNV
BCOR	?		SNV

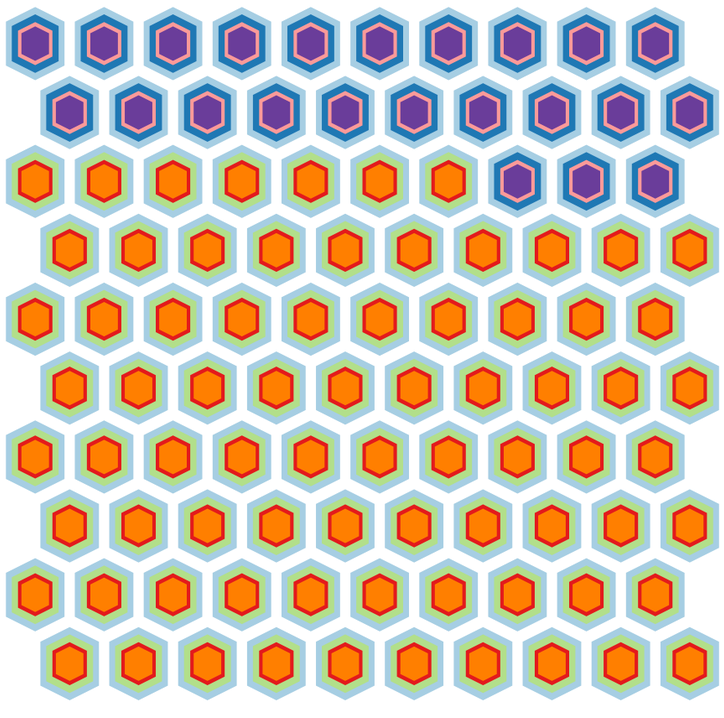
Histology:Adeno, Age:72, PackYears:73.5, Size:28
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

R1



R2





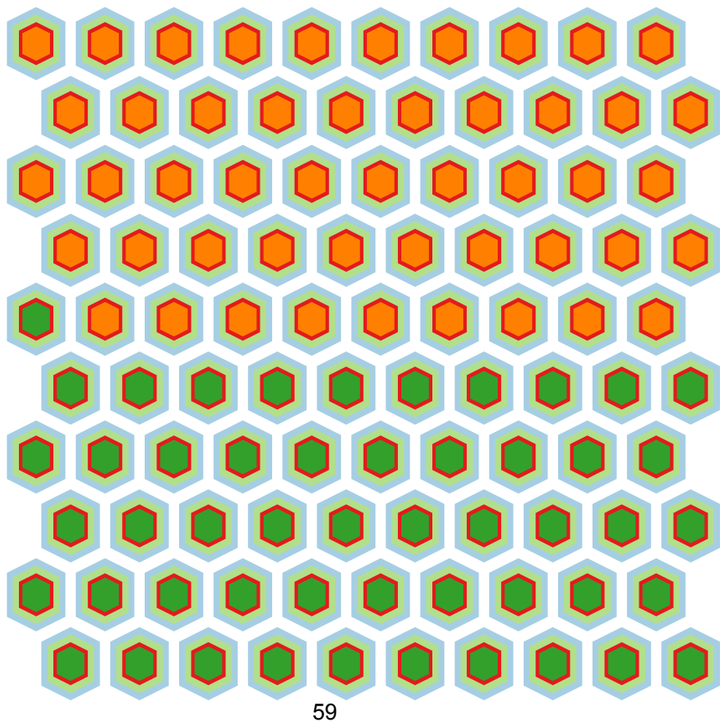
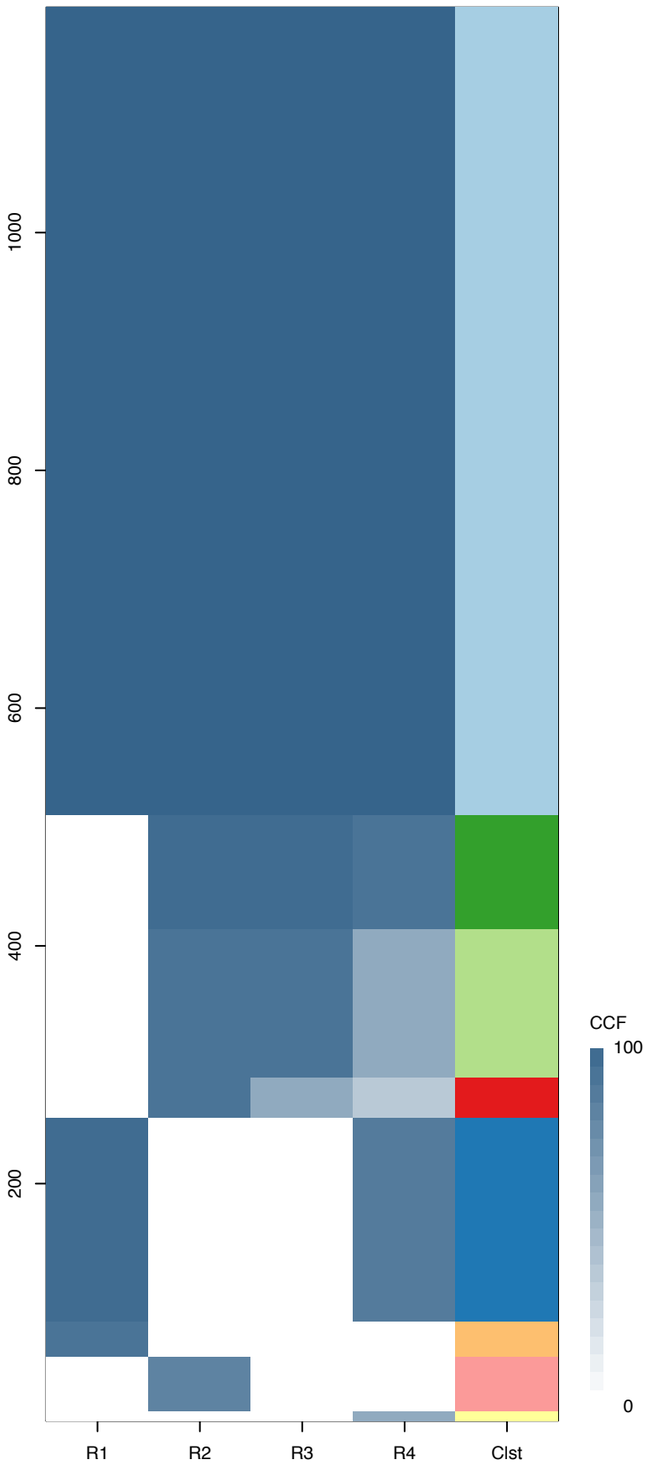
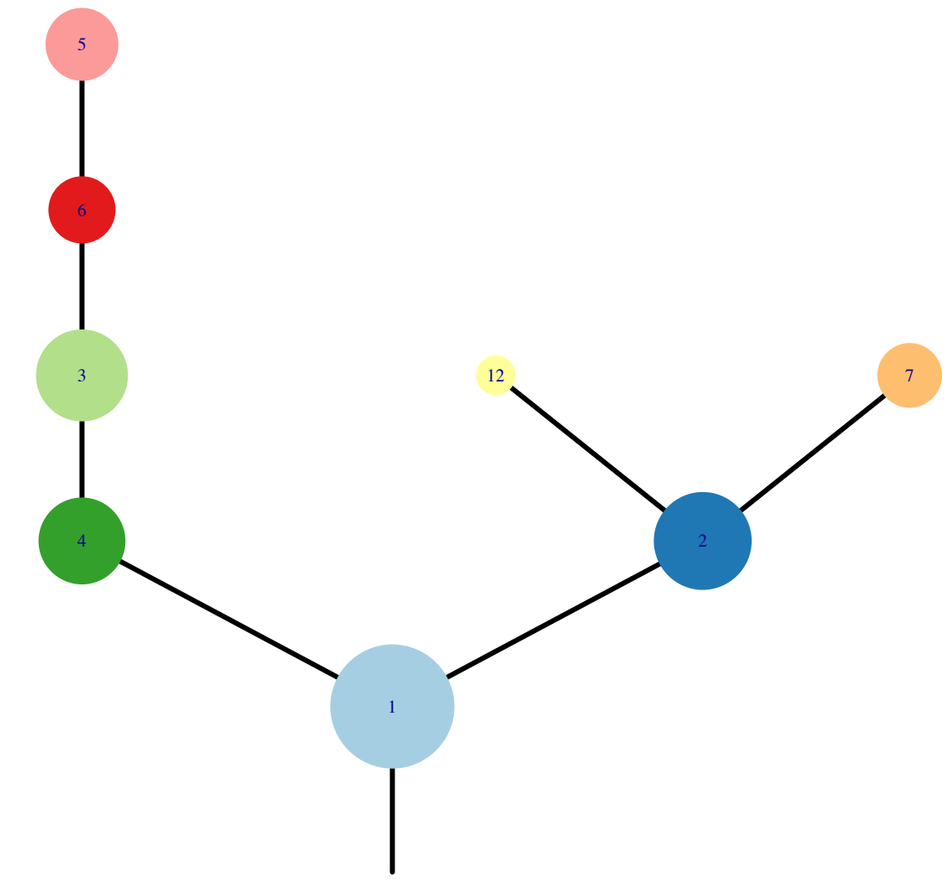


Fig.S12Q

CRUK0017





Histology:Adeno, Age:79, PackYears:42, Size:18
Stage:1b, Gender:Female, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
ARID1B	1	6q25.3	SNV
MYC	1	8q24.21	Amp
NDRG1	1	8q24.22	Amp
EXT2	1	11p11.2	SNV
ARID2	1	12q12	SNV
DDIT3	1	12q13.3	Amp
CDK4	1	12q14.1	Amp
TP53	1	17p13.1	SNV
KEAP1	1	19p13.2	SNV
TPM4	1	19p13.12	Amp
CCNE1	1	19q12	Amp
CEP89	1	19q13.11	Amp
CEBPA	1	19q13.11	Amp
LSM14A	1	19q13.11	Amp
TCEA1	2	8q11.23	Amp
PLAG1	2	8q12.1	Amp
CHCHD7	2	8q12.1	Amp
NCOA2	2	8q13.3	Amp
HEY1	2	8q21.13	Amp
CNBD1	2	8q21.3	Amp
RUNX1T1	2	8q21.3	Amp
COX6C	2	8q22.2	Amp
RSPO2	2	8q23.1	Amp
EIF3E	2	8q23.1	Amp
SS18L1	2	20q13.33	Amp
KRAS	3	12p12.1	SNV
CLTCL1	3	22q11.21	Amp
SEPT5	3	22q11.21	Amp
EWSR1	3	22q12.2	Amp
PDGFB	3	22q13.1	Amp
DICER1	12	14q32.13	SNV
NOTCH2	?	1p12	SNV
MSH2	?	2p21	Del

R1 R2 R3 R4

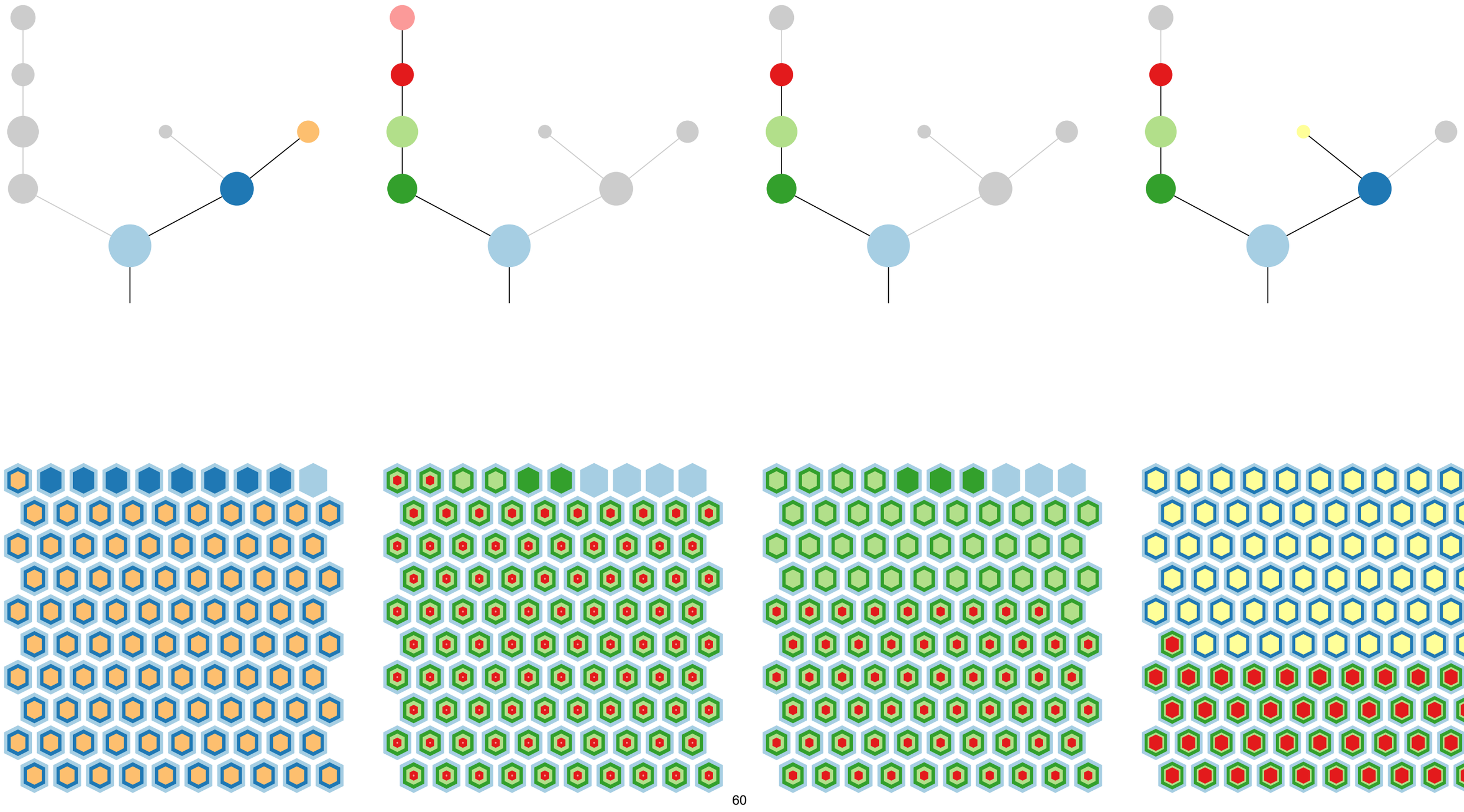
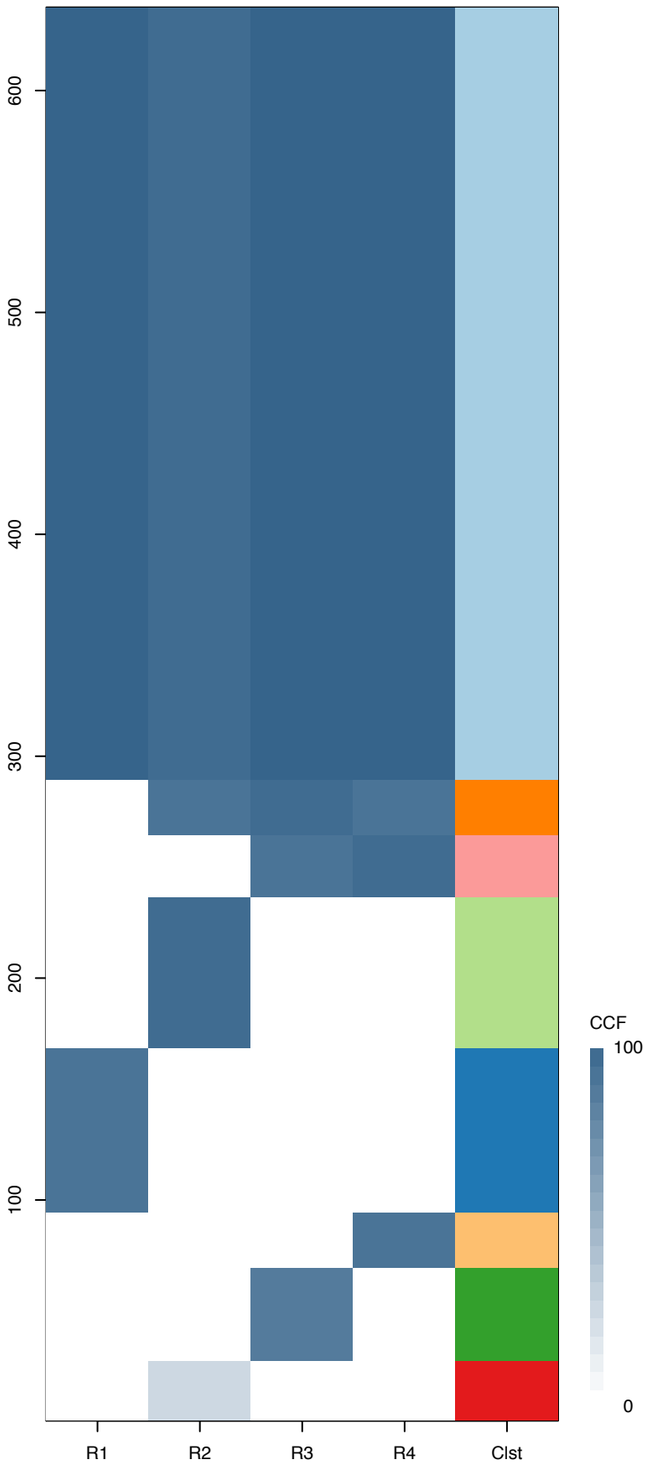
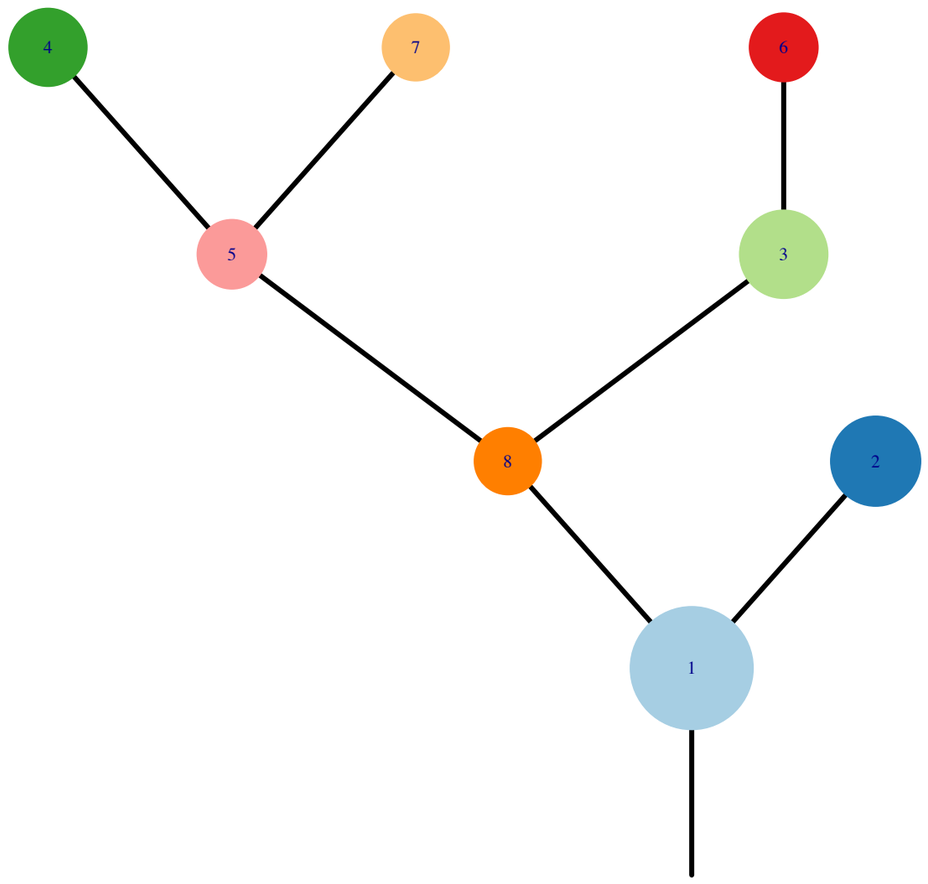


Fig.S12R



CRUK0018



Histology:Adeno, Age:60, PackYears:27.75, Size:24
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
COL5A2	1	2q32.2	SNV
KRAS	1	12p12.1	SNV
MGA	1	15q15.1	SNV
CMTR2	1	16q22.2	SNV
CARD11	8	7p22.2	Amp
RAC1	8	7p22.1	Amp
DNMT3A	?	2p23.3	SNV
WAS	?		SNV

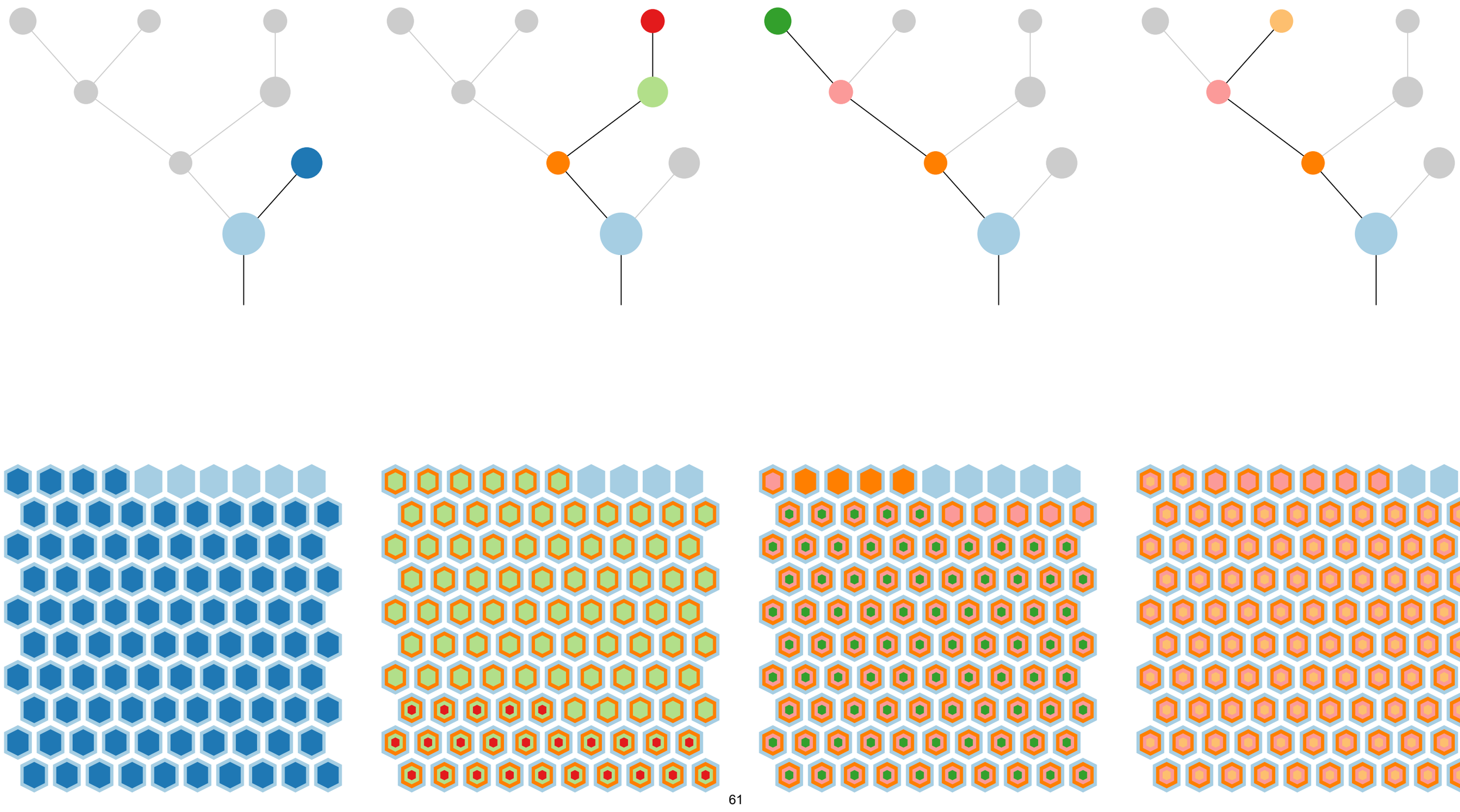
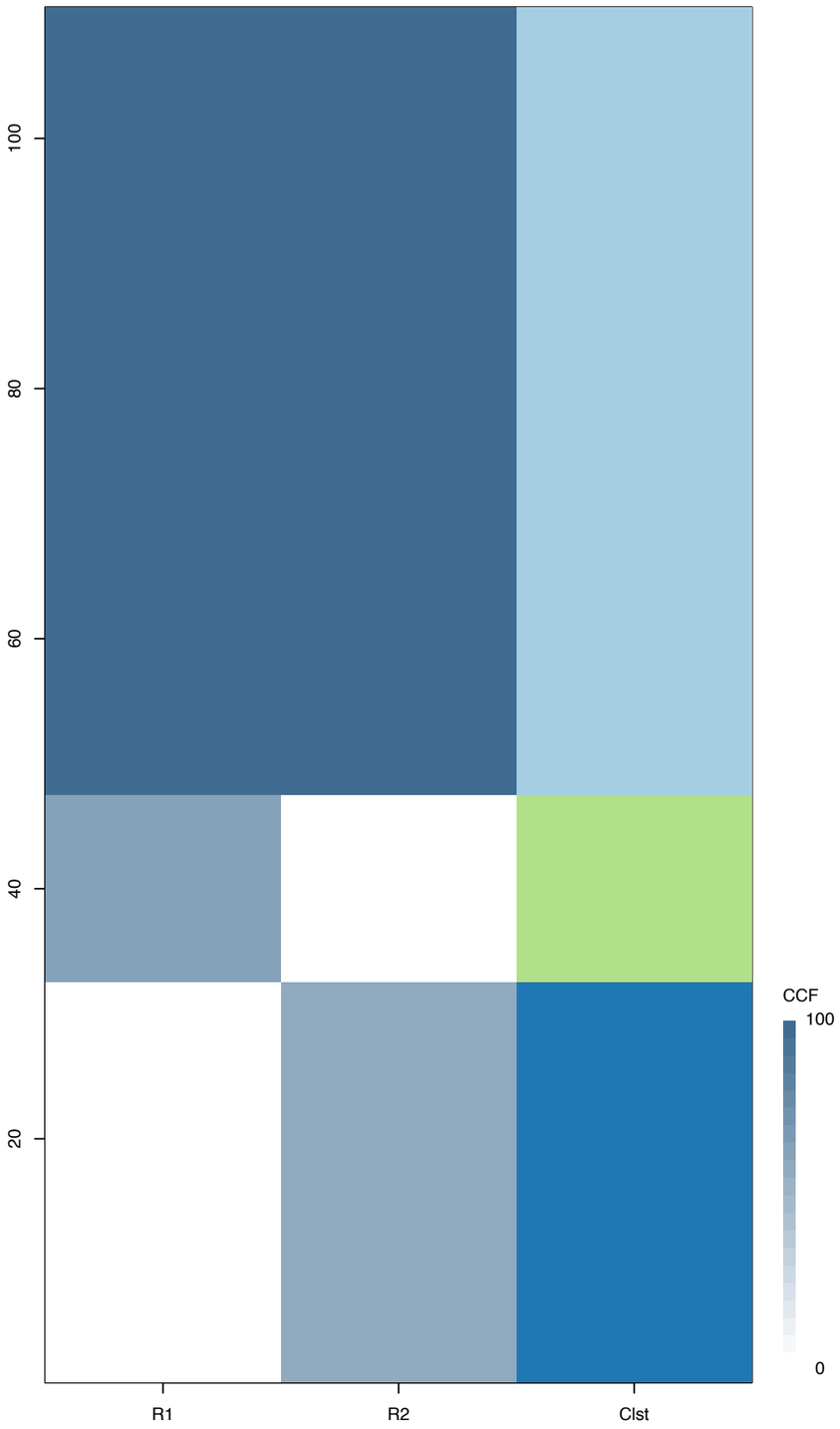
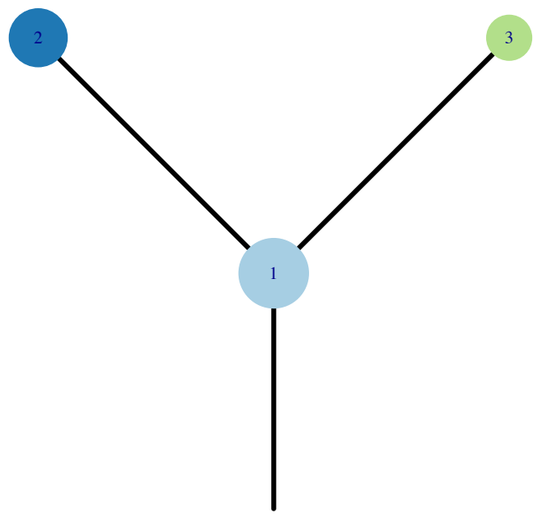


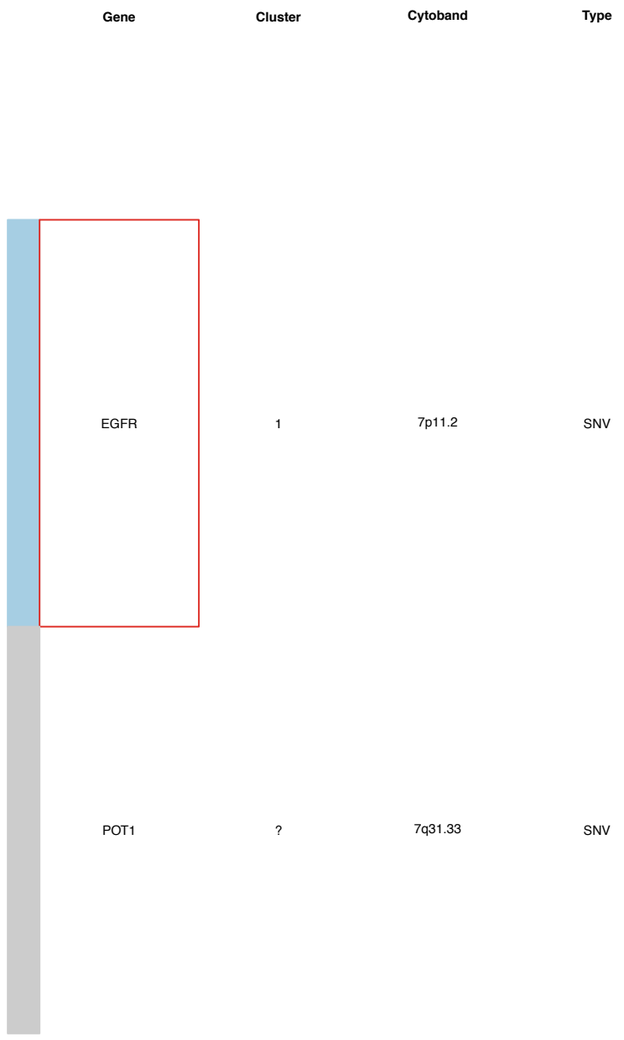
Fig.S12S



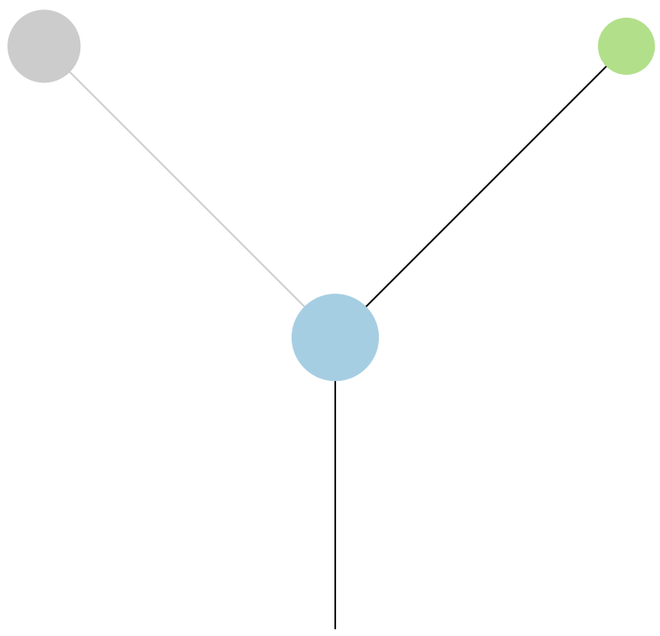
CRUK0019

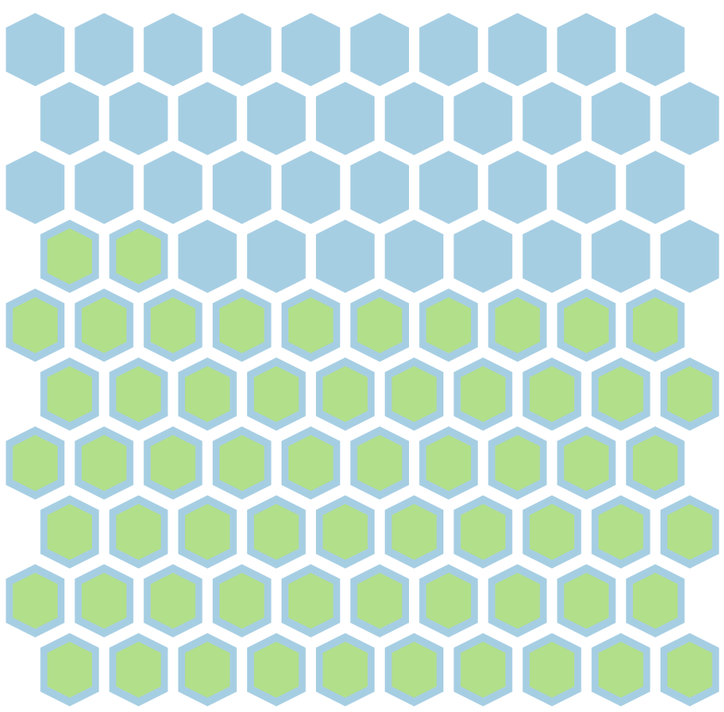


Histology:Adeno, Age:55, PackYears:NA, Size:32
Stage:1b, Gender:Female, GD:Clonal GD, Recur:no

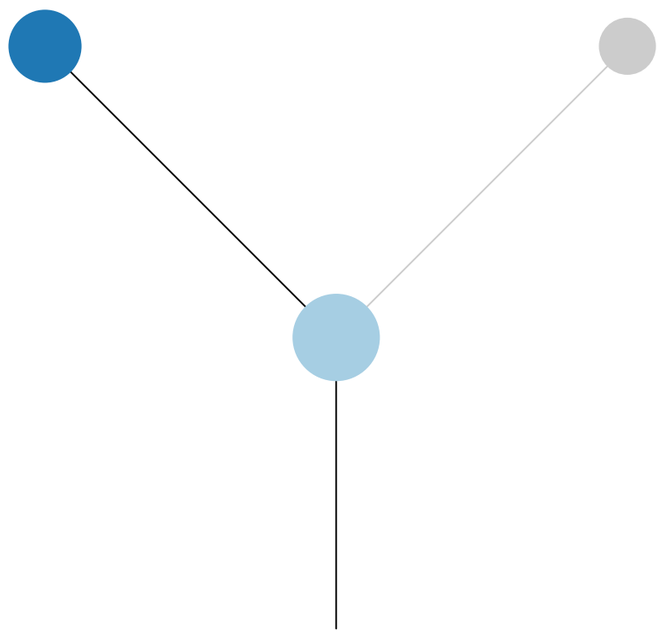


R1





R2



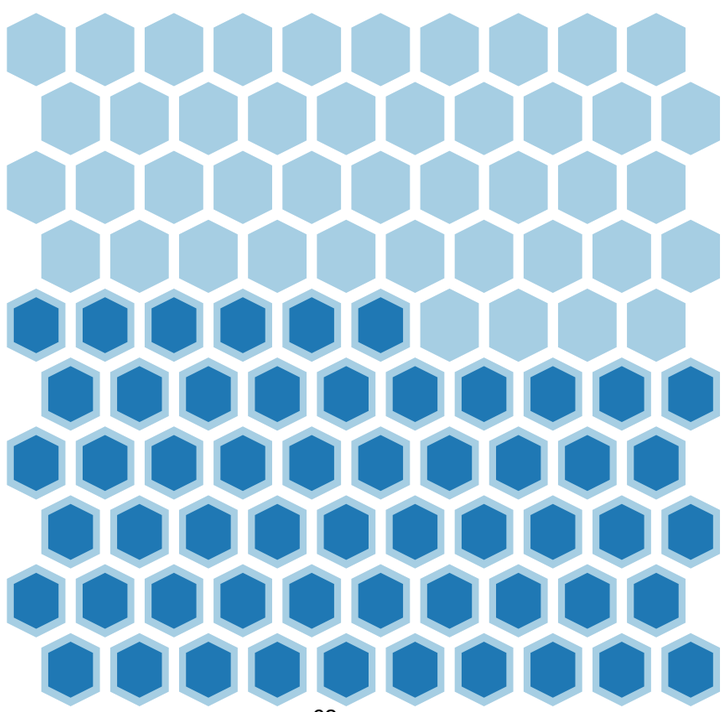
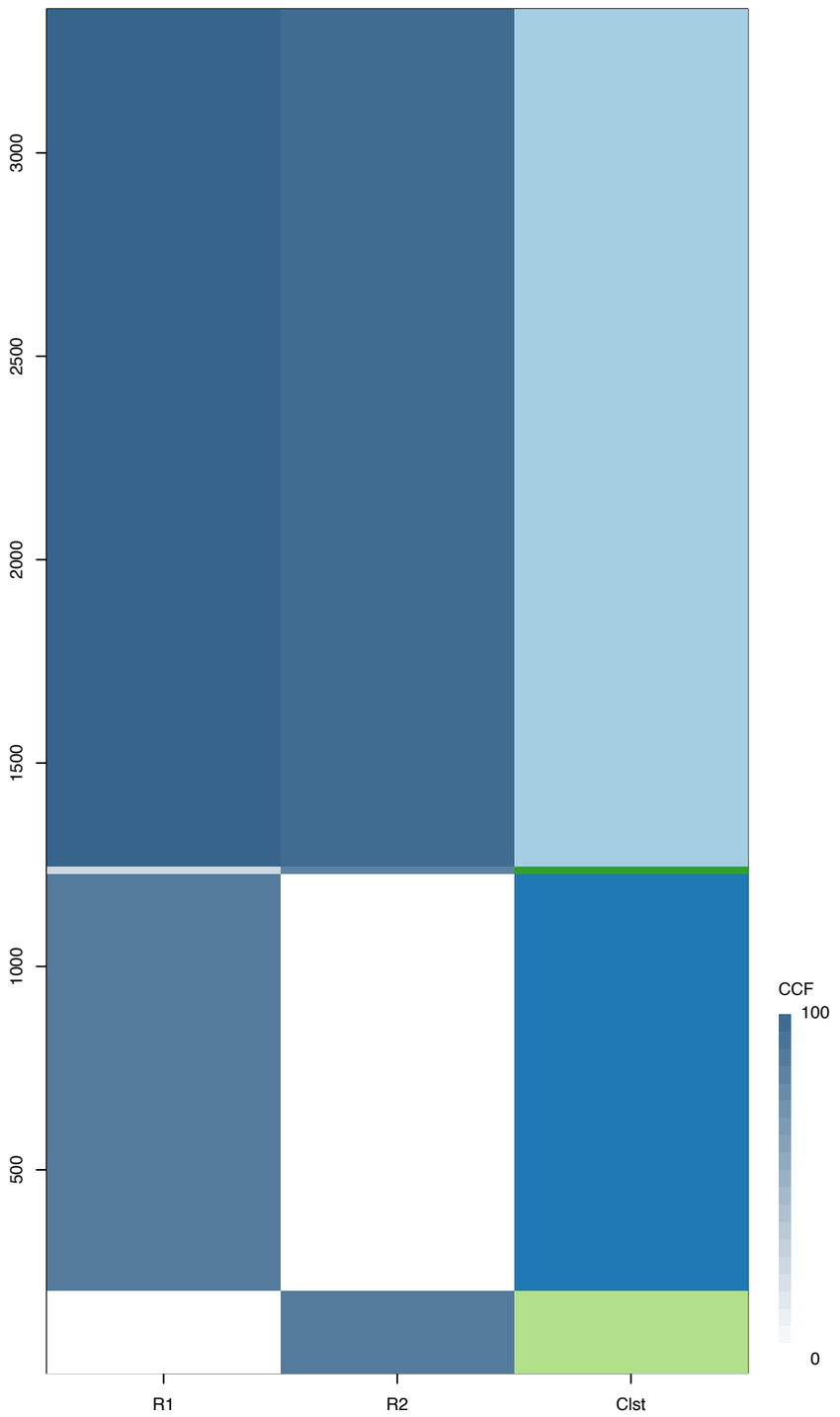
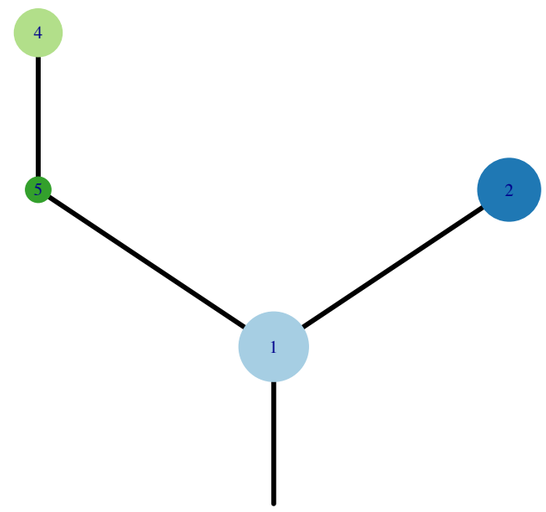


Fig.S12T



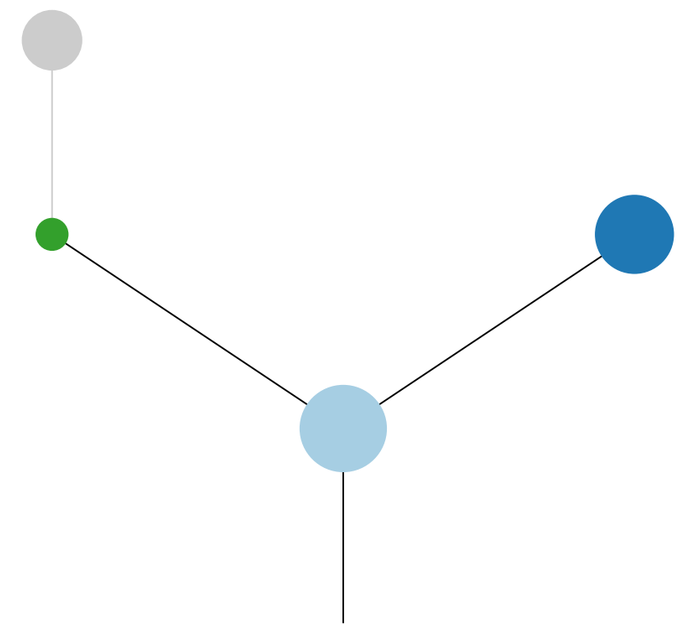
CRUK0020

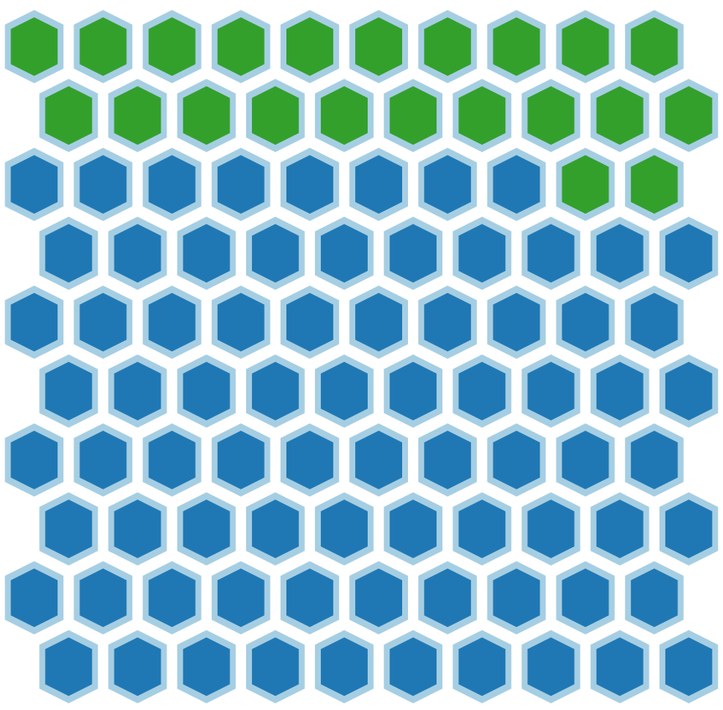


Histology:Adeno, Age:84, PackYears:102, Size:19
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

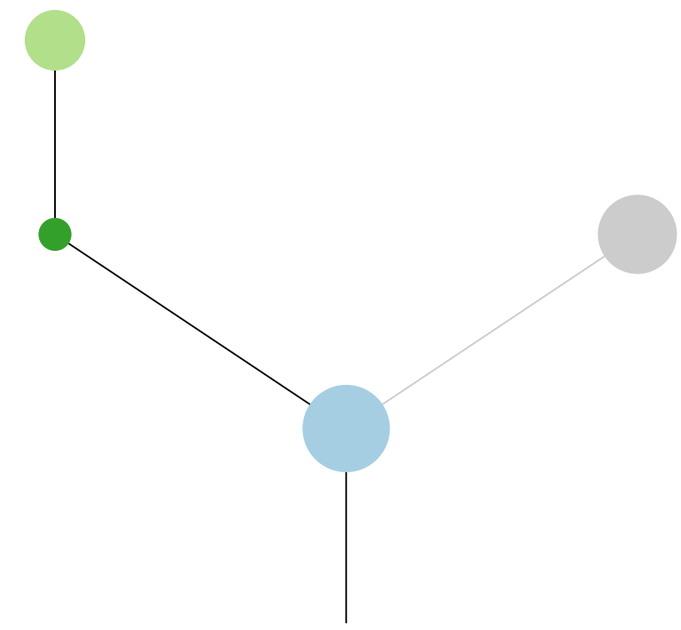
Gene	Cluster	Cytoband	Type
FANCD2	1	3p25.3	SNV
TFEB	1	6p21.1	Amp
CCND3	1	6p21.1	Amp
PRF1	1	10q22.1	SNV
ETNK1	1	12p12.1	Amp
KRAS	1	12p12.1	Amp
ARID2	1	12q12	SNV
COL2A1	1	12q13.11	SNV
MGA	1	15q15.1	SNV
TP53	1	17p13.1	SNV
TAF15	1	17q12	Amp
KEAP1	1	19p13.2	SNV
CDC73	2	1q31.2	SNV
BAP1	2	3p21.1	SNV
PIK3CA	2	3q26.32	SNV
JAK2	4	9p24.1	Amp
CD274	4	9p24.1	Amp
PDCD1LG2	4	9p24.1	Amp
NCOR1	4	17p12	SNV

R1





R2



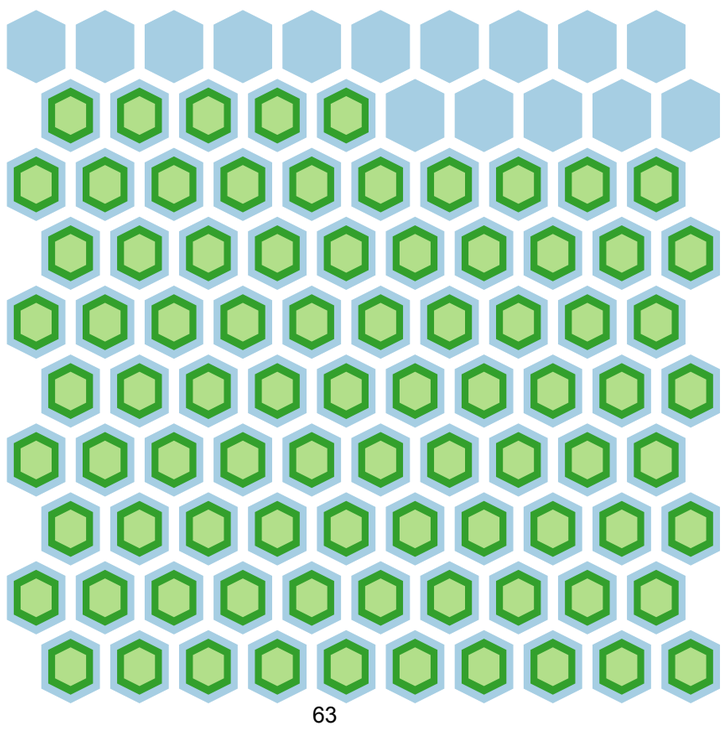
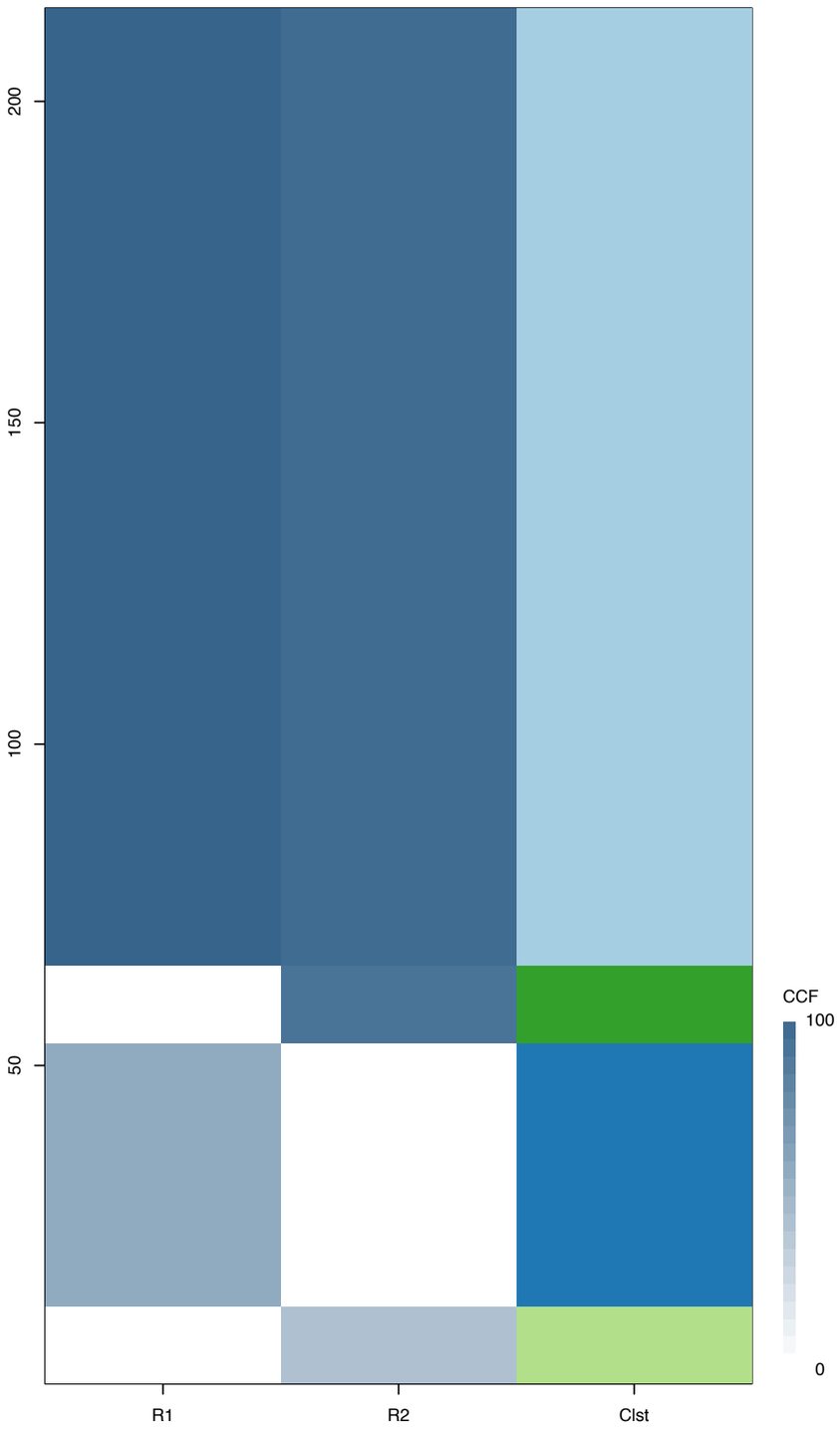
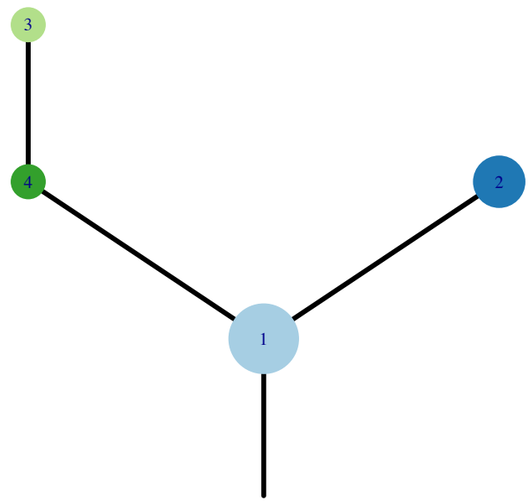


Fig.S12U



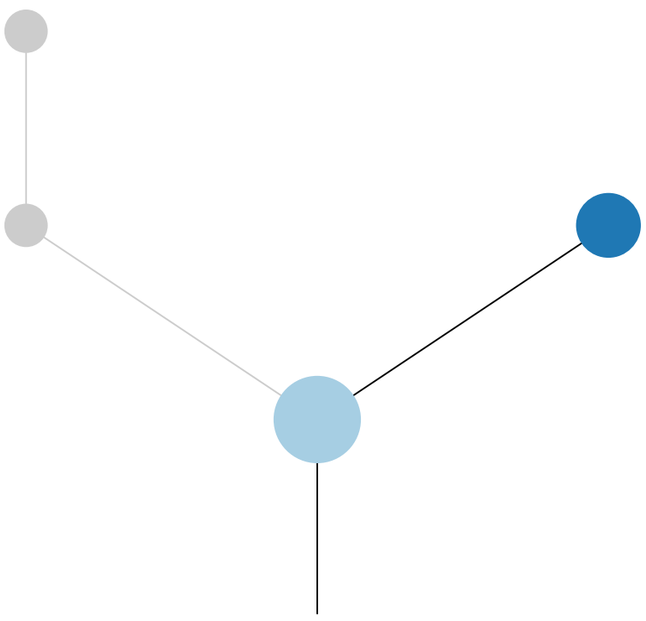
CRUK0021



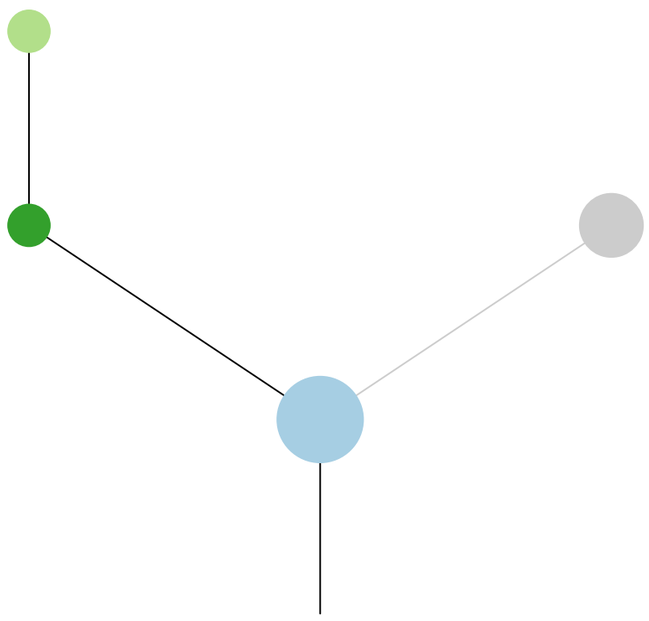
Histology:Adeno, Age:84, PackYears:NA, Size:18
Stage:1a, Gender:Female, GD:Clonal GD, Recur:yes

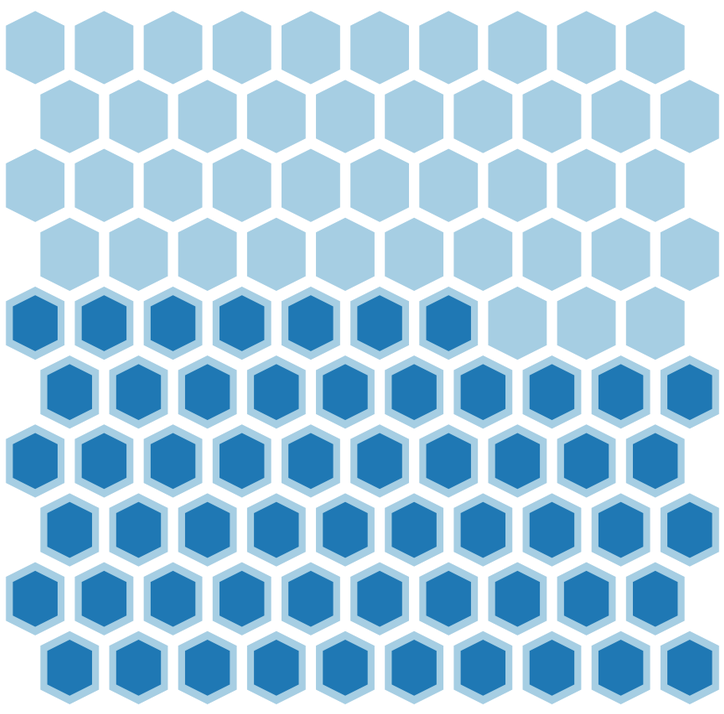
Gene	Cluster	Cytoband	Type
EBF1	1	5q33.3	Amp
PWWP2A	1	5q33.3	Amp
RANBP17	1	5q35.1	Amp
TLX3	1	5q35.1	Amp
NPM1	1	5q35.1	Amp
EGFR	1	7p11.2	Amp
EGFR	1	7p11.2	SNV
CDKN2A	1	9p21.3	Del
TP53	1	17p13.1	SNV
CHEK2	1	22q12.1	SNV
RBM10	?		SNV

R1



R2





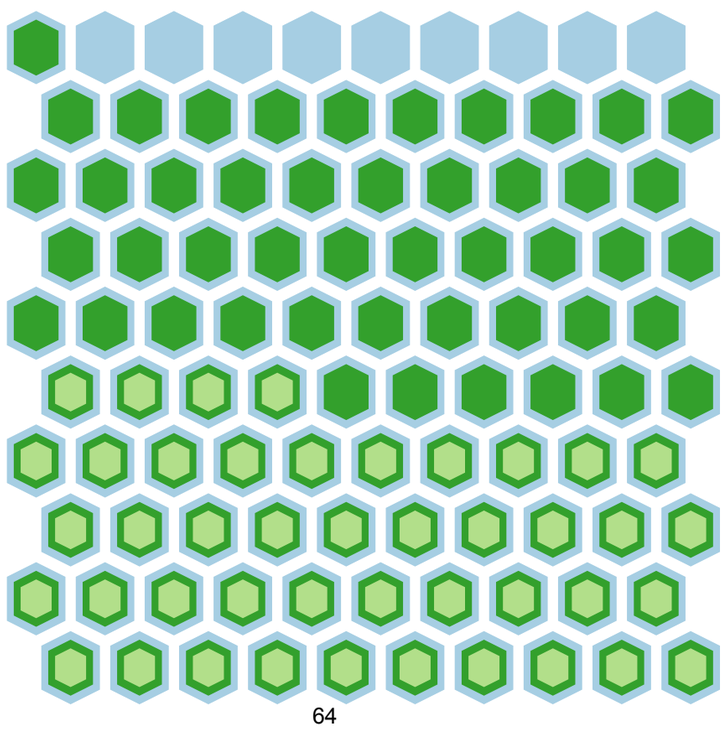
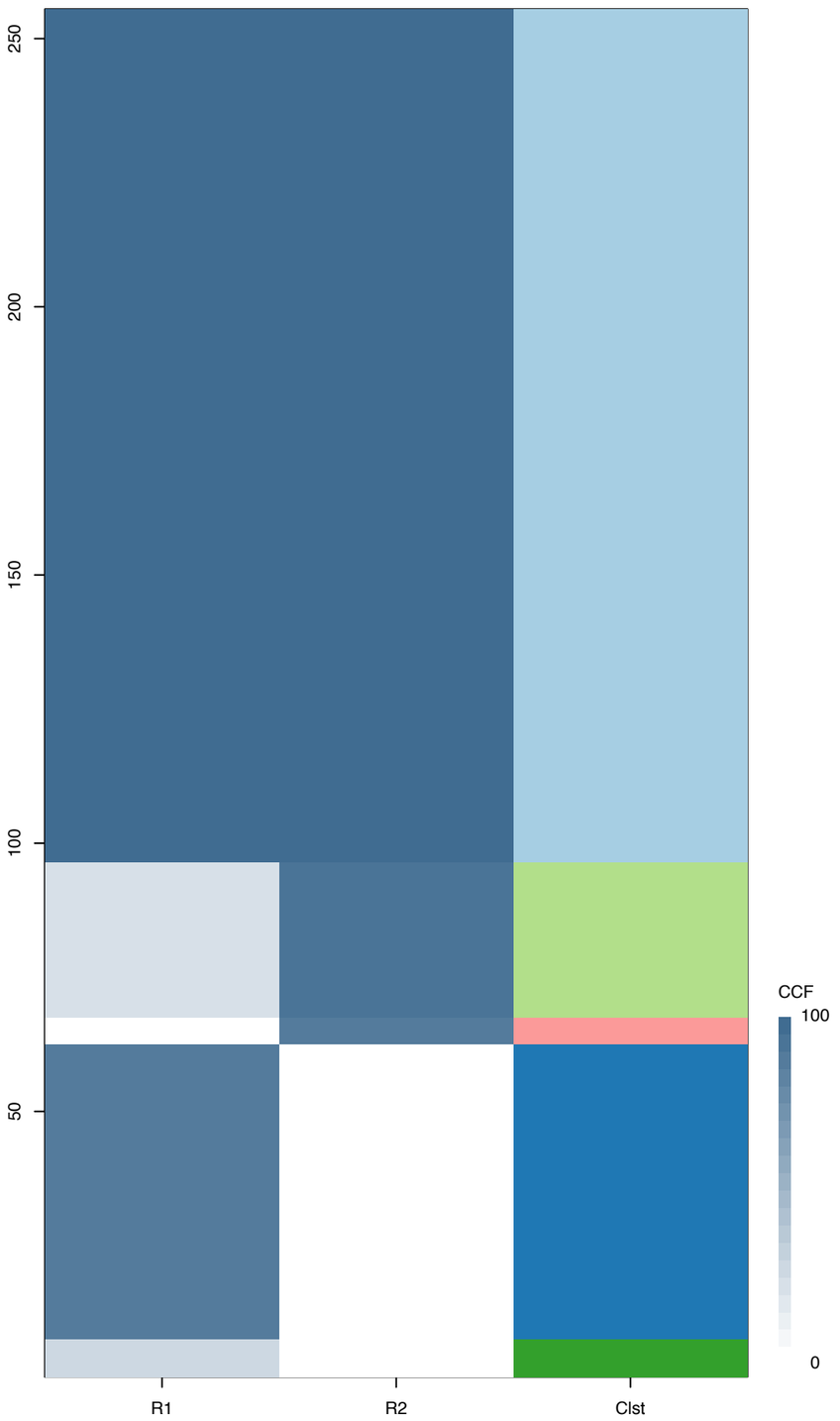
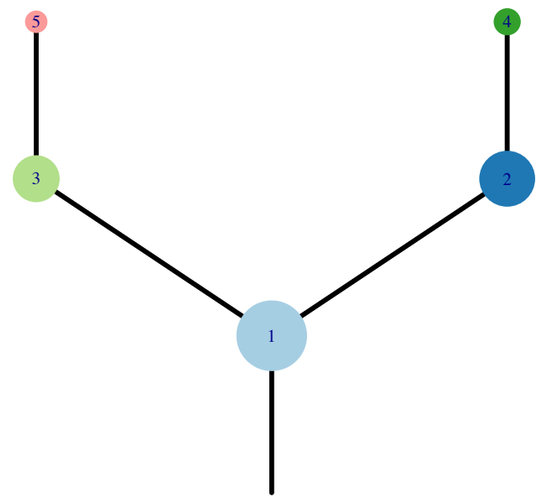


Fig.S12V



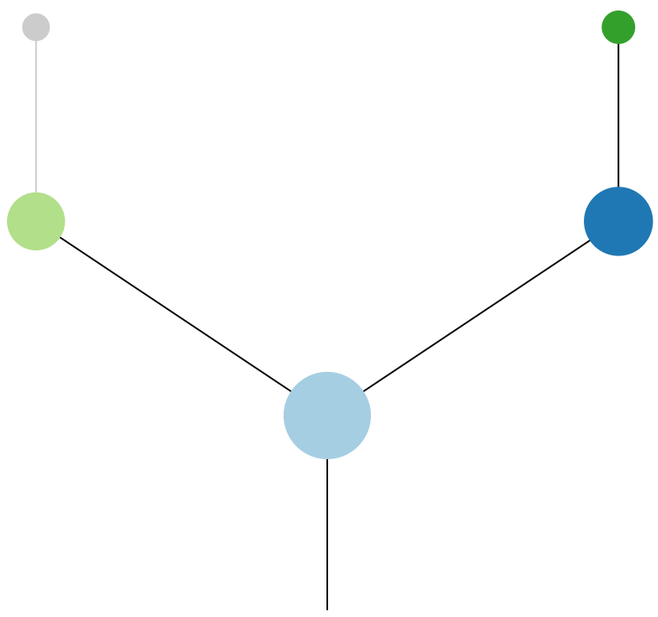
CRUK0022_A



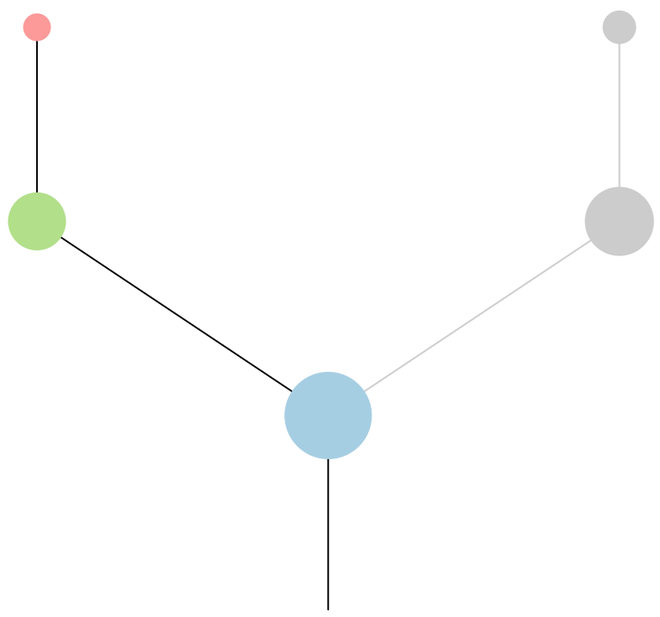
Histology:Adeno, Age:66, PackYears:32.25, Size:27
Stage:3a, Gender:Female, GD:Clonal GD, Recur:no

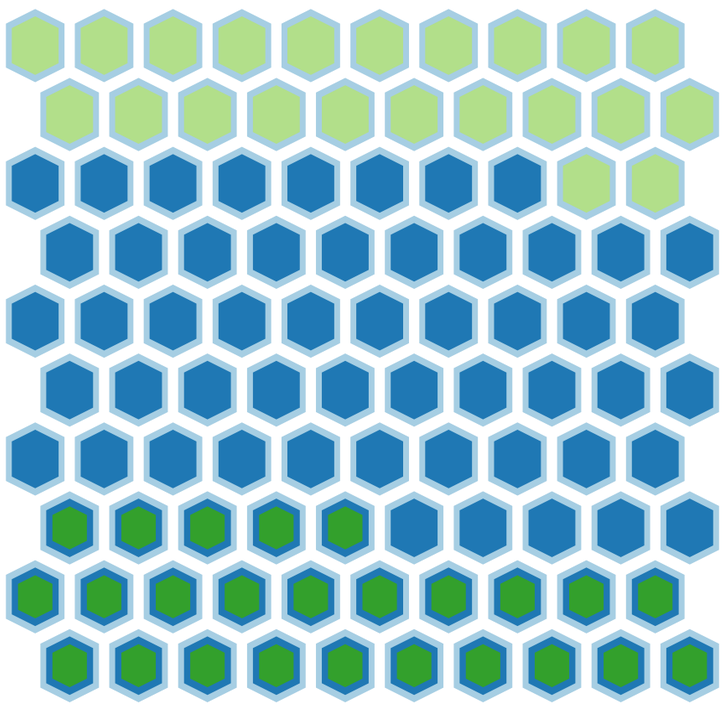
Gene	Cluster	Cytoband	Type
ETV1	1	7p21.2	Amp
HNRNPA2B1	1	7p15.2	Amp
HOXA9	1	7p15.2	Amp
HOXA11	1	7p15.2	Amp
HOXA13	1	7p15.2	Amp
JAZF1	1	7p15.2	Amp
EGFR	1	7p11.2	SNV
AKAP9	1	7q21.2	Amp
CDK6	1	7q21.2	Amp
TRRAP	1	7q22.1	Amp
EZH2	1	7q36.1	Amp
MNX1	1	7q36.3	Amp
RUNX1T1	1	8q21.3	Amp
COX6C	1	8q22.2	Amp
RSPO2	1	8q23.1	Amp
EIF3E	1	8q23.1	Amp
NKX2−1	1	14q13.3	Amp
FOXA1	1	14q21.1	Amp
TP53	1	17p13.1	SNV
CCNE1	1	19q12	Amp
AKT2	1	19q13.2	Amp
CIC	2	19q13.2	SNV

R1



R2





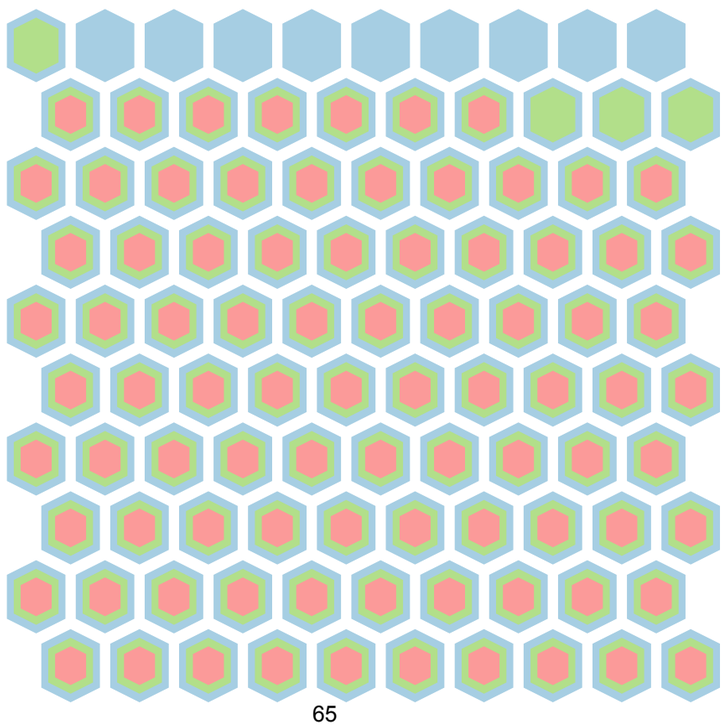
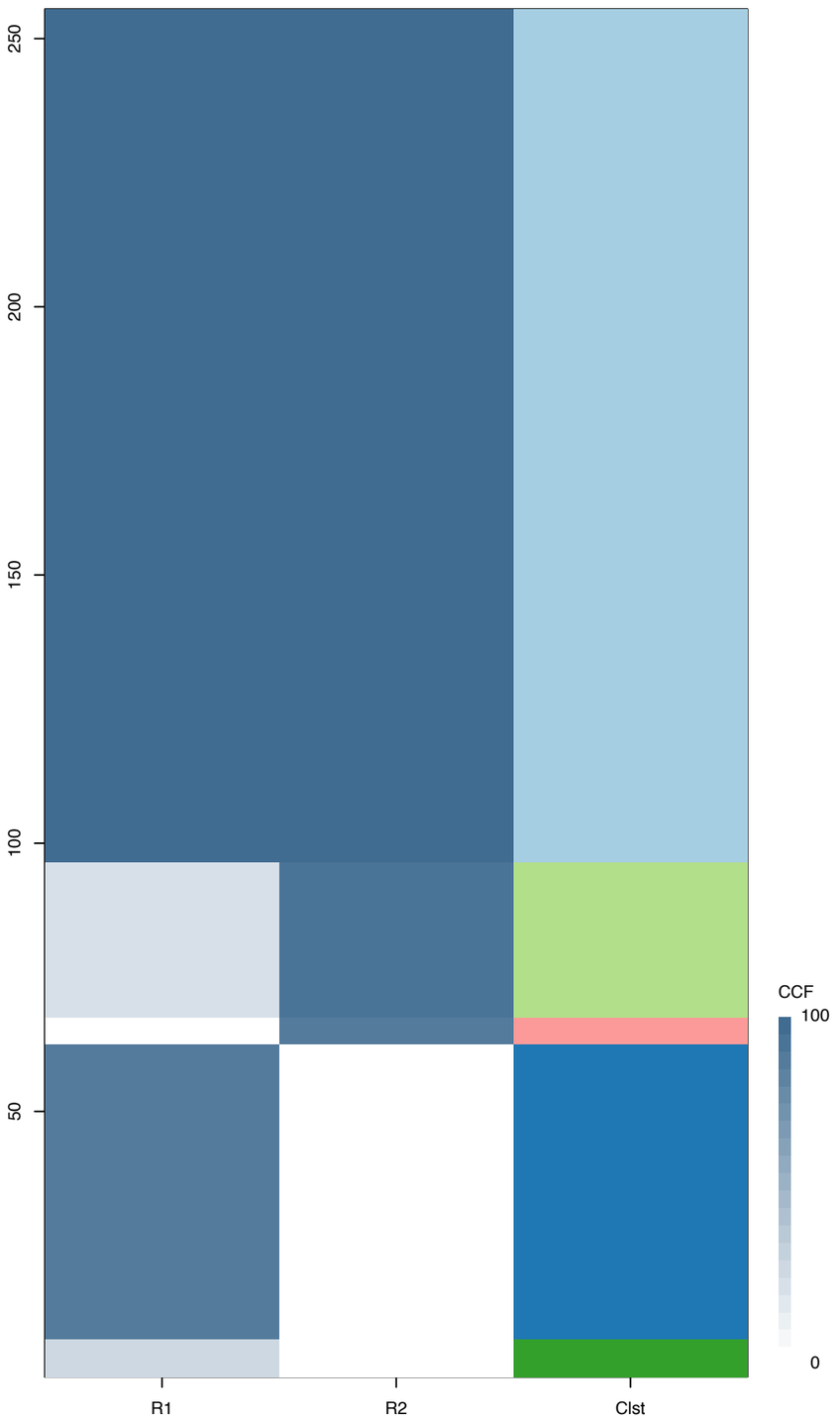
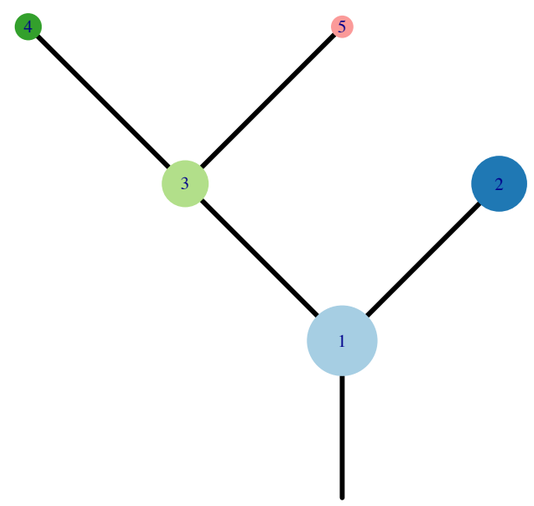


Fig.S12V



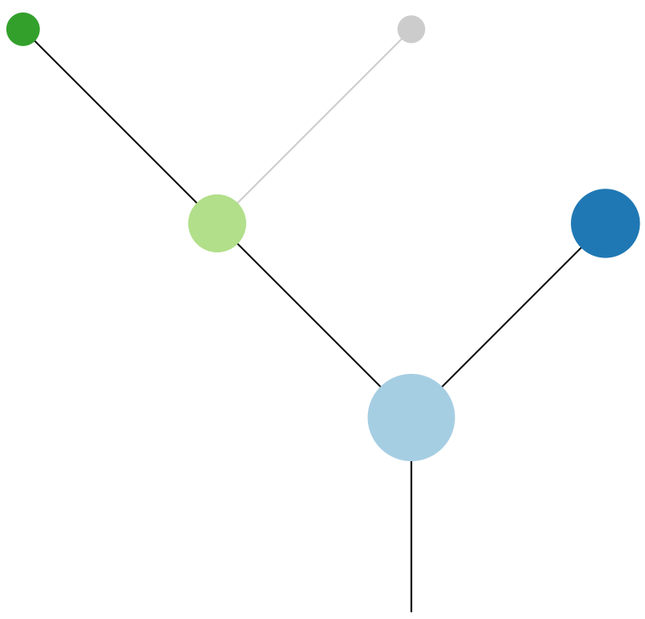
CRUK0022_B



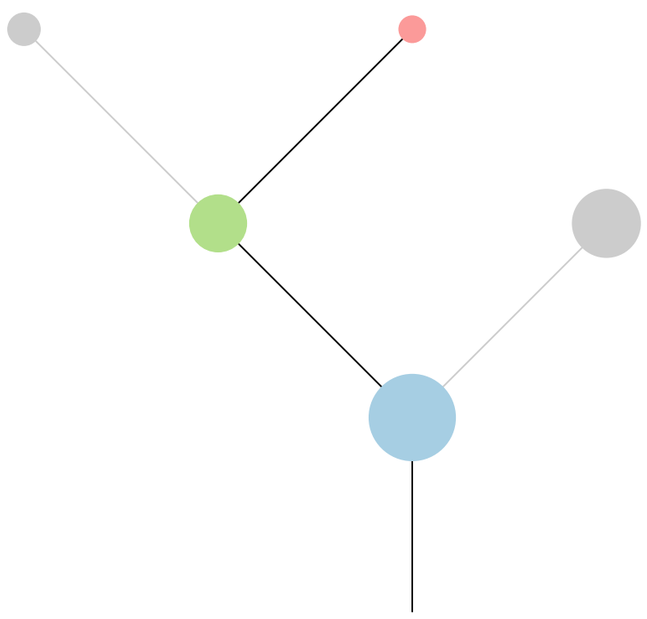
Histology:Adeno, Age:66, PackYears:32.25, Size:27
Stage:3a, Gender:Female, GD:Clonal GD, Recur:no

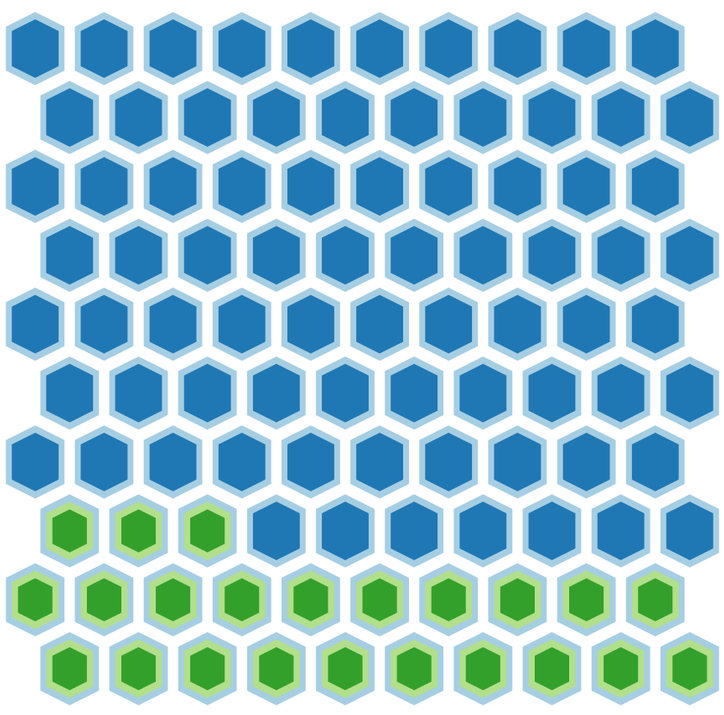
Gene	Cluster	Cytoband	Type
ETV1	1	7p21.2	Amp
HNRNPA2B1	1	7p15.2	Amp
HOXA9	1	7p15.2	Amp
HOXA11	1	7p15.2	Amp
HOXA13	1	7p15.2	Amp
JAZF1	1	7p15.2	Amp
EGFR	1	7p11.2	SNV
AKAP9	1	7q21.2	Amp
CDK6	1	7q21.2	Amp
TRRAP	1	7q22.1	Amp
EZH2	1	7q36.1	Amp
MNX1	1	7q36.3	Amp
RUNX1T1	1	8q21.3	Amp
COX6C	1	8q22.2	Amp
RSPO2	1	8q23.1	Amp
EIF3E	1	8q23.1	Amp
NKX2−1	1	14q13.3	Amp
FOXA1	1	14q21.1	Amp
TP53	1	17p13.1	SNV
CCNE1	1	19q12	Amp
AKT2	1	19q13.2	Amp
CIC	2	19q13.2	SNV

R1



R2





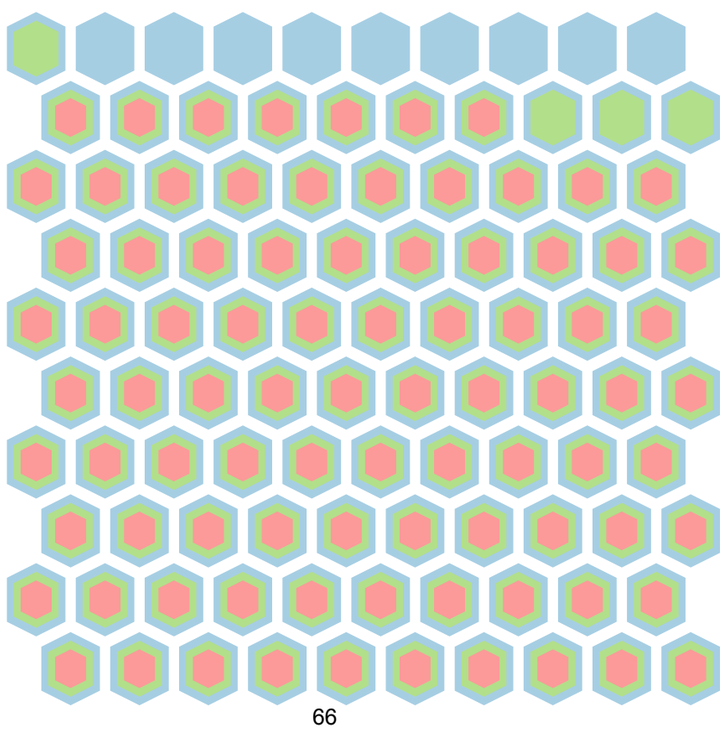
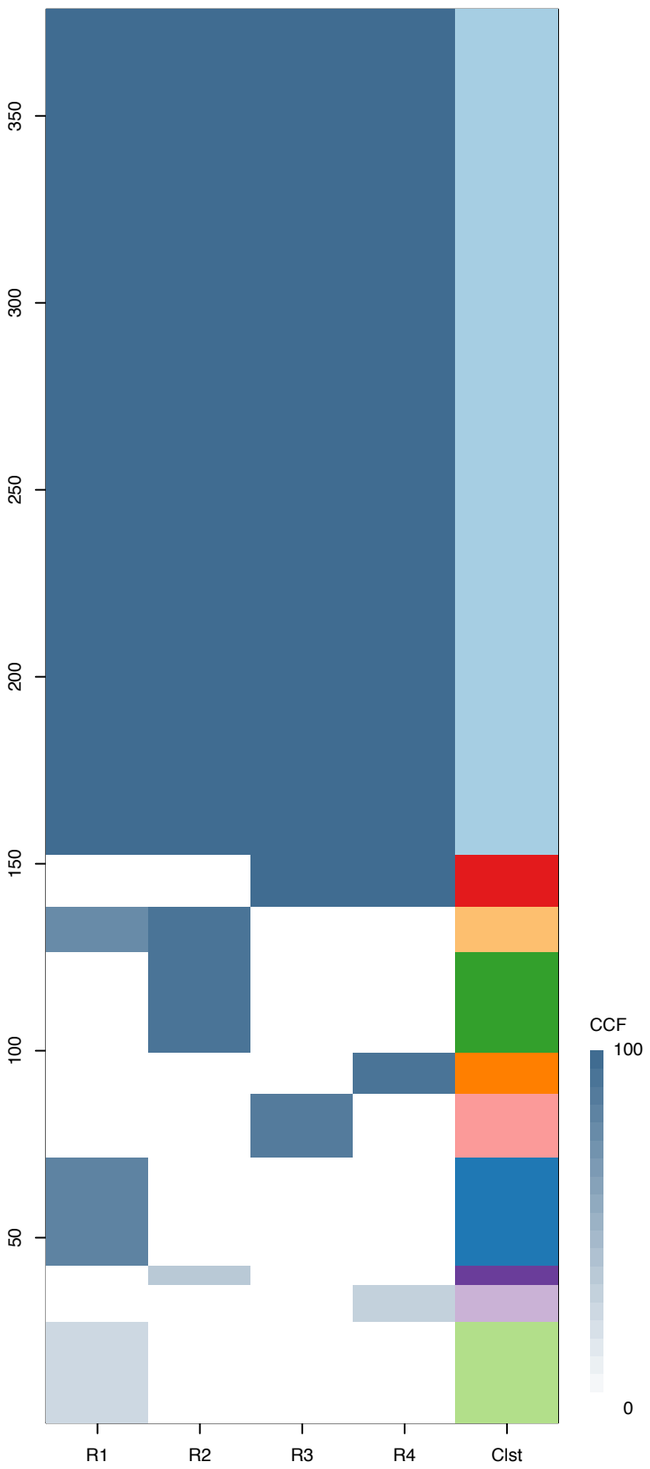
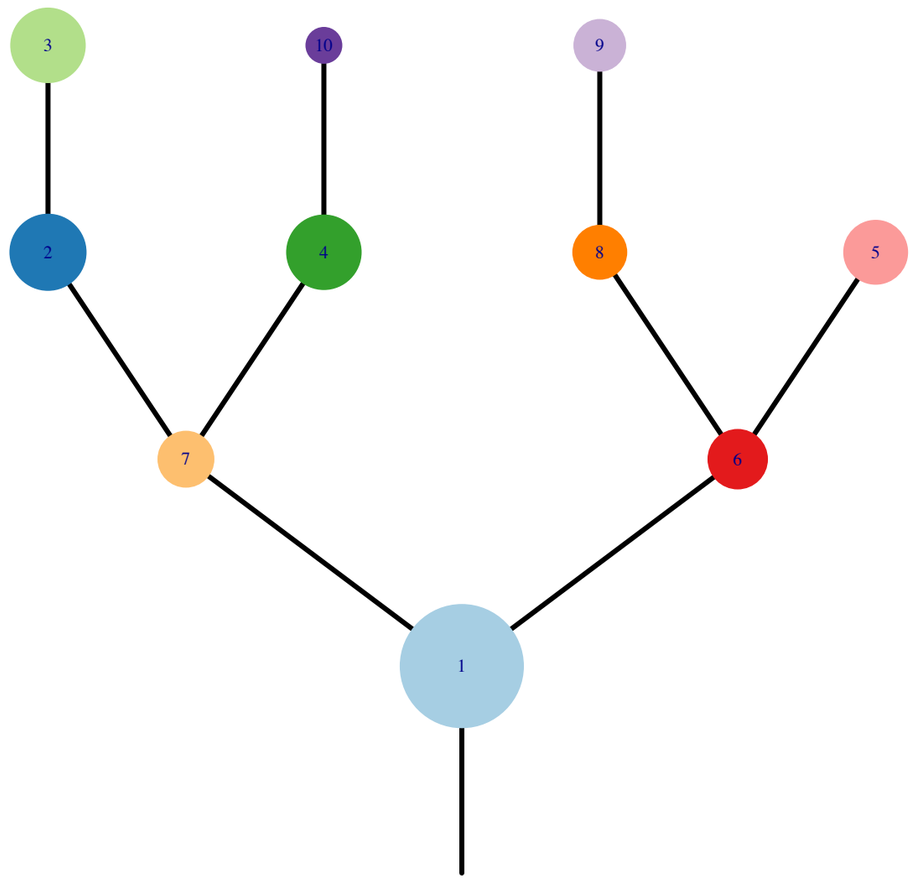


Fig.S12W



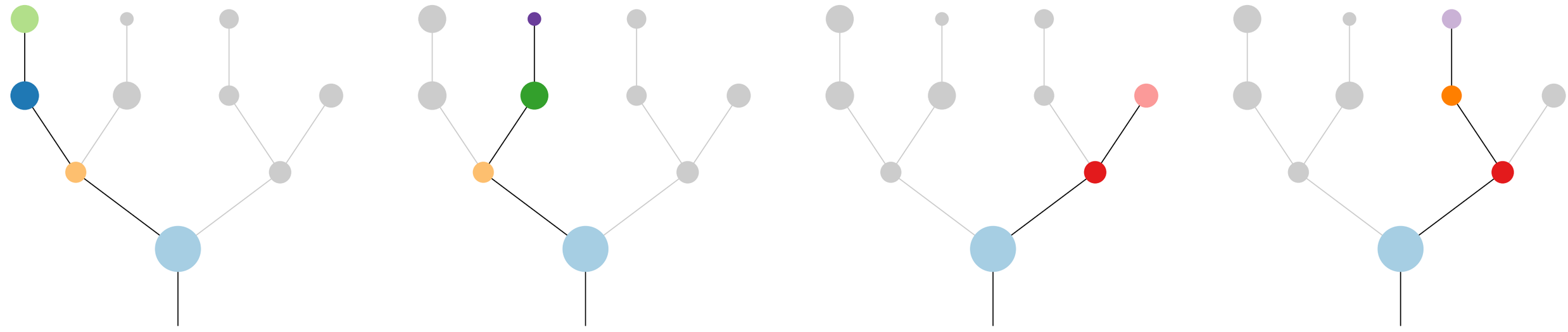
CRUK0023_A



Histology:Adeno, Age:69, PackYears:39, Size:71
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
WRN	1	8p12	SNV
EXT1	1	8q24.11	SNV
CDKN2A	1	9p21.3	Del
KRAS	1	12p12.1	SNV
TP53	2	17p13.1	SNV
PTPRC	4	1q32.1	SNV
KMT2D	4	12q13.12	SNV

R1 R2 R3 R4



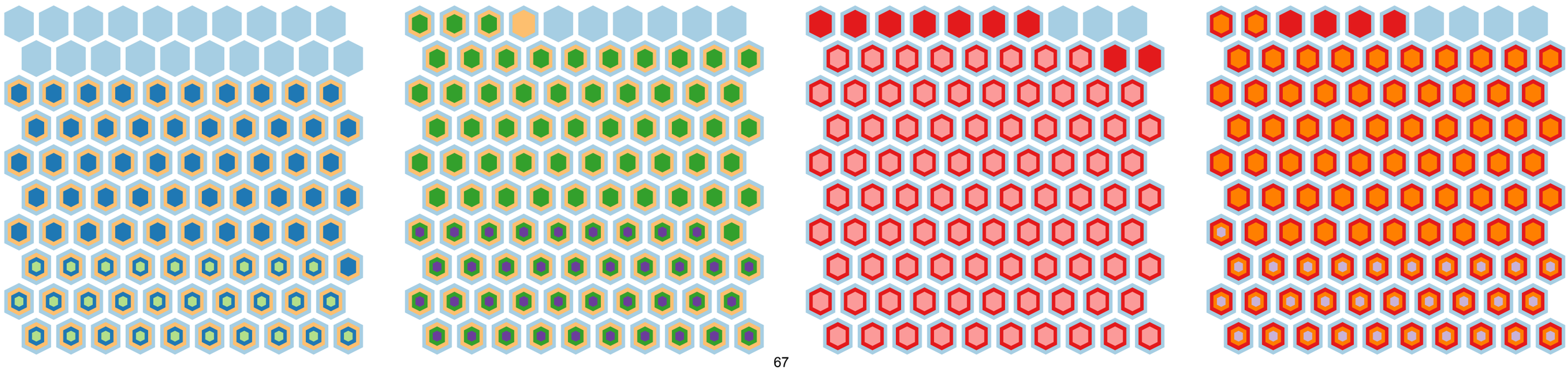
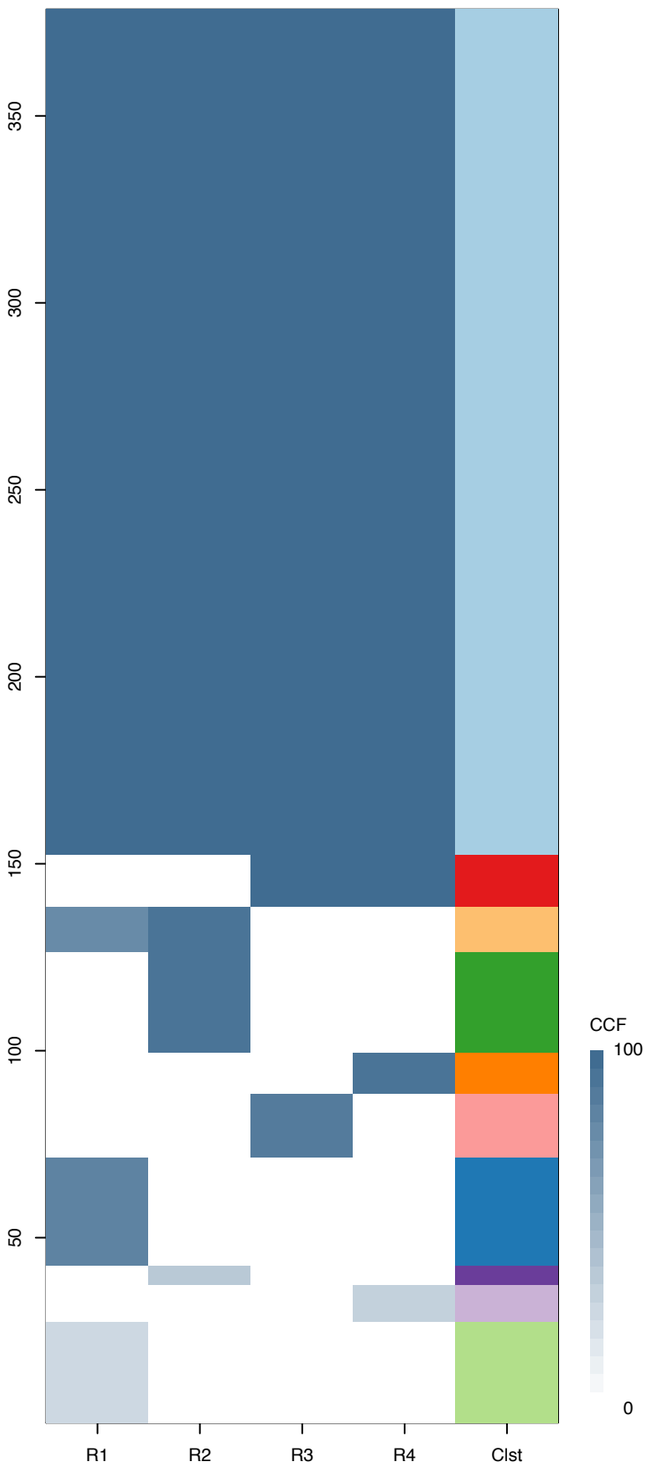
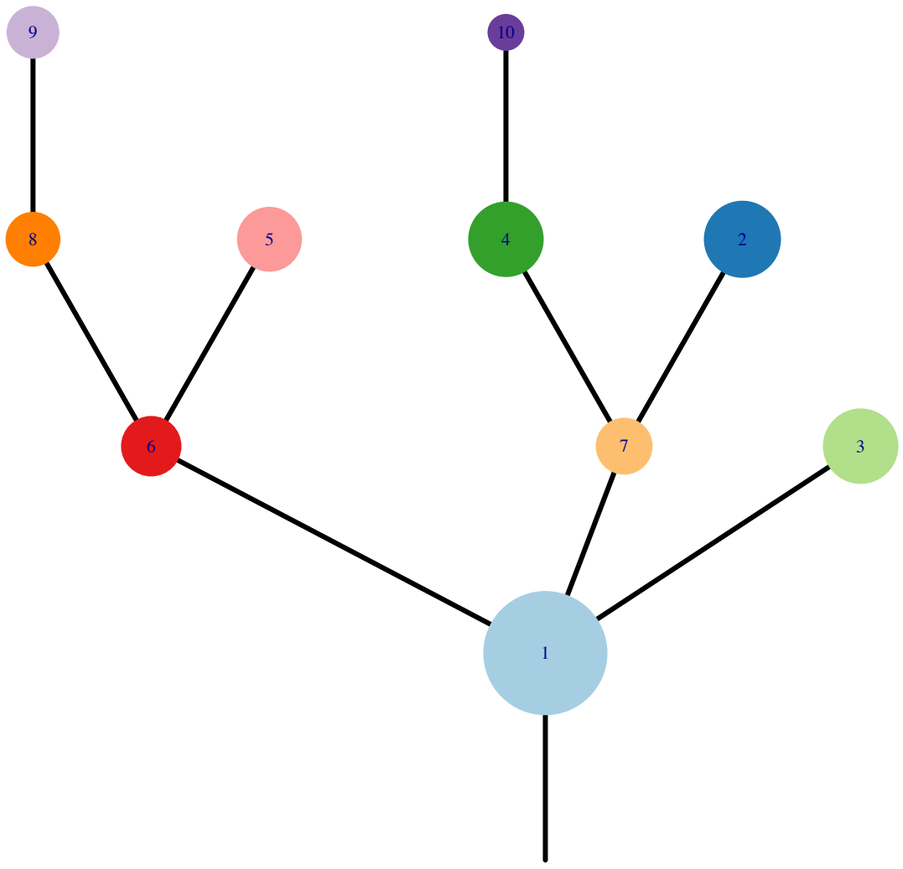


Fig.S12W



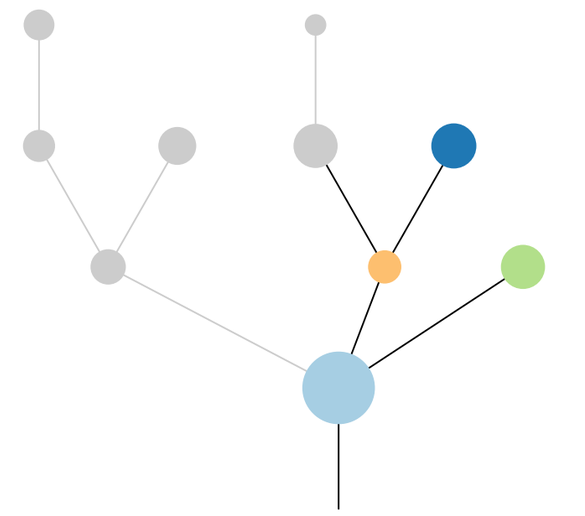
CRUK0023_B



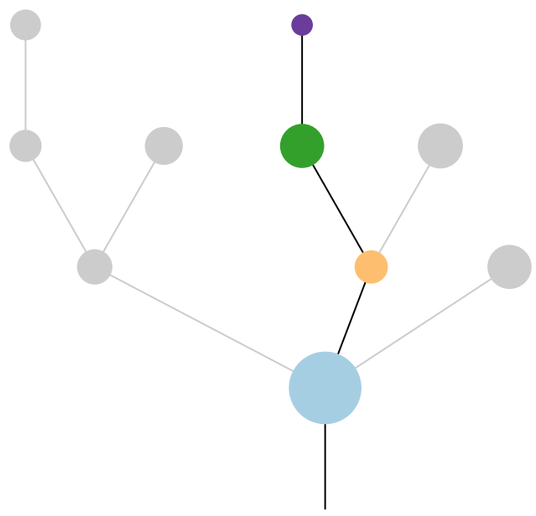
Gene	Cluster	Cytoband	Type
WRN	1	8p12	SNV
EXT1	1	8q24.11	SNV
CDKN2A	1	9p21.3	Del
KRAS	1	12p12.1	SNV
TP53	2	17p13.1	SNV
PTPRC	4	1q32.1	SNV
KMT2D	4	12q13.12	SNV

Histology:Adeno, Age:69, PackYears:39, Size:71
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

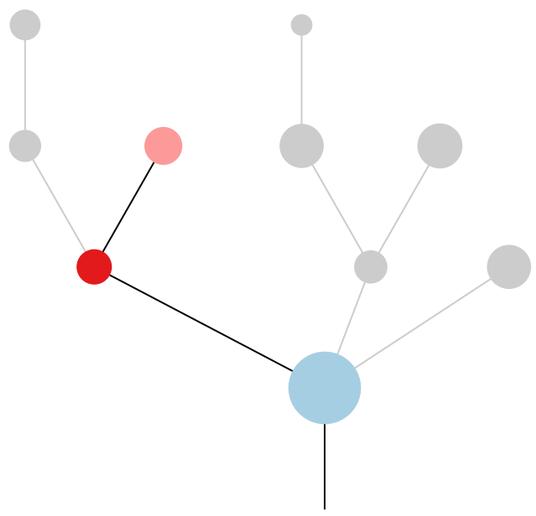
R1



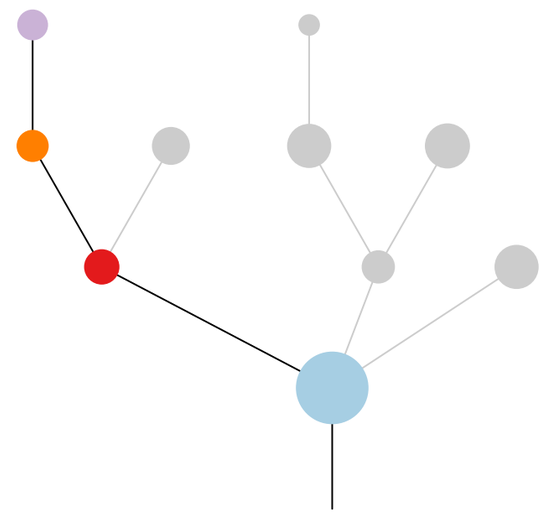
R2



R3



R4



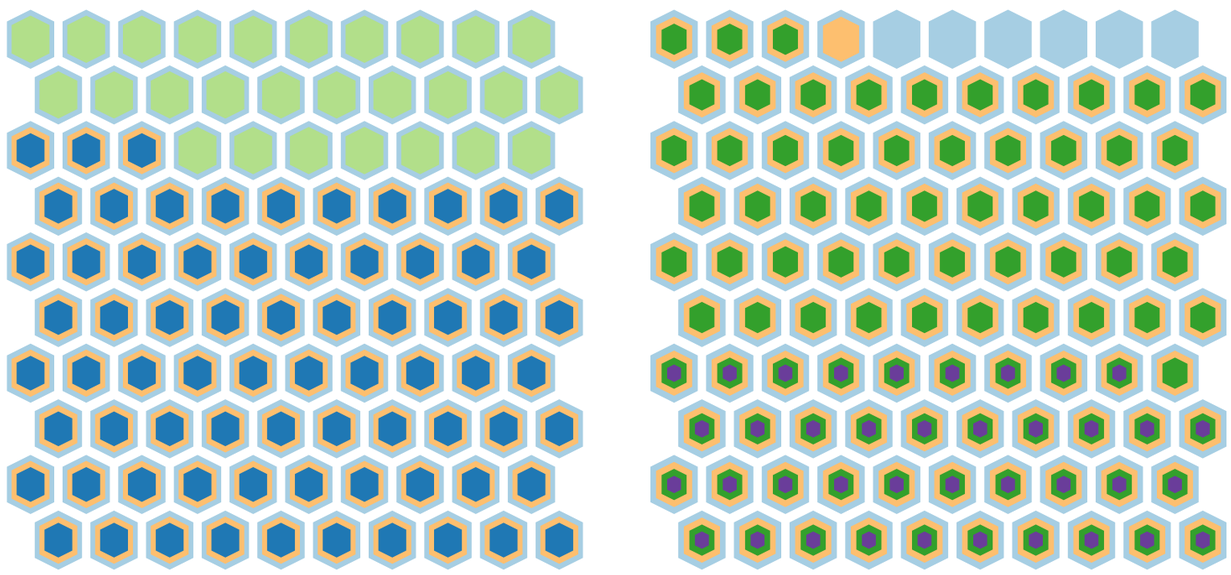
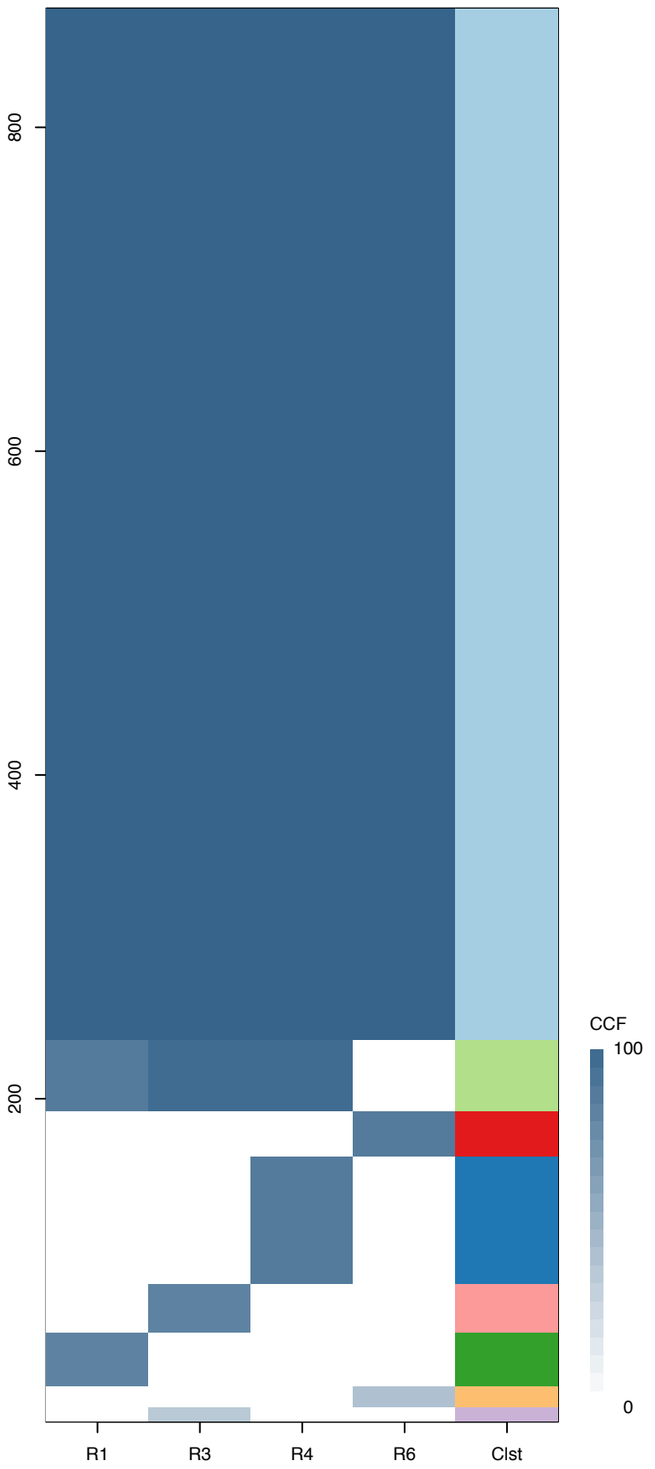
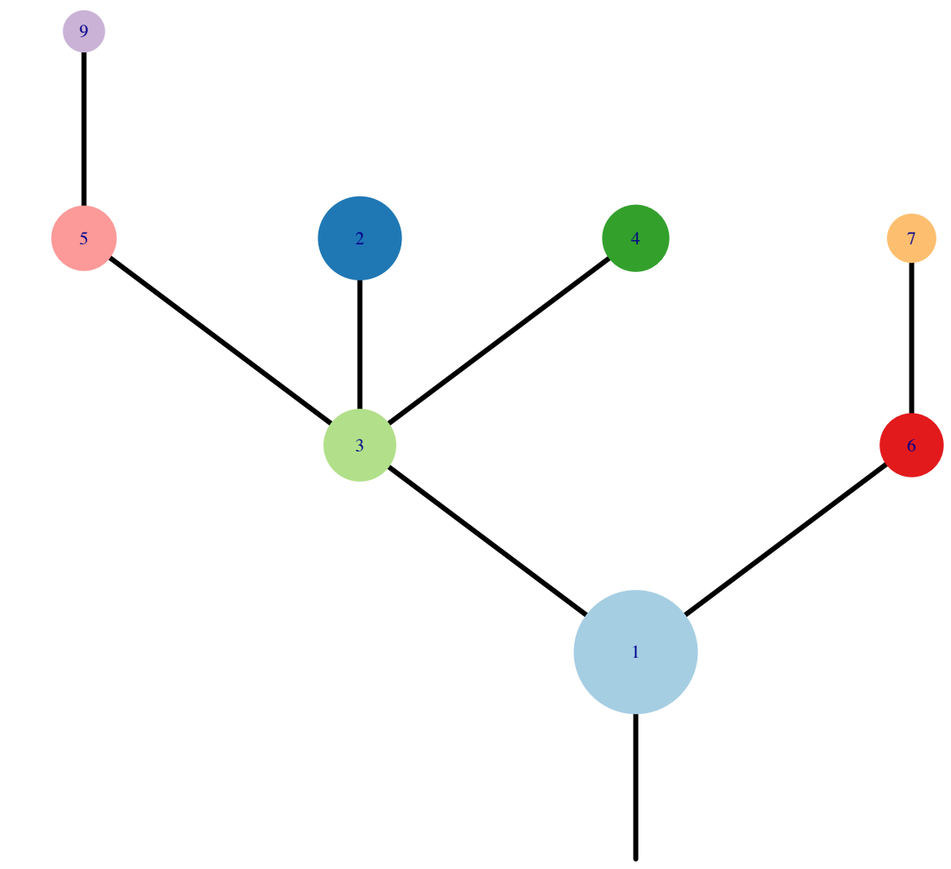


Fig.S12X

CRUK0024

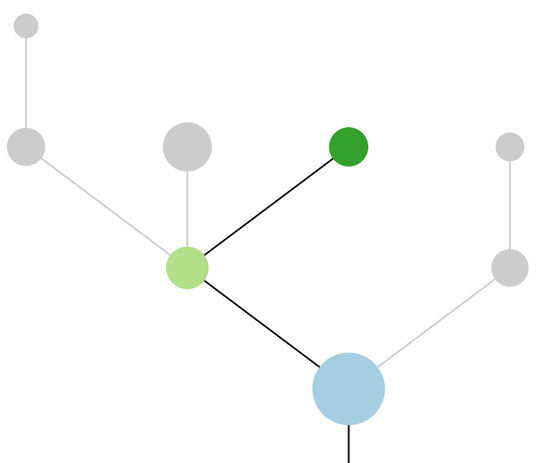




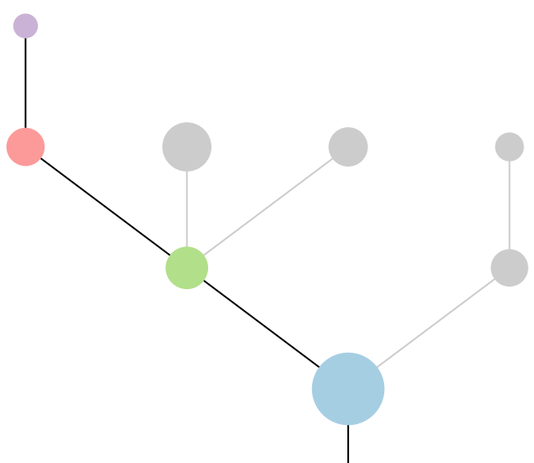
Histology:Adeno, Age:69, PackYears:65, Size:52
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
MECOM	1	3q26.2	Amp
EGFR	1	7p11.2	Amp
TCEA1	1	8q11.23	Amp
PLAG1	1	8q12.1	Amp
CHCHD7	1	8q12.1	Amp
NCOA2	1	8q13.3	Amp
HEY1	1	8q21.13	Amp
CNBD1	1	8q21.3	Amp
RUNX1T1	1	8q21.3	Amp
COX6C	1	8q22.2	Amp
POLE	1	12q24.33	SNV
TP53	1	17p13.1	SNV
STK11	1	19p13.3	SNV
ATM	3	11q22.3	SNV
NCOR1	4	17p12	SNV
STIM1	?	11p15.4	SNV

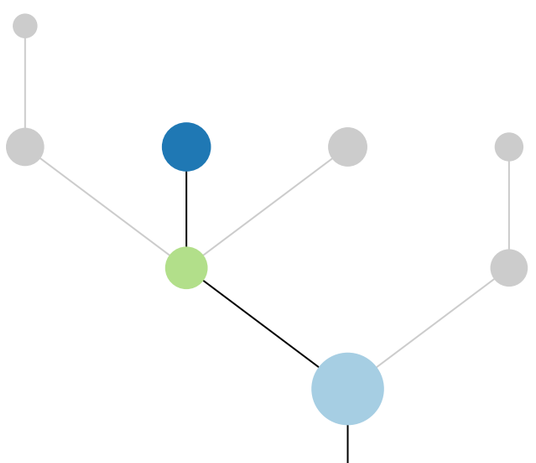
R1



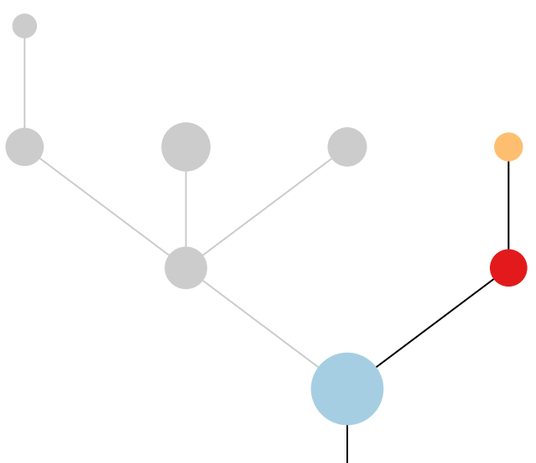
R3

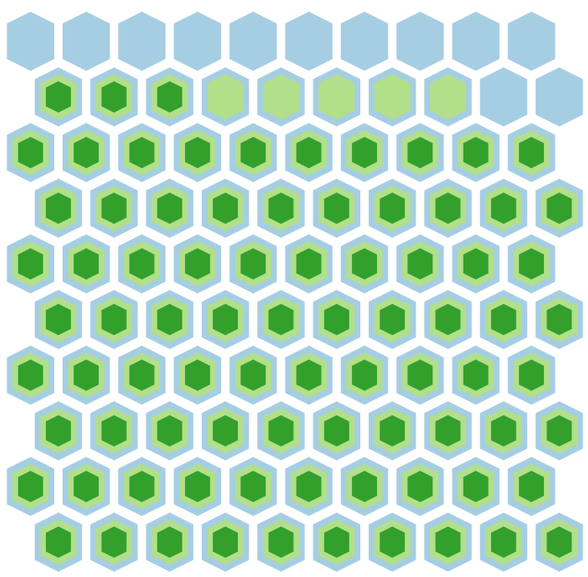


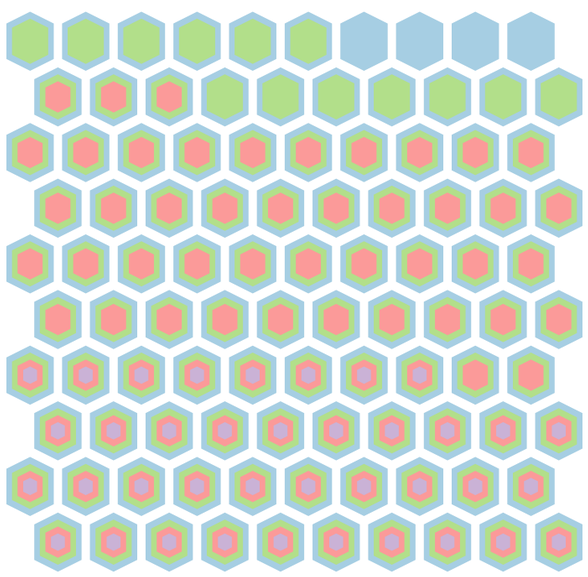
R4

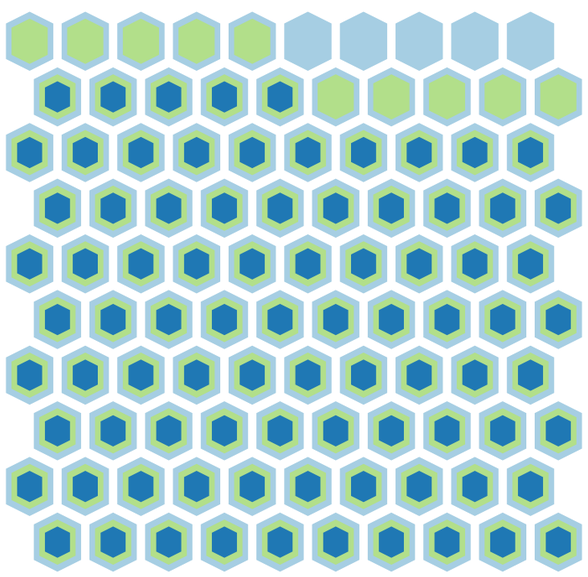


R6









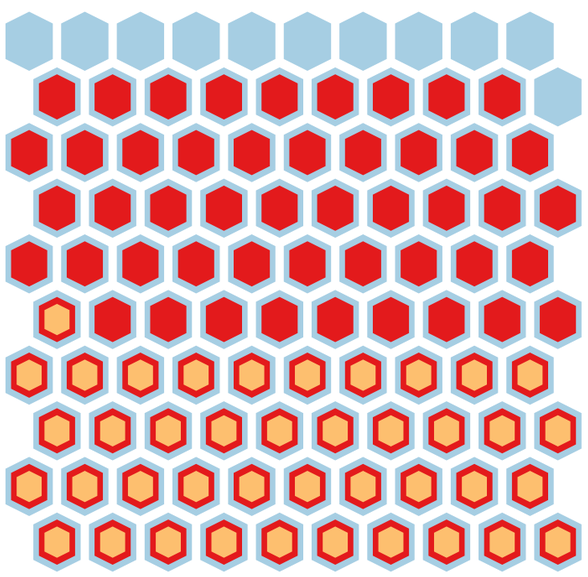
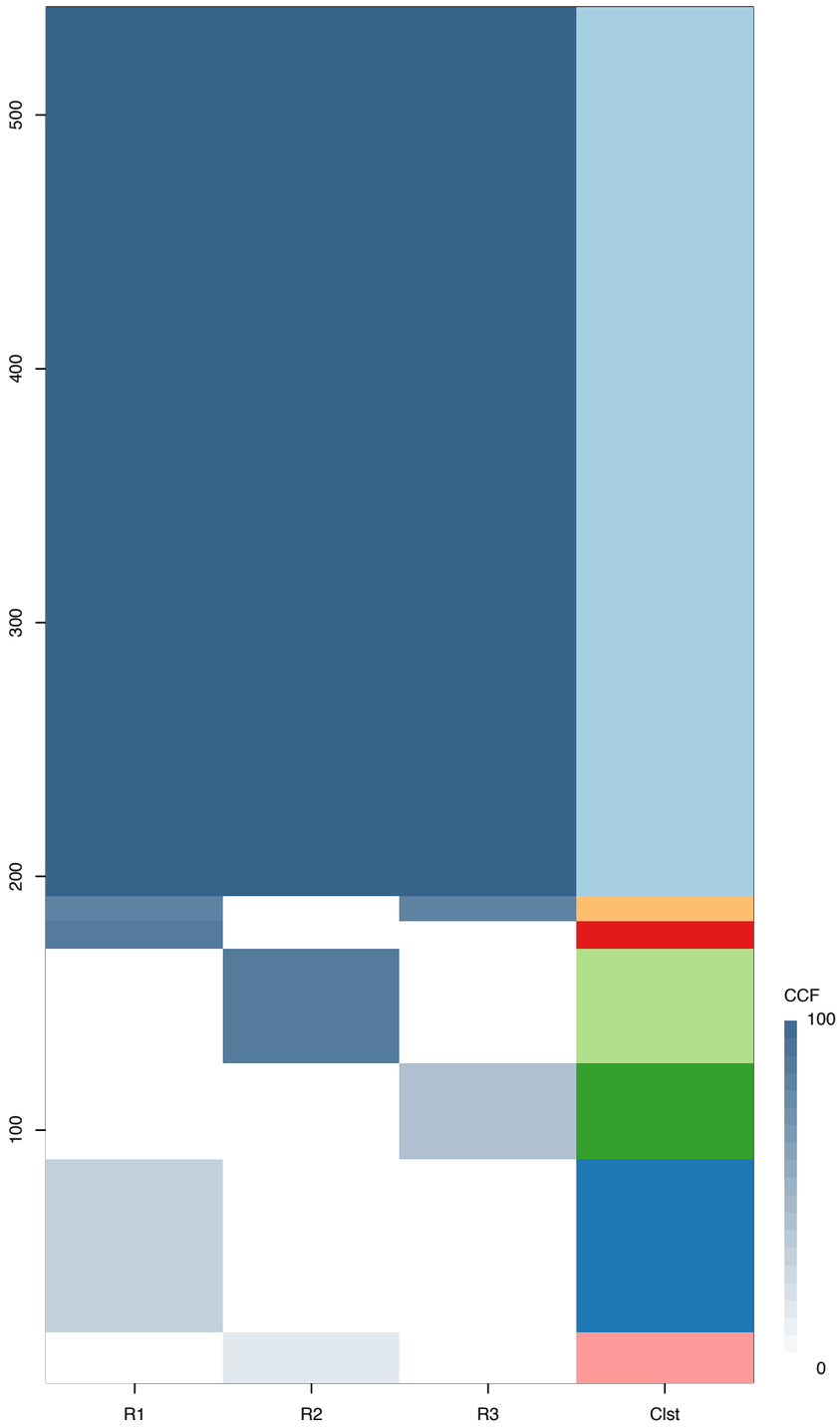
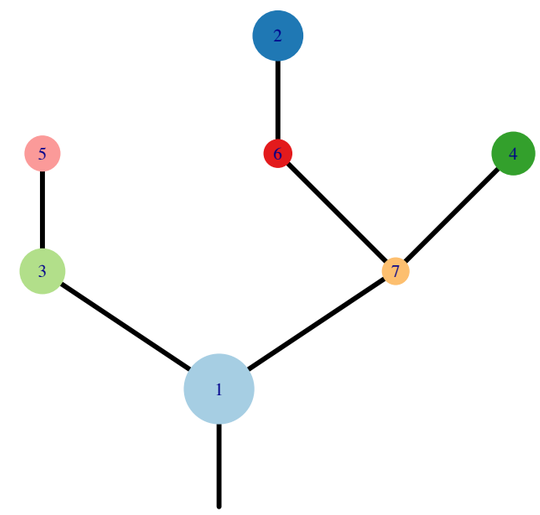


Fig.S12Y



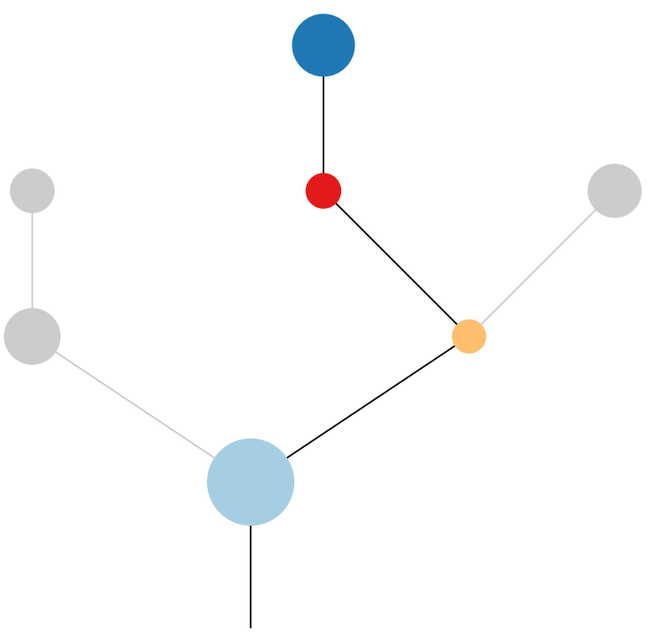
CRUK0025_A



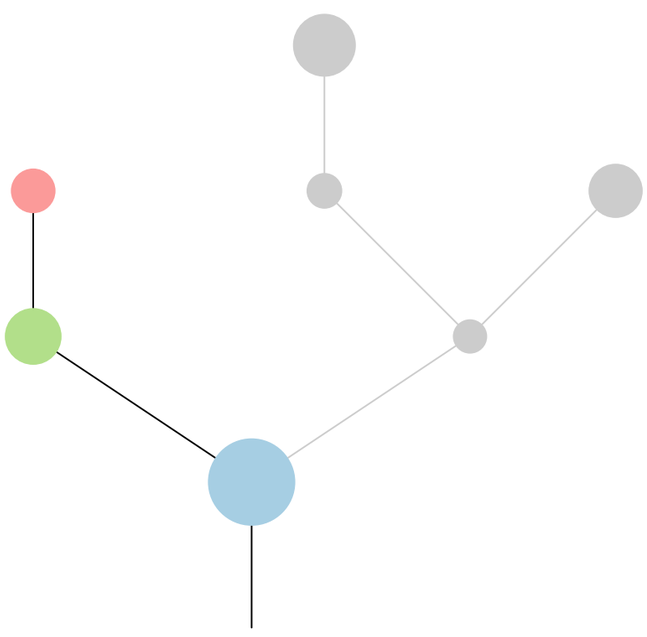
Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
MGA	1	15q15.1	SNV
TP53	1	17p13.1	SNV
RBM10	?		SNV

Histology:Adeno, Age:53, PackYears:55.5, Size:33
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

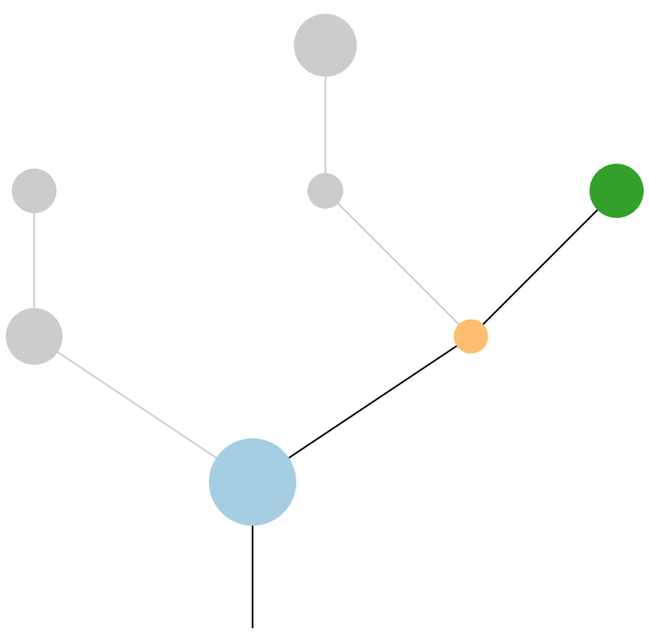
R1

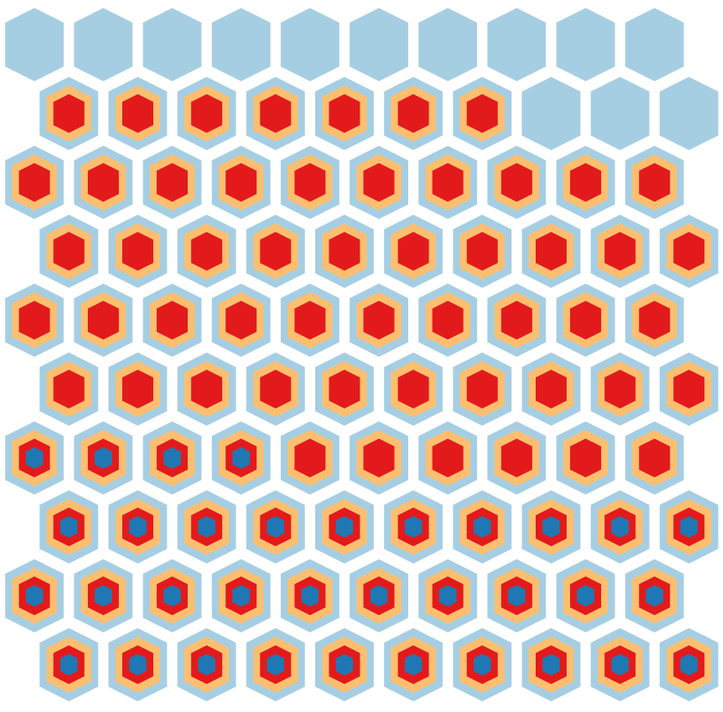


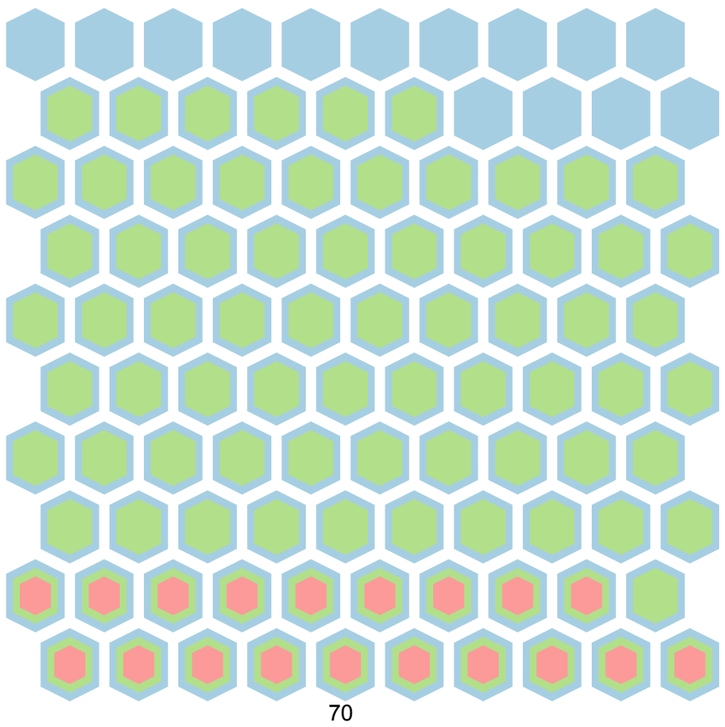
R2



R3







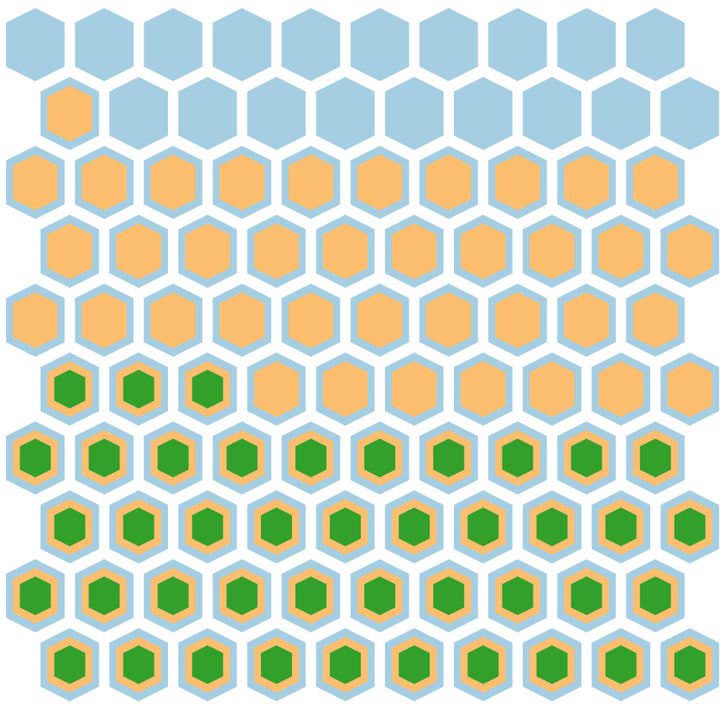
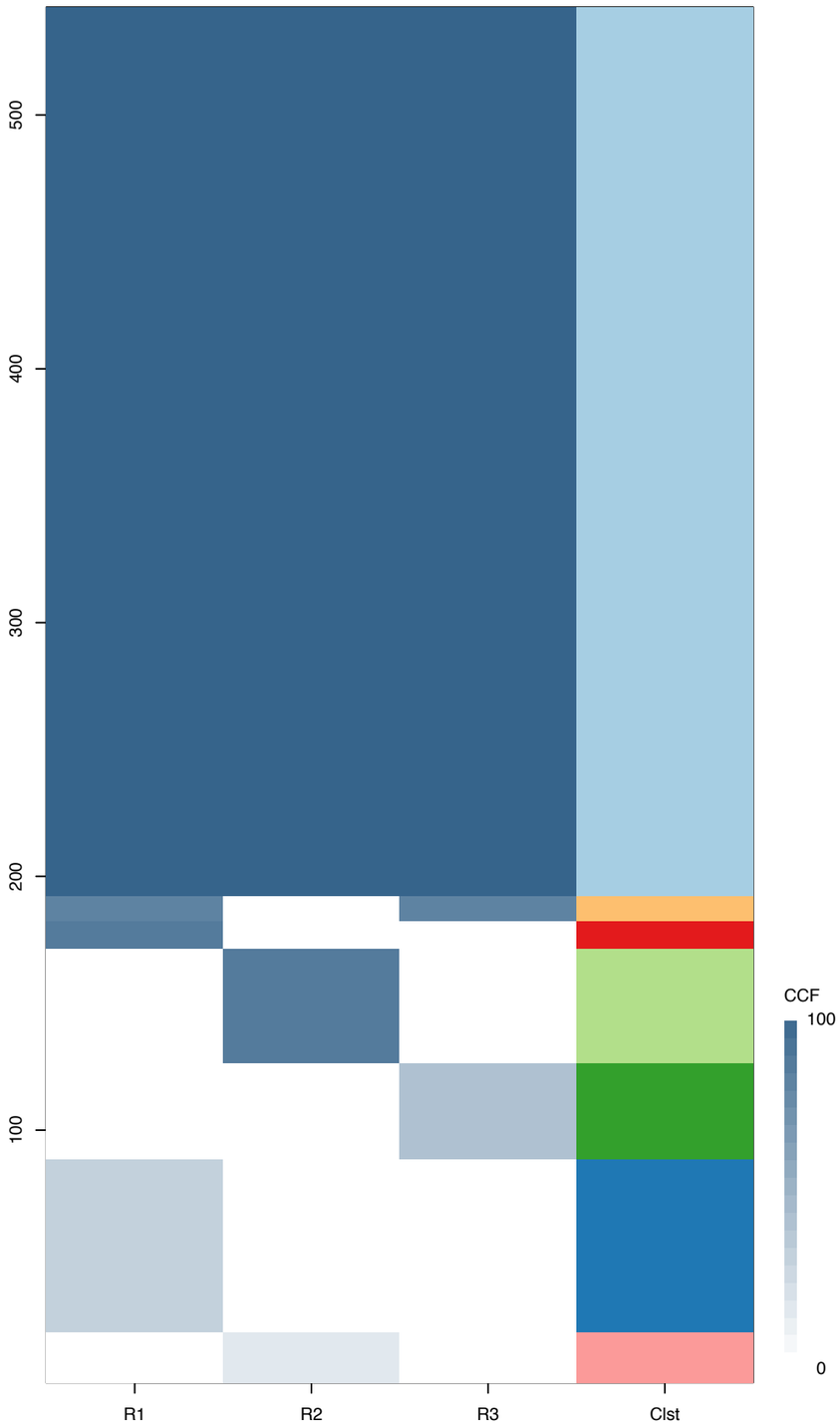
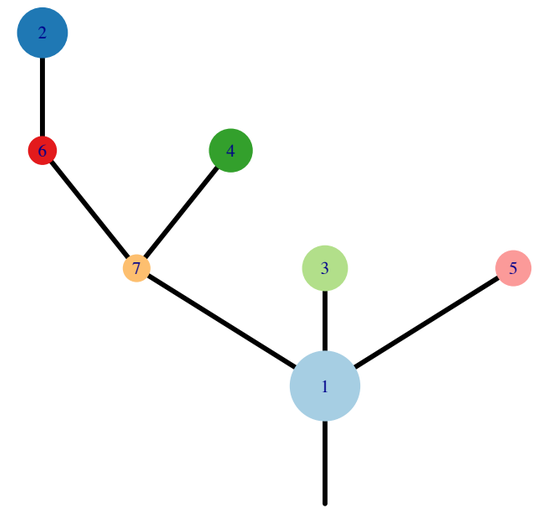


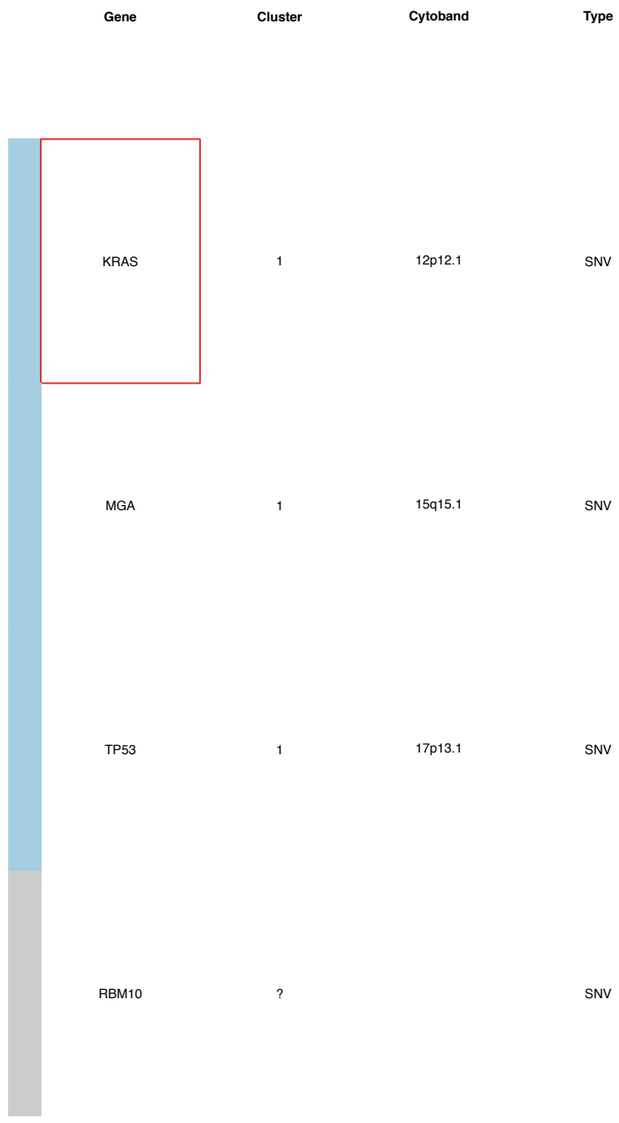
Fig.S12Y



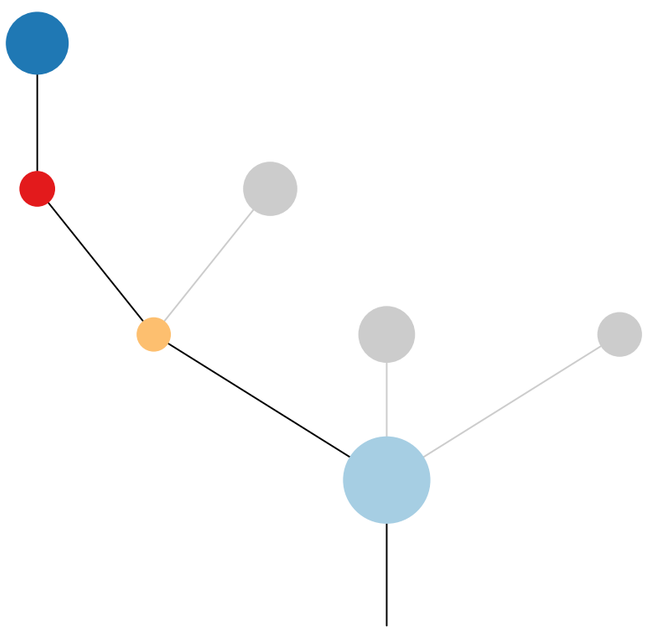
CRUK0025_B



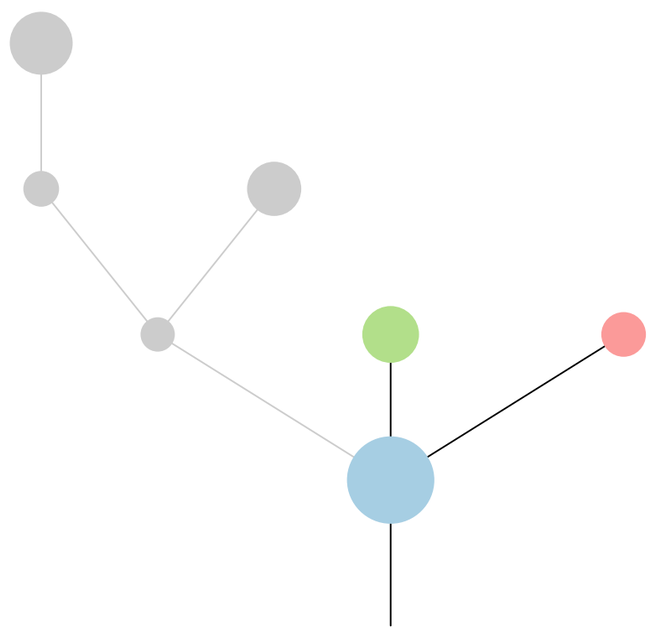
Histology:Adeno, Age:53, PackYears:55.5, Size:33
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no



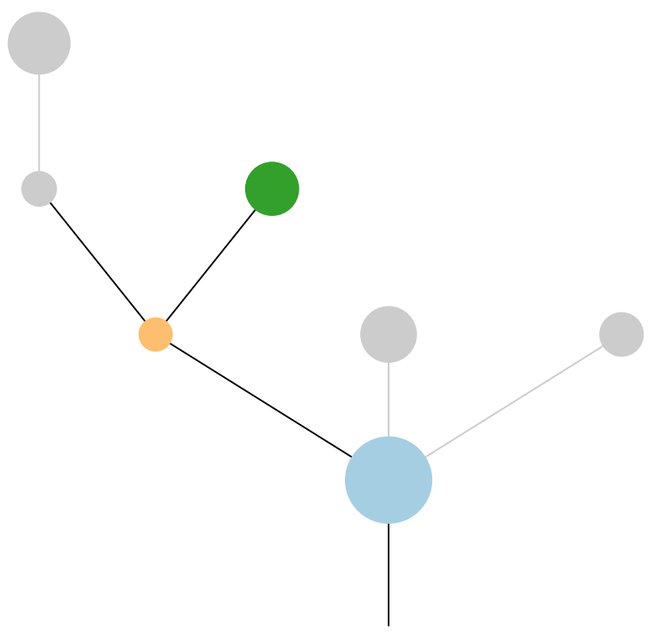
R1

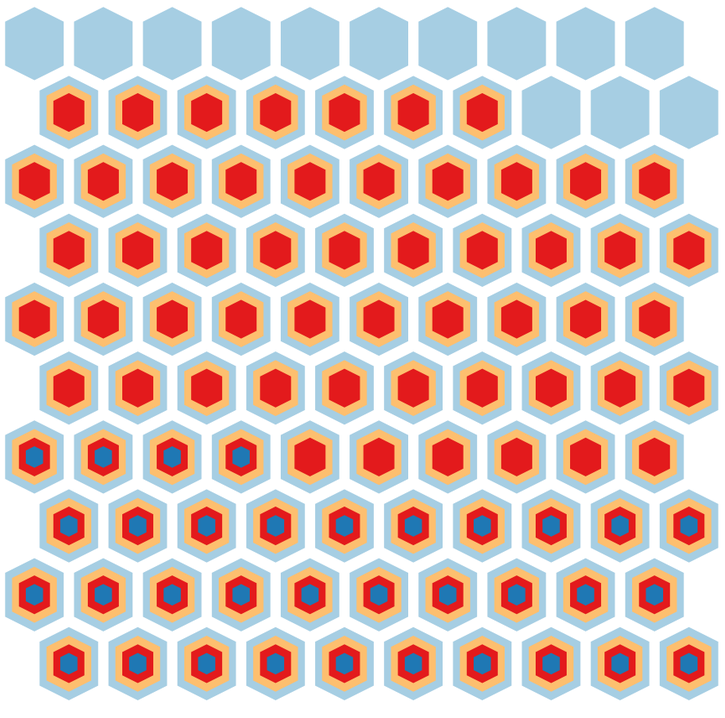


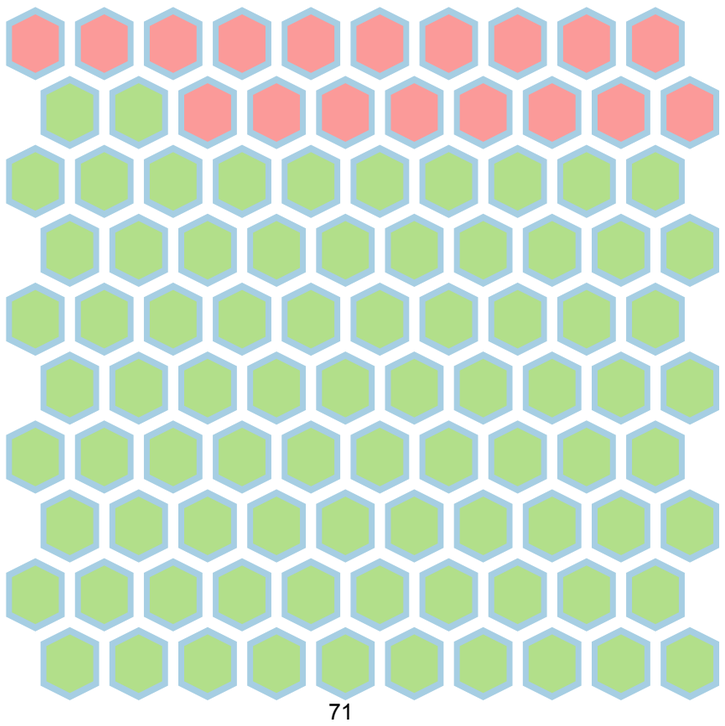
R2



R3







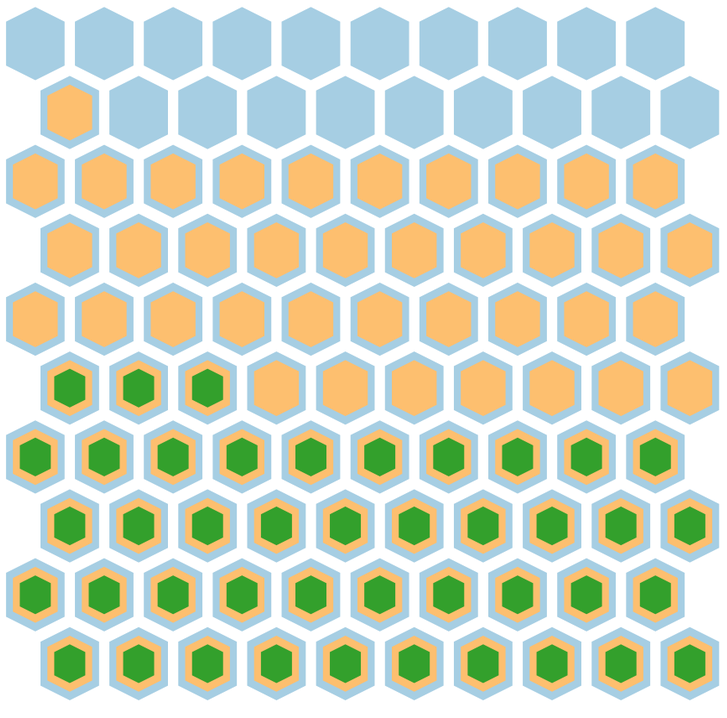
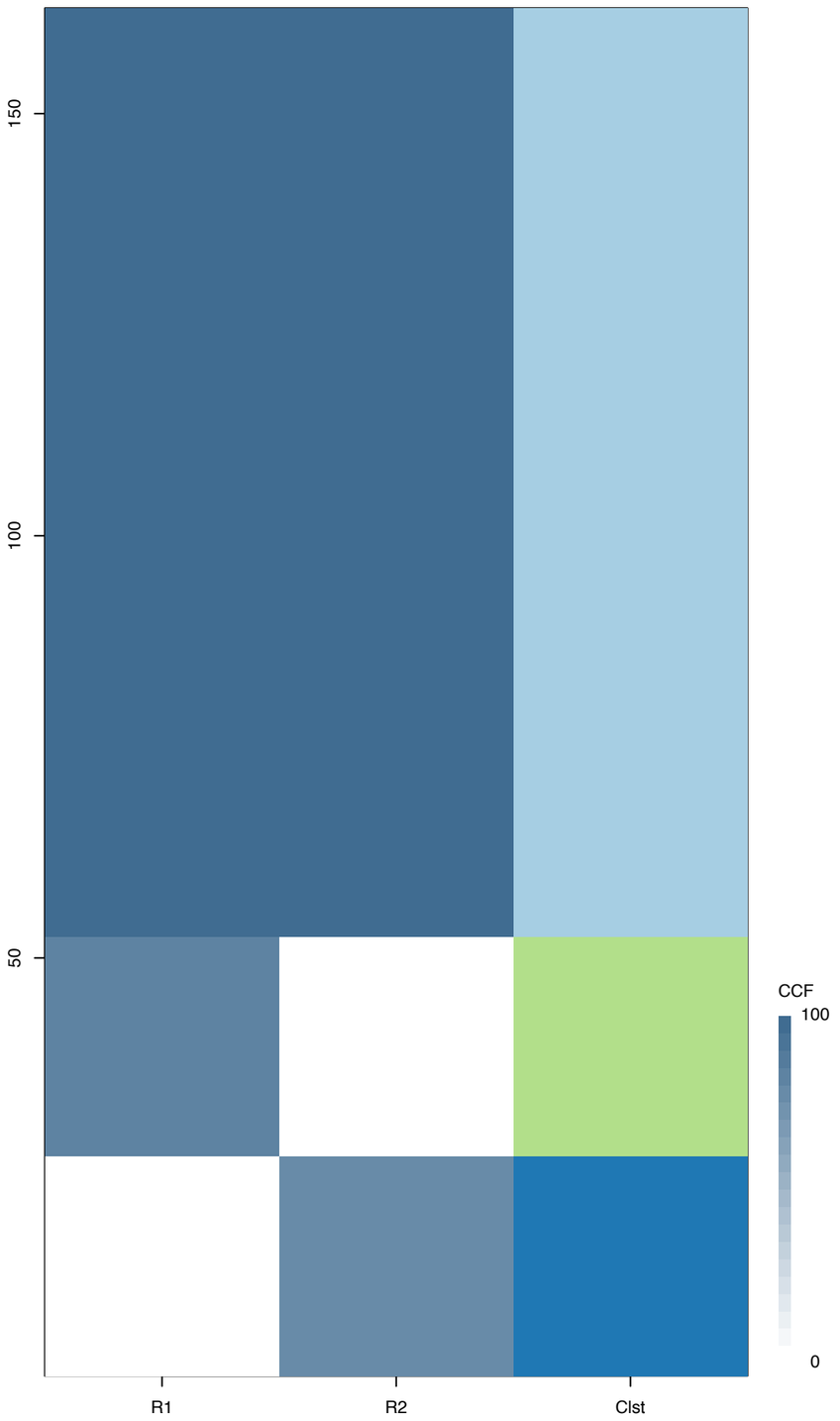
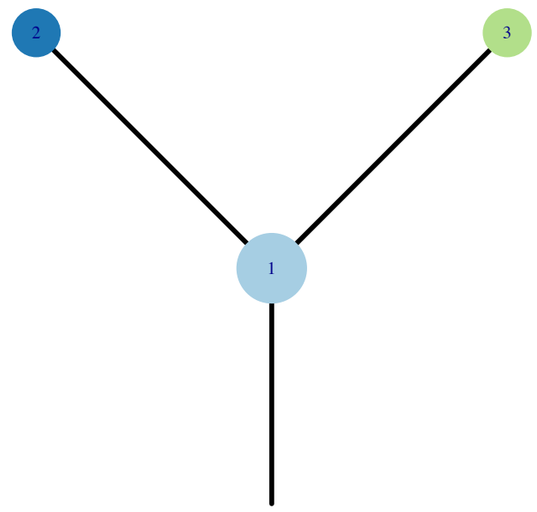


Fig.S12Z



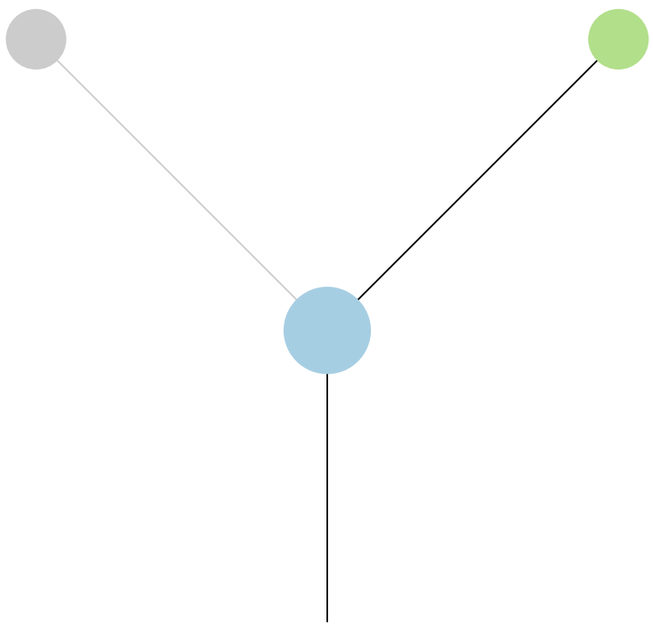
CRUK0026



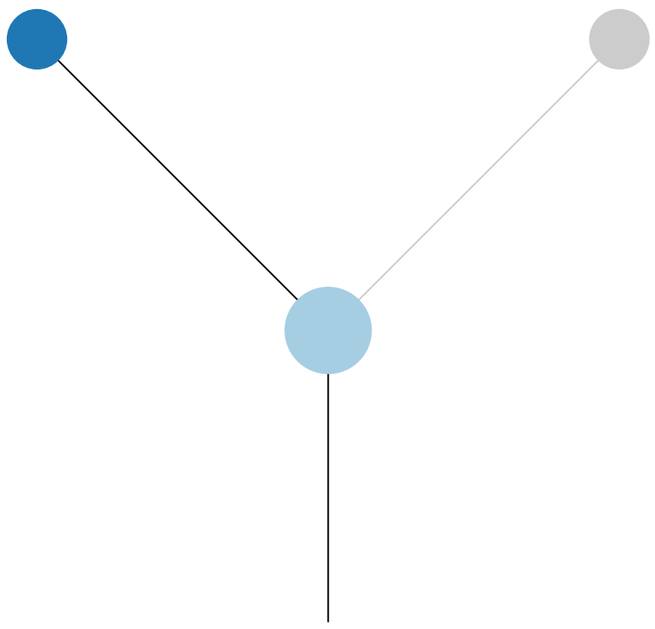
Gene	Cluster	Cytoband	Type
EGFR	1	7p11.2	SNV
RB1	1	13q14.2	SNV
TP53	1	17p13.1	SNV
SERPINB13	1	18q21.33	SNV

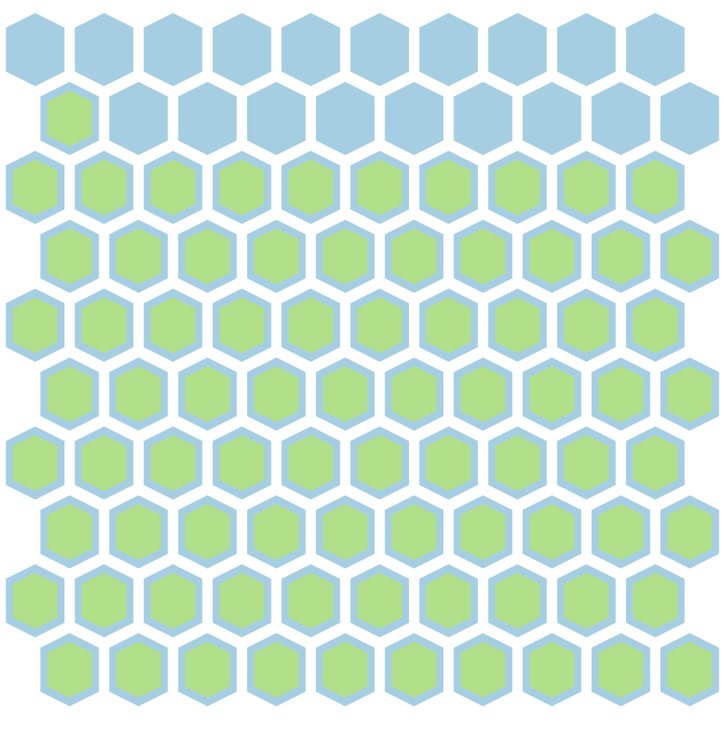
Histology:Adeno, Age:69, PackYears:0.05, Size:26
Stage:1a, Gender:Female, GD:Clonal GD, Recur:no

R1



R2





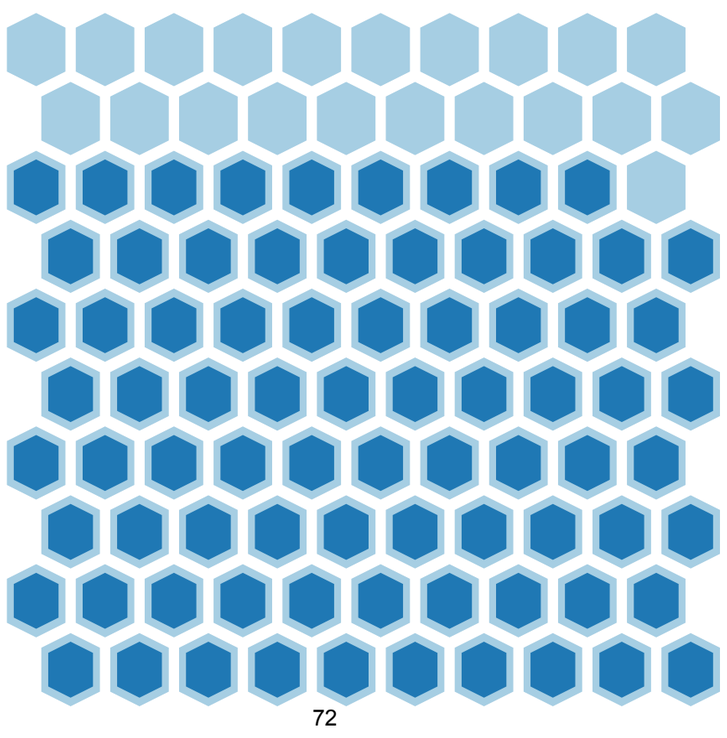
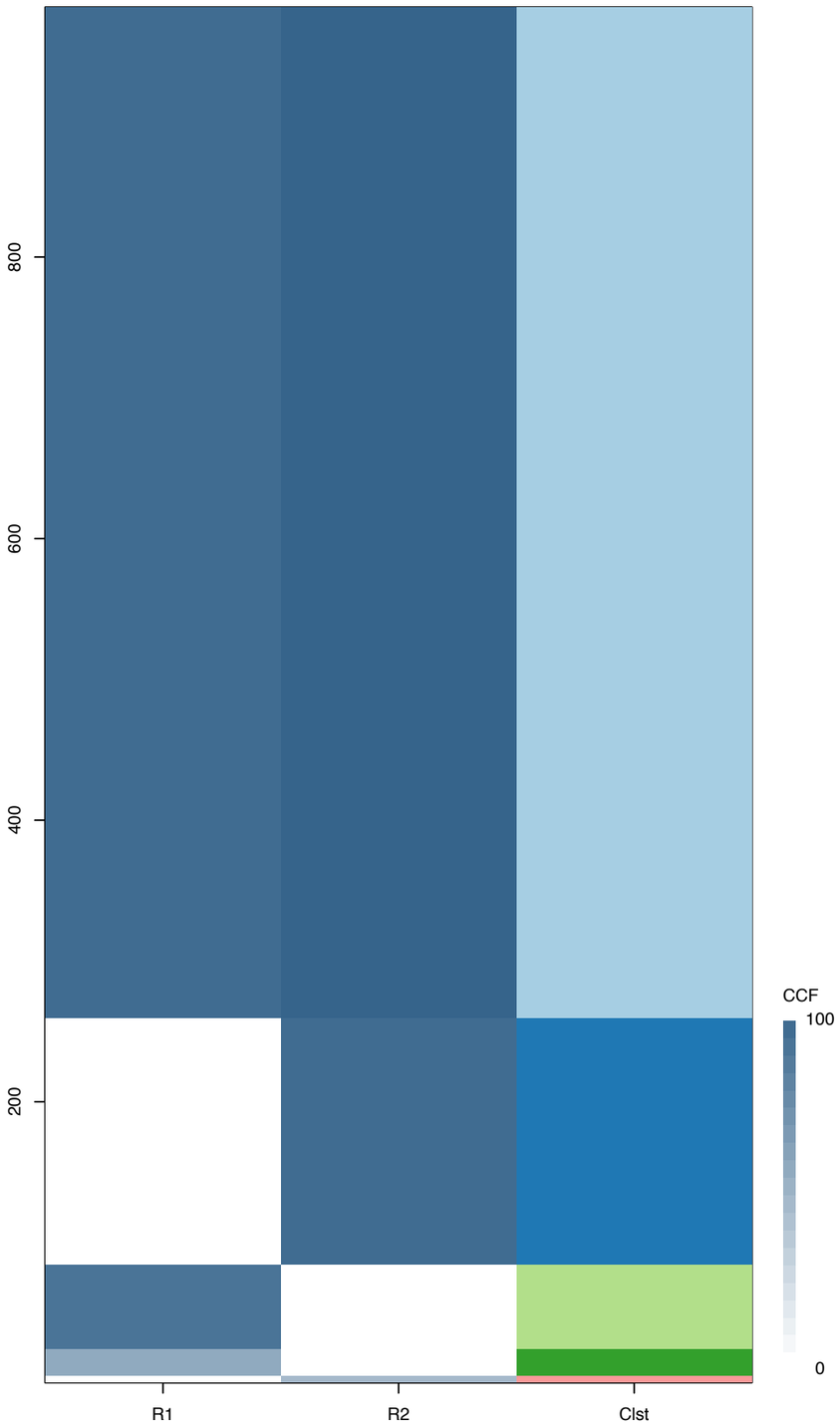
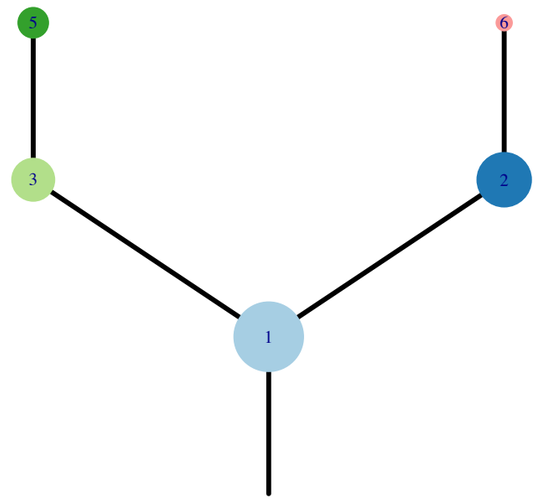


Fig.S12AA



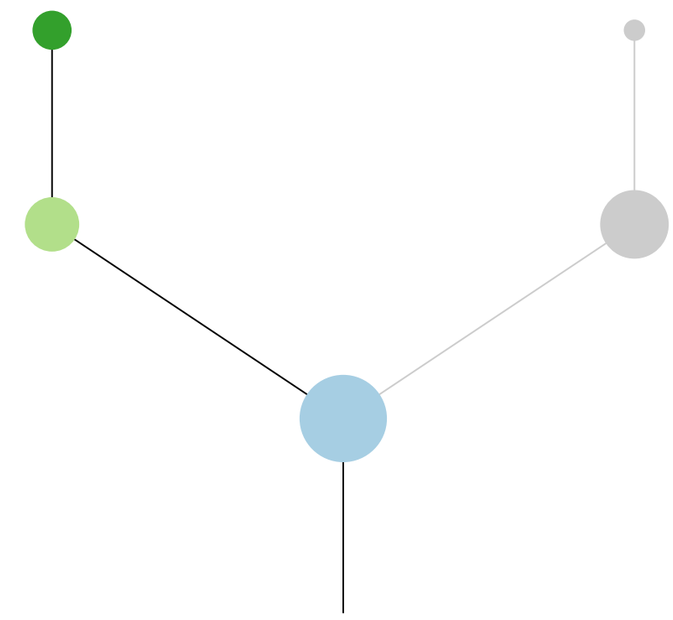
CRUK0027



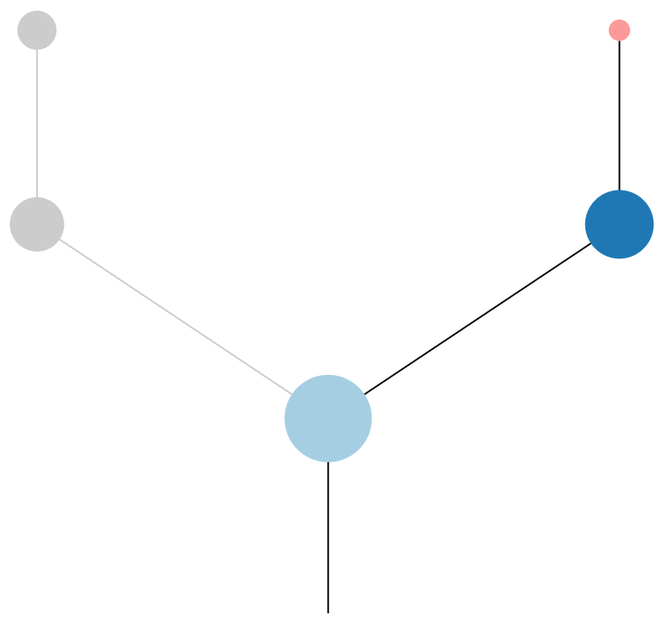
Histology:Adeno, Age:59, PackYears:30, Size:22
Stage:2a, Gender:Female, GD:Not GD, Recur:no

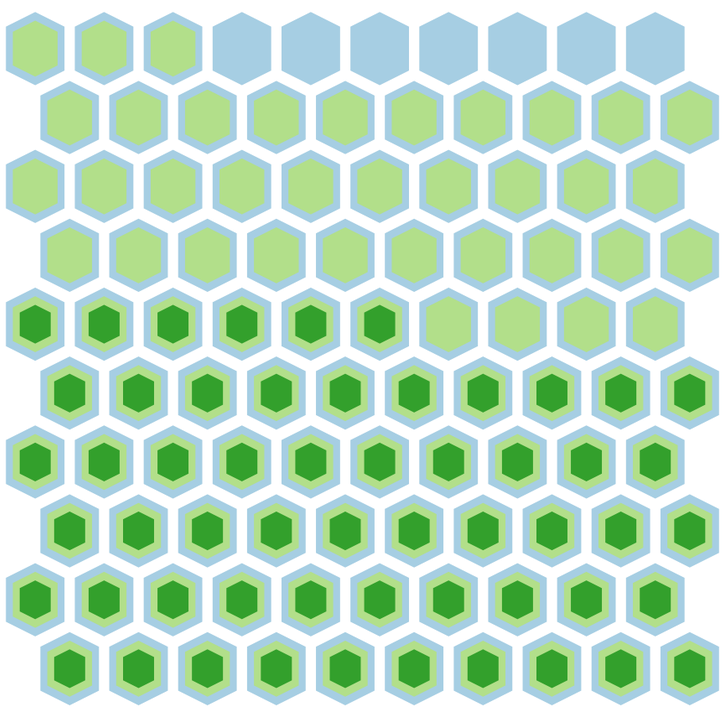
Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
TP53	1	17p13.1	SNV
DOT1L	1	19p13.3	SNV
PLXNB2	2	22q13.33	SNV

R1



R2





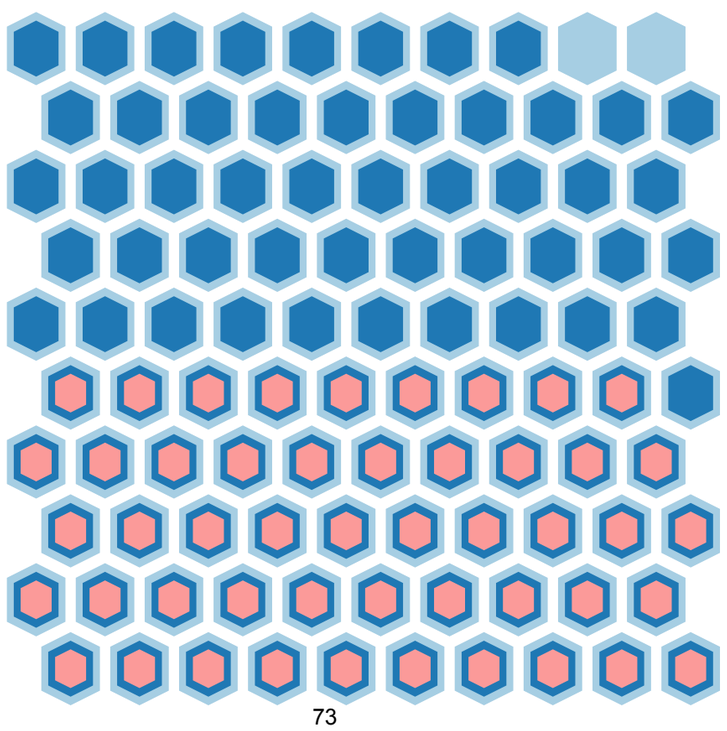
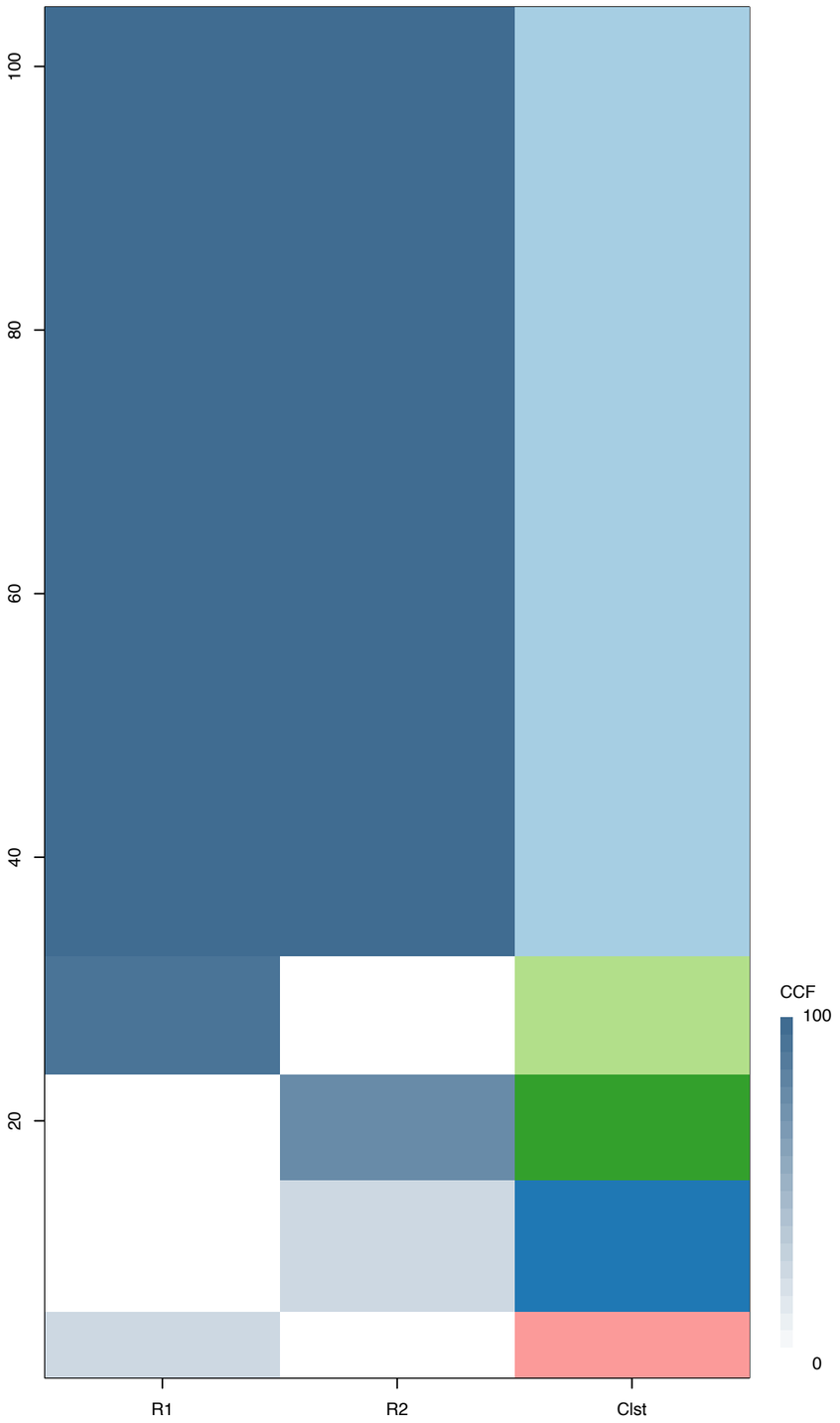
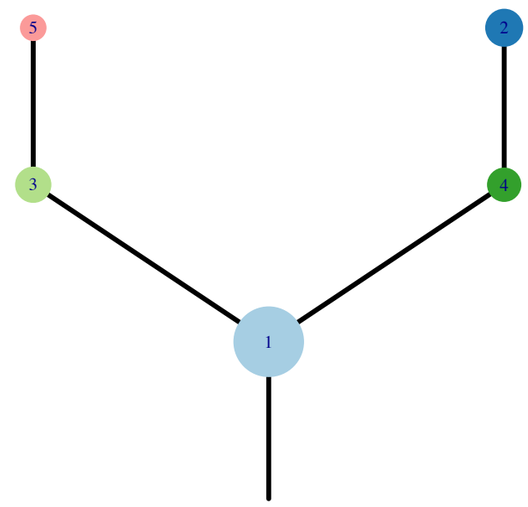


Fig.S12AB



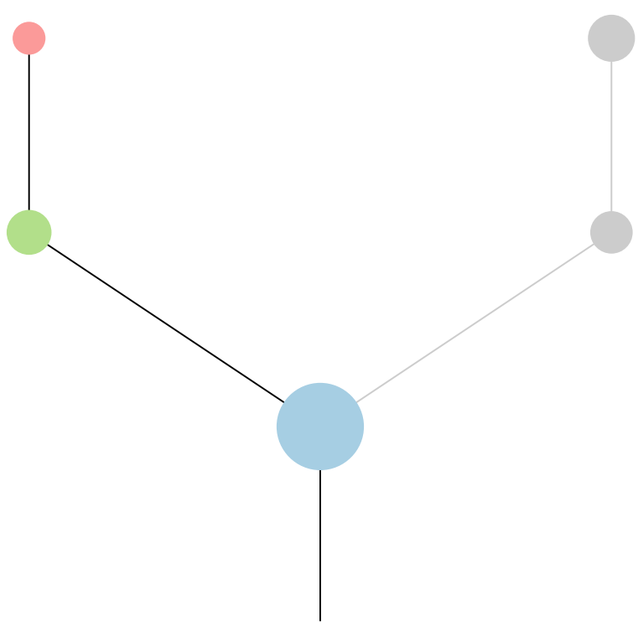
CRUK0028_A



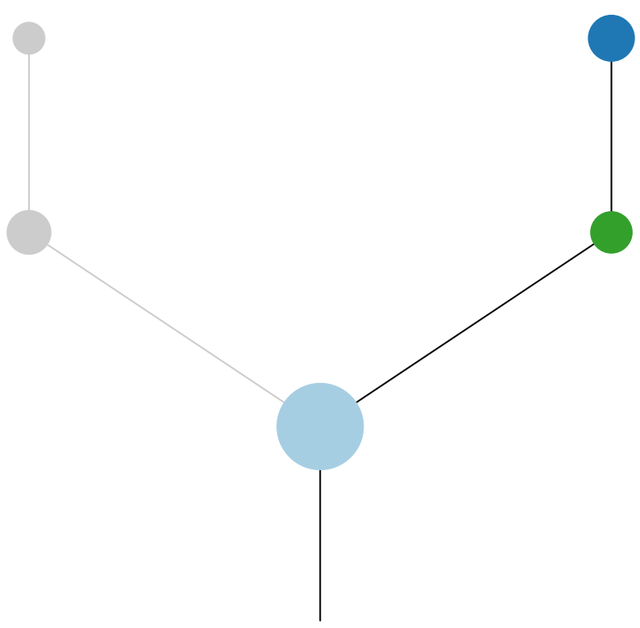
Gene	Cluster	Cytoband	Type
APC	1	5q22.2	SNV
RAC1	1	7p22.1	Amp
EGFR	1	7p11.2	Amp
NCOA6	2	20q11.22	SNV
FOXL2	3	3q22.3	Amp
TP53	3	17p13.1	SNV
PIK3CA	5	3q26.32	SNV

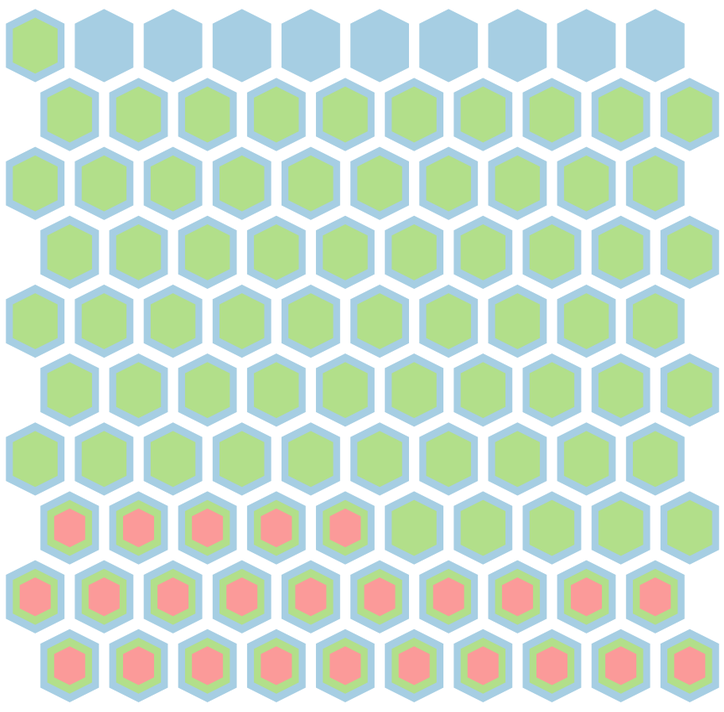
Histology:Adeno, Age:80, PackYears:17.5, Size:22
Stage:1b, Gender:Female, GD:Not GD, Recur:no

R1



R2





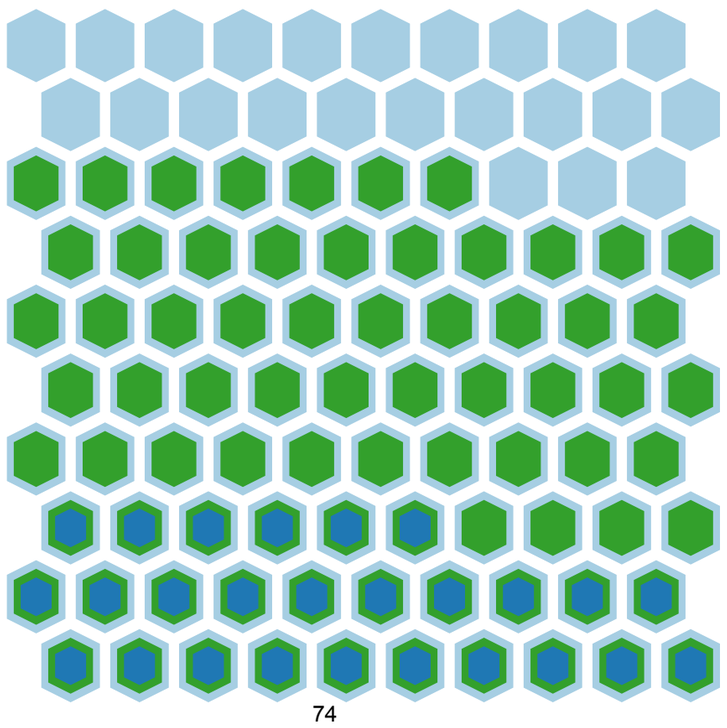
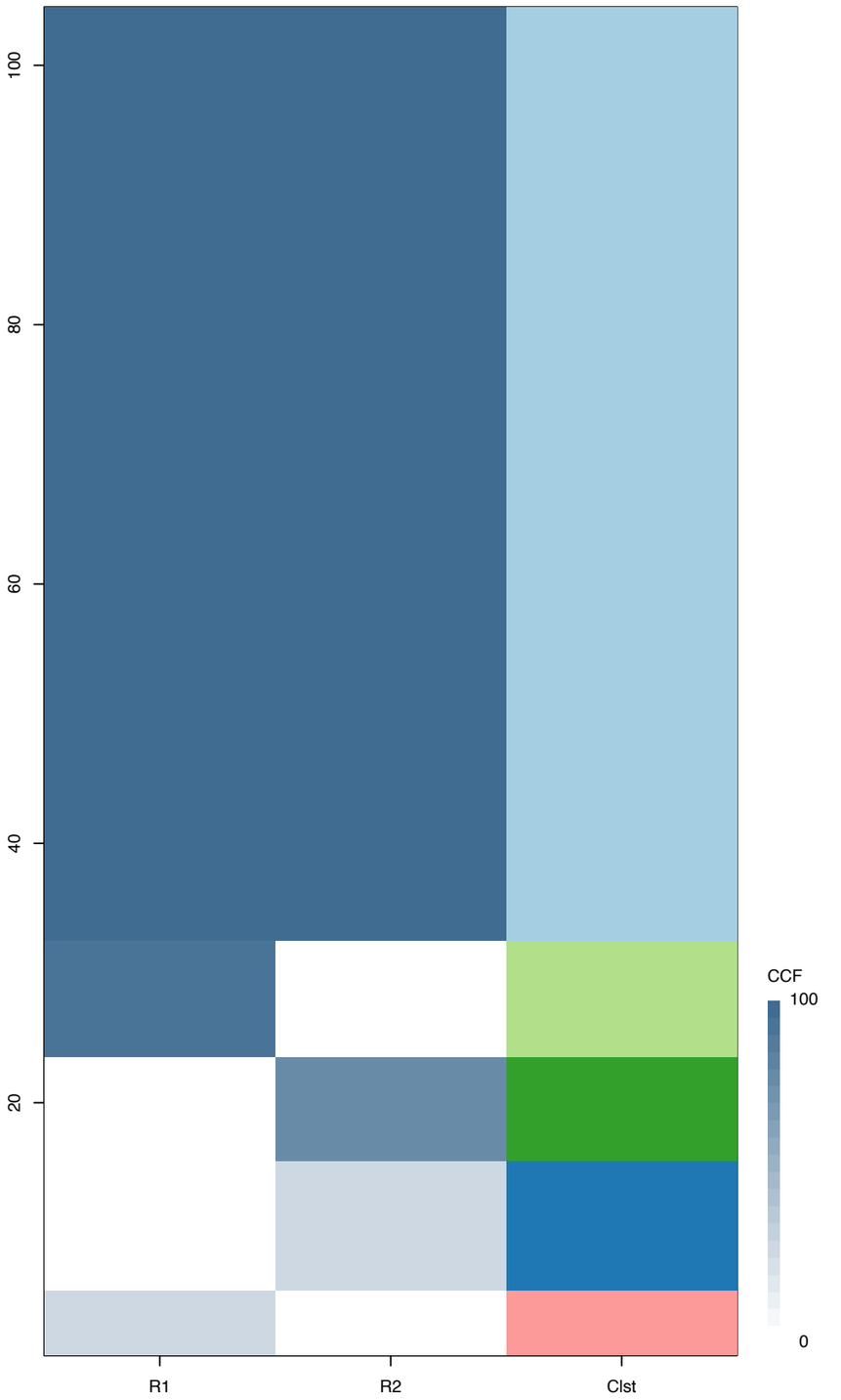
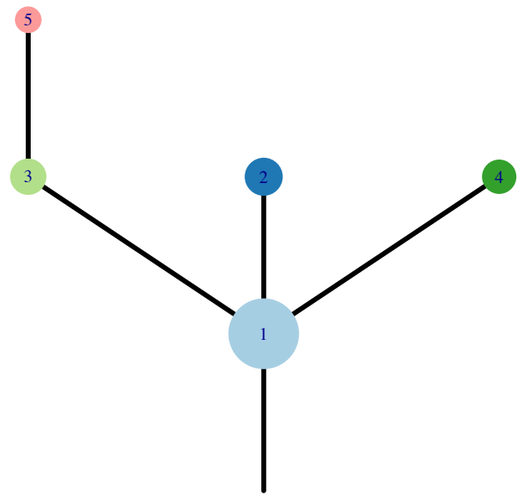


Fig.S12AB



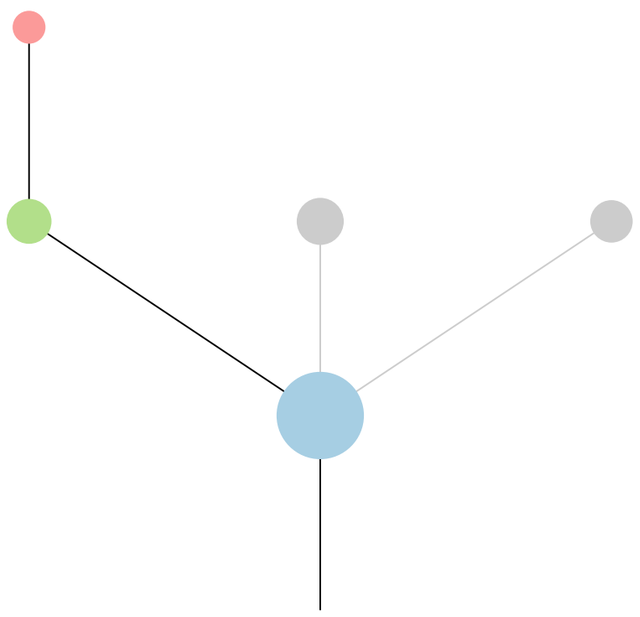
CRUK0028_B



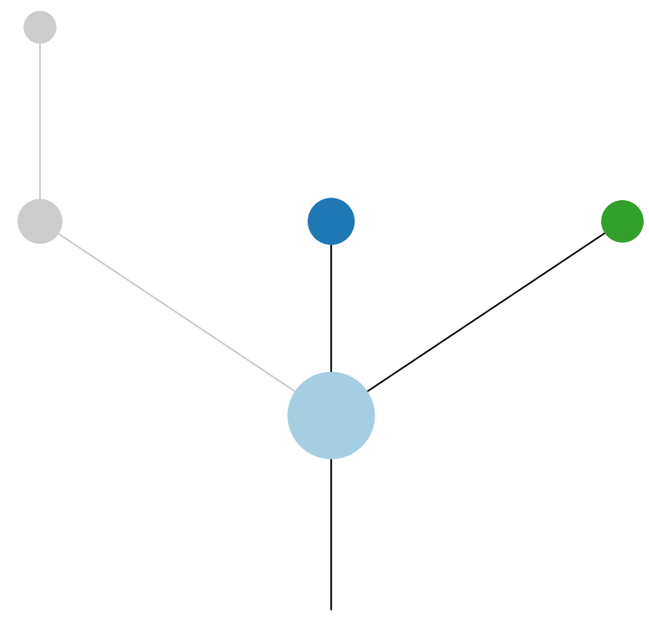
Gene	Cluster	Cytoband	Type
APC	1	5q22.2	SNV
RAC1	1	7p22.1	Amp
EGFR	1	7p11.2	Amp
NCOA6	2	20q11.22	SNV
FOXL2	3	3q22.3	Amp
TP53	3	17p13.1	SNV
PIK3CA	5	3q26.32	SNV

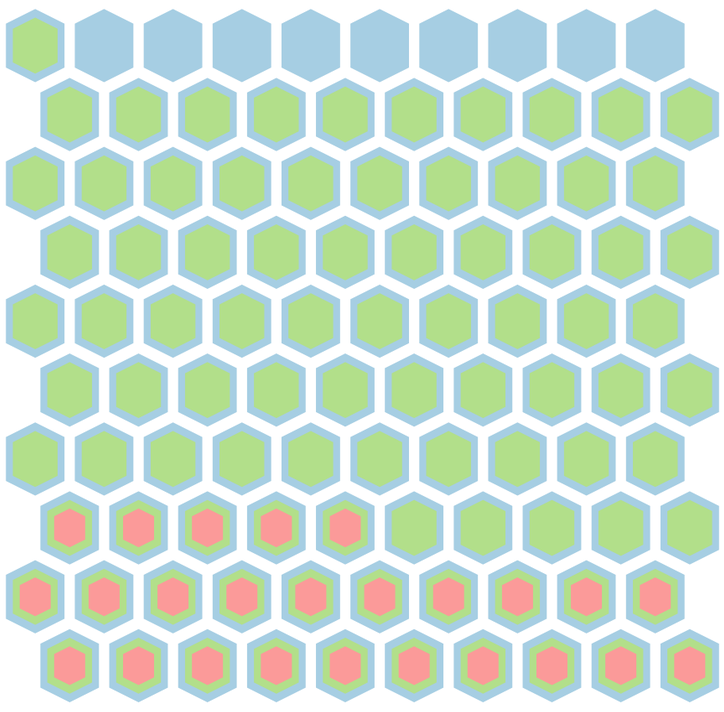
Histology:Adeno, Age:80, PackYears:17.5, Size:22
Stage:1b, Gender:Female, GD:Not GD, Recur:no

R1



R2





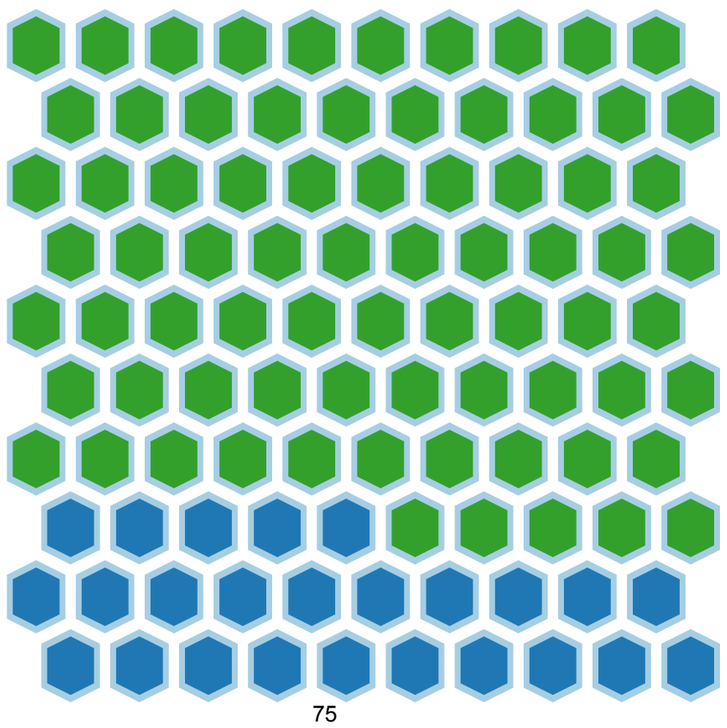
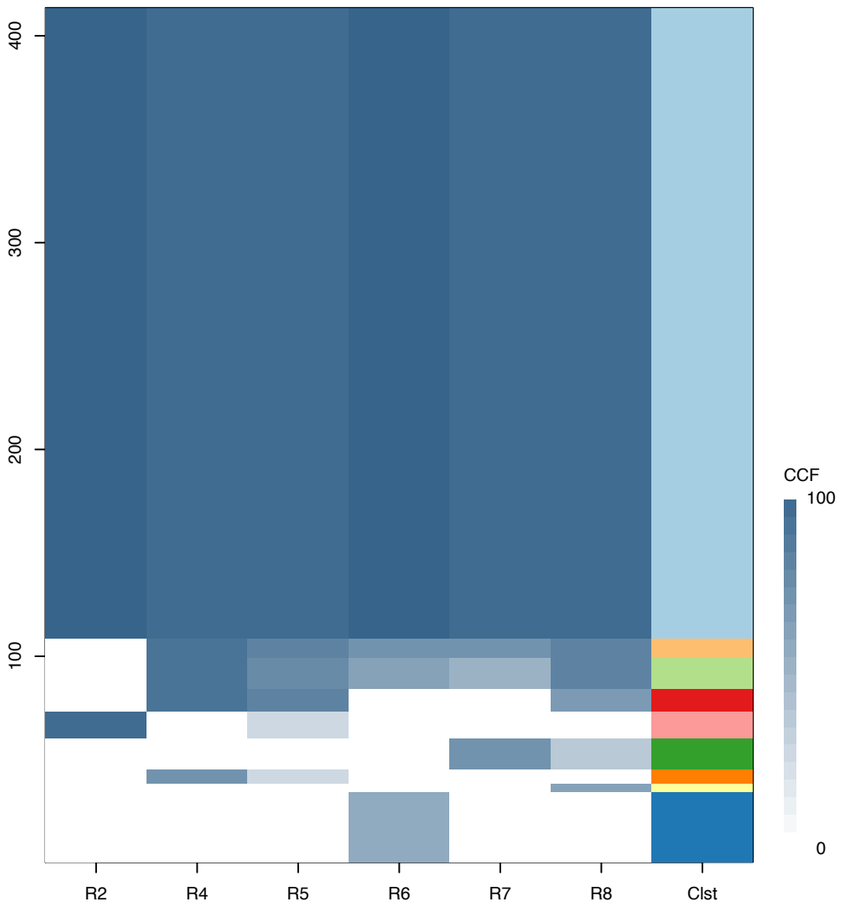
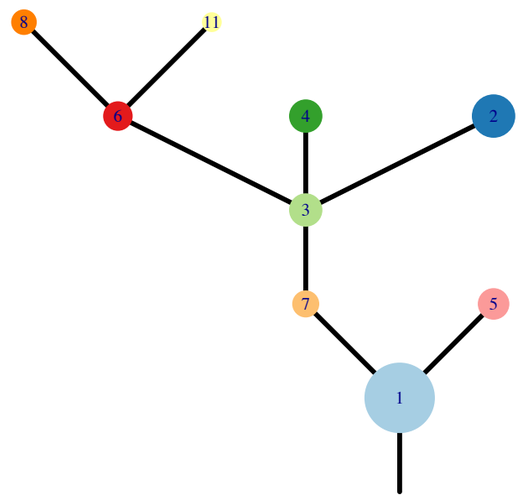


Fig.S12AC



CRUK0029



Histology:Adeno, Age:57, PackYears:35, Size:70
Stage:3a, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
NRAS	1	1p13.2	SNV
MALAT1	1	11q13.1	Amp
CCND1	1	11q13.3	Amp
NUMA1	1	11q13.4	Amp
NKX2−1	1	14q13.3	Amp
FOXA1	1	14q21.1	Amp
MGA	1	15q15.1	SNV
TP53	1	17p13.1	SNV
AKAP9	6	7q21.2	Amp

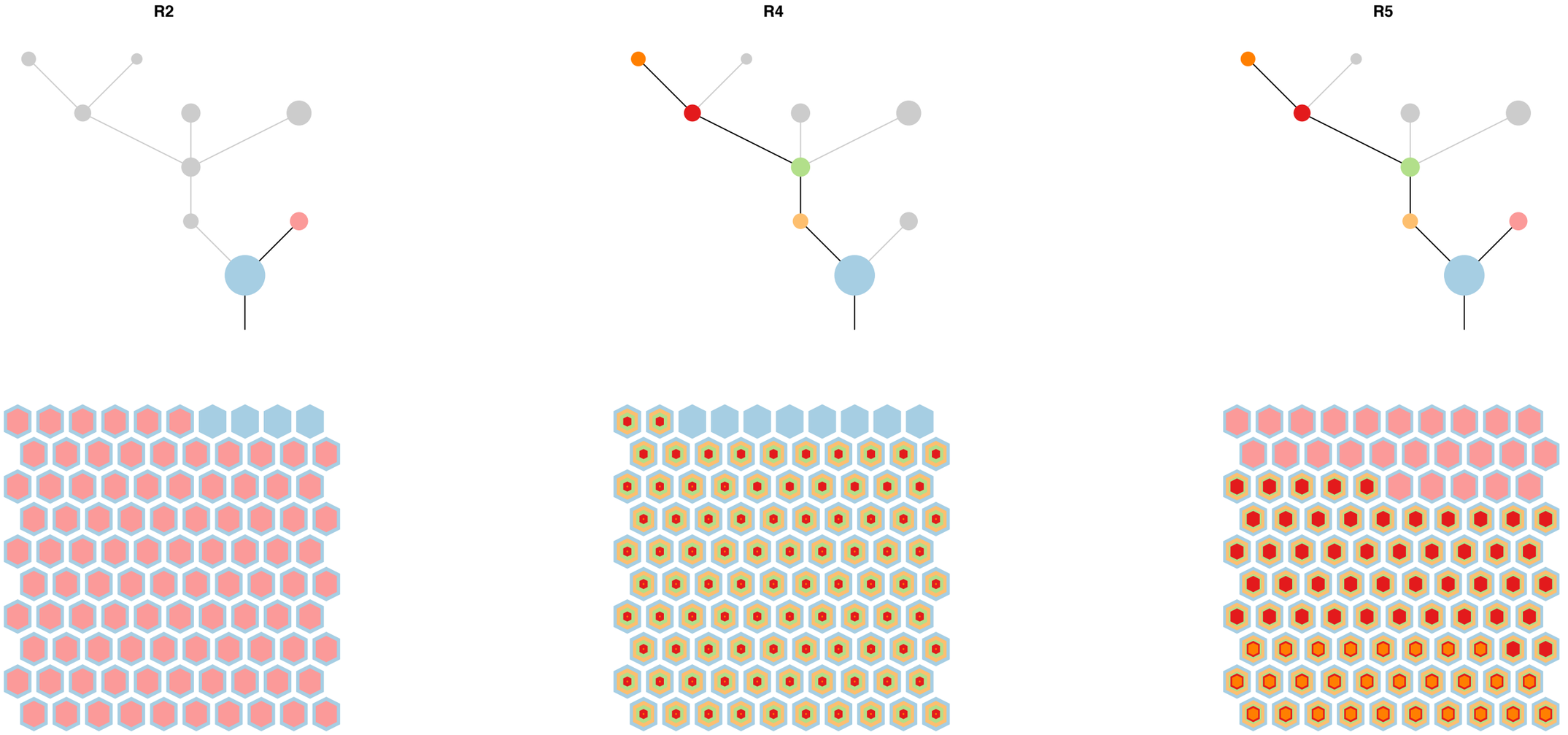
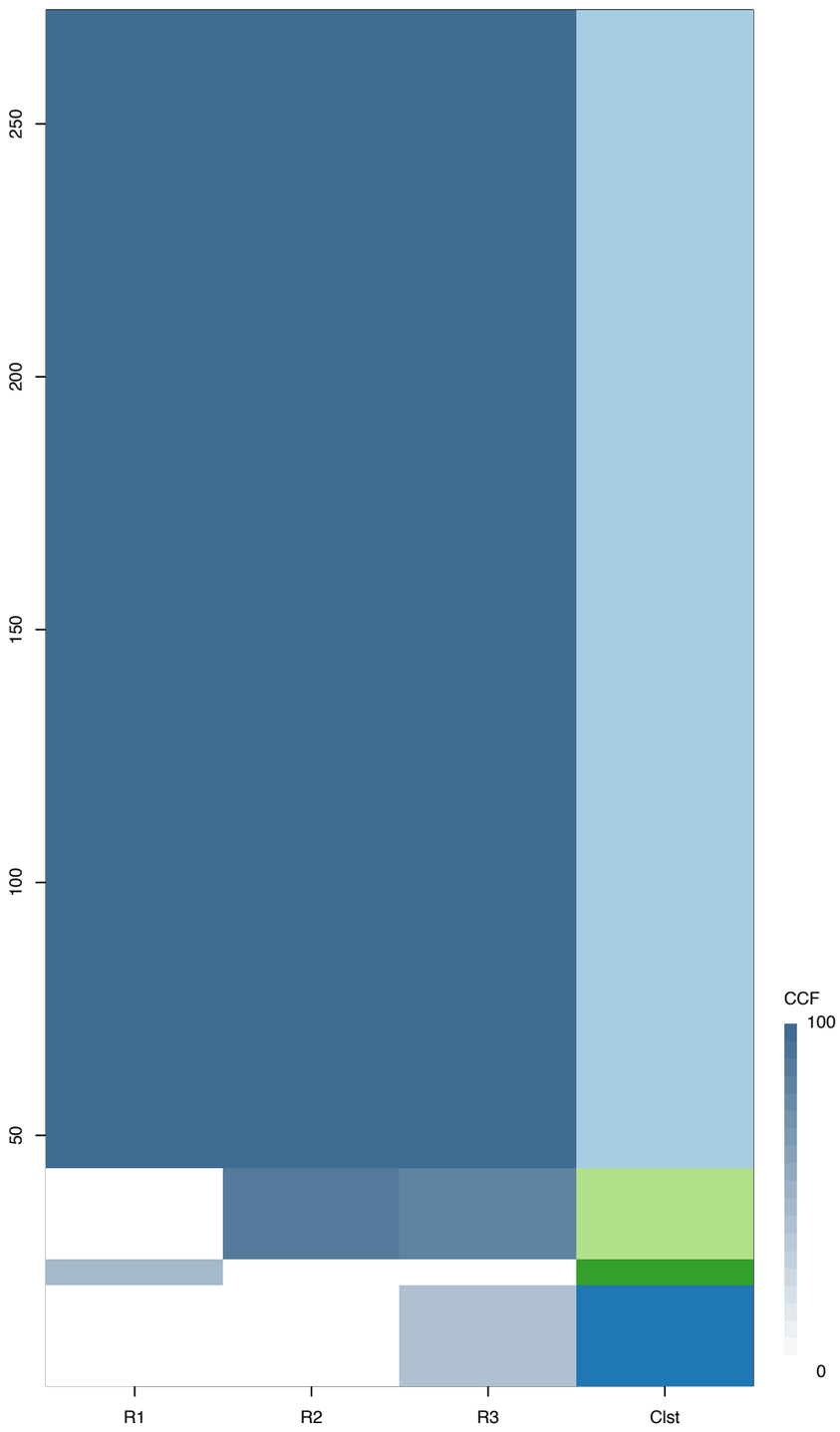
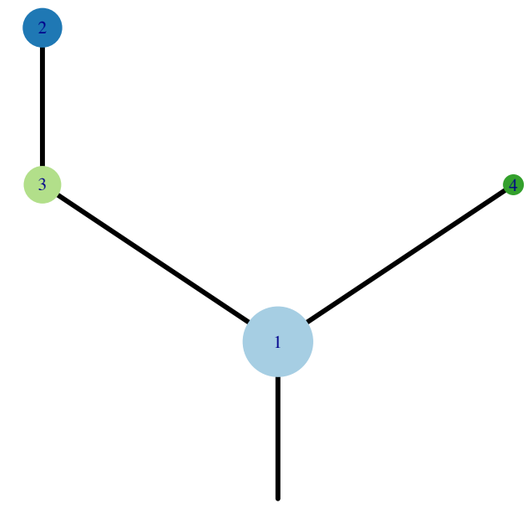


Fig.S12AD



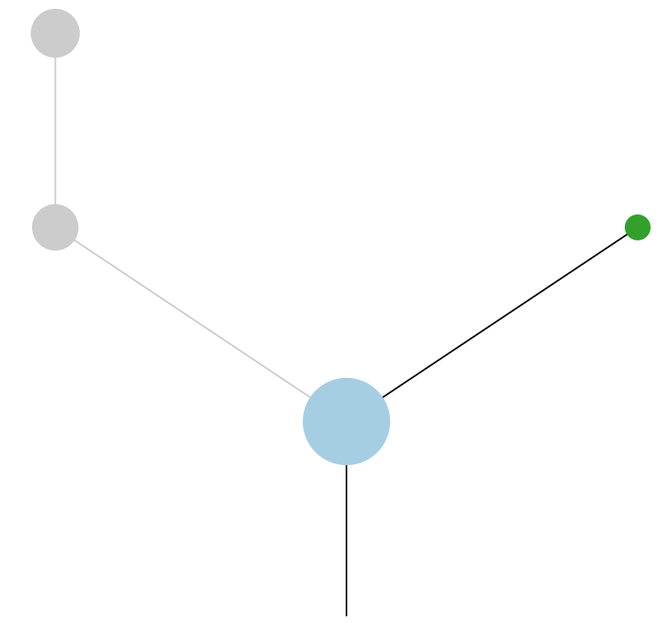
CRUK0030



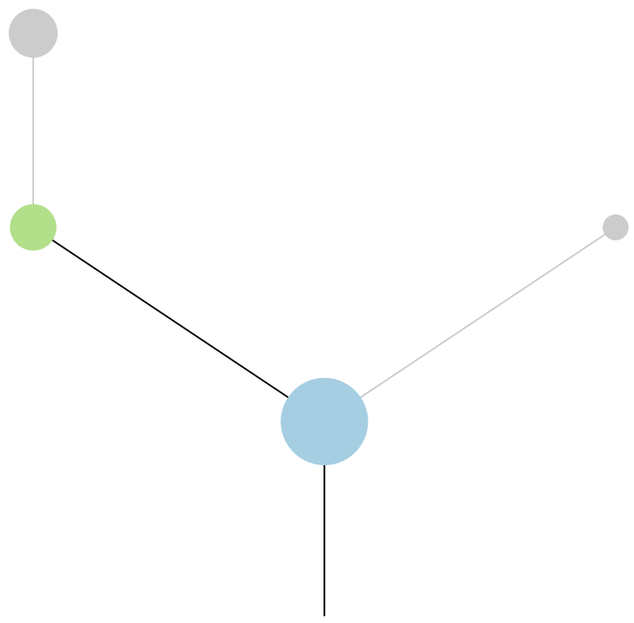
Gene	Cluster	Cytoband	Type
FBXW7	1	4q31.3	SNV
KRAS	1	12p12.1	Amp
KRAS	1	12p12.1	SNV
TSC2	1	16p13.3	SNV
TP53	1	17p13.1	SNV
NF1	1	17q11.2	SNV
U2AF1	1	21q22.3	SNV

Histology:Adeno, Age:76, PackYears:60, Size:14
Stage:3a, Gender:Male, GD:Not GD, Recur:no

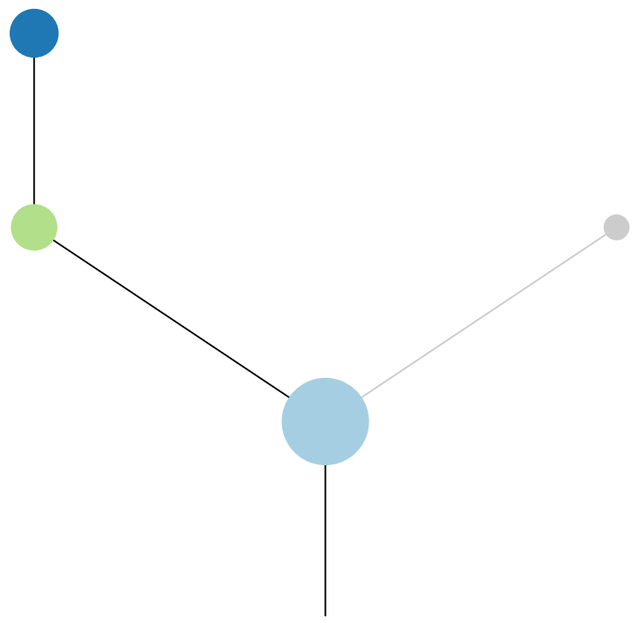
R1

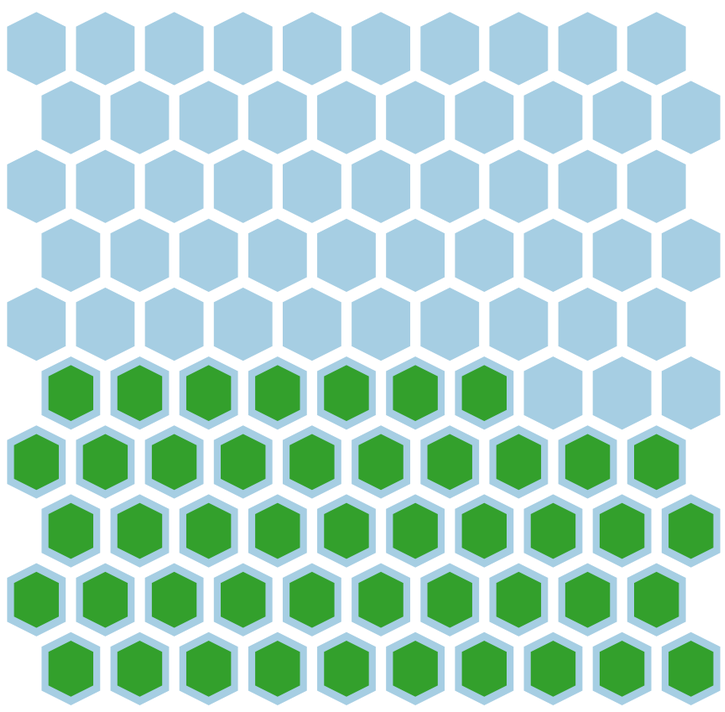


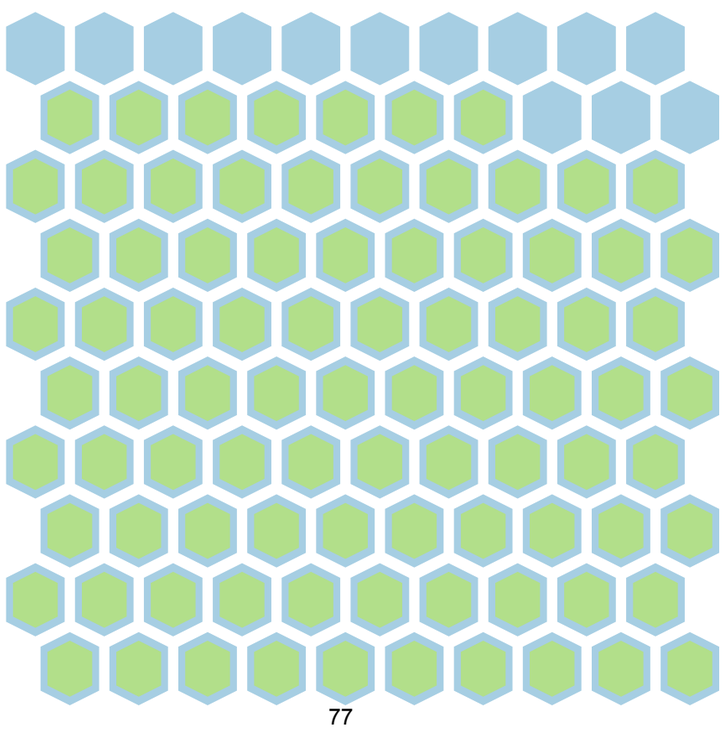
R2



R3







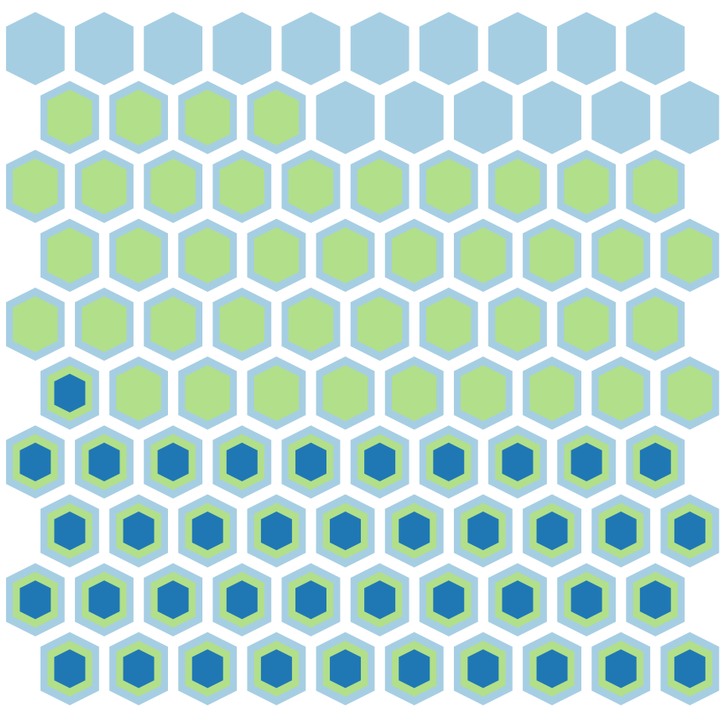
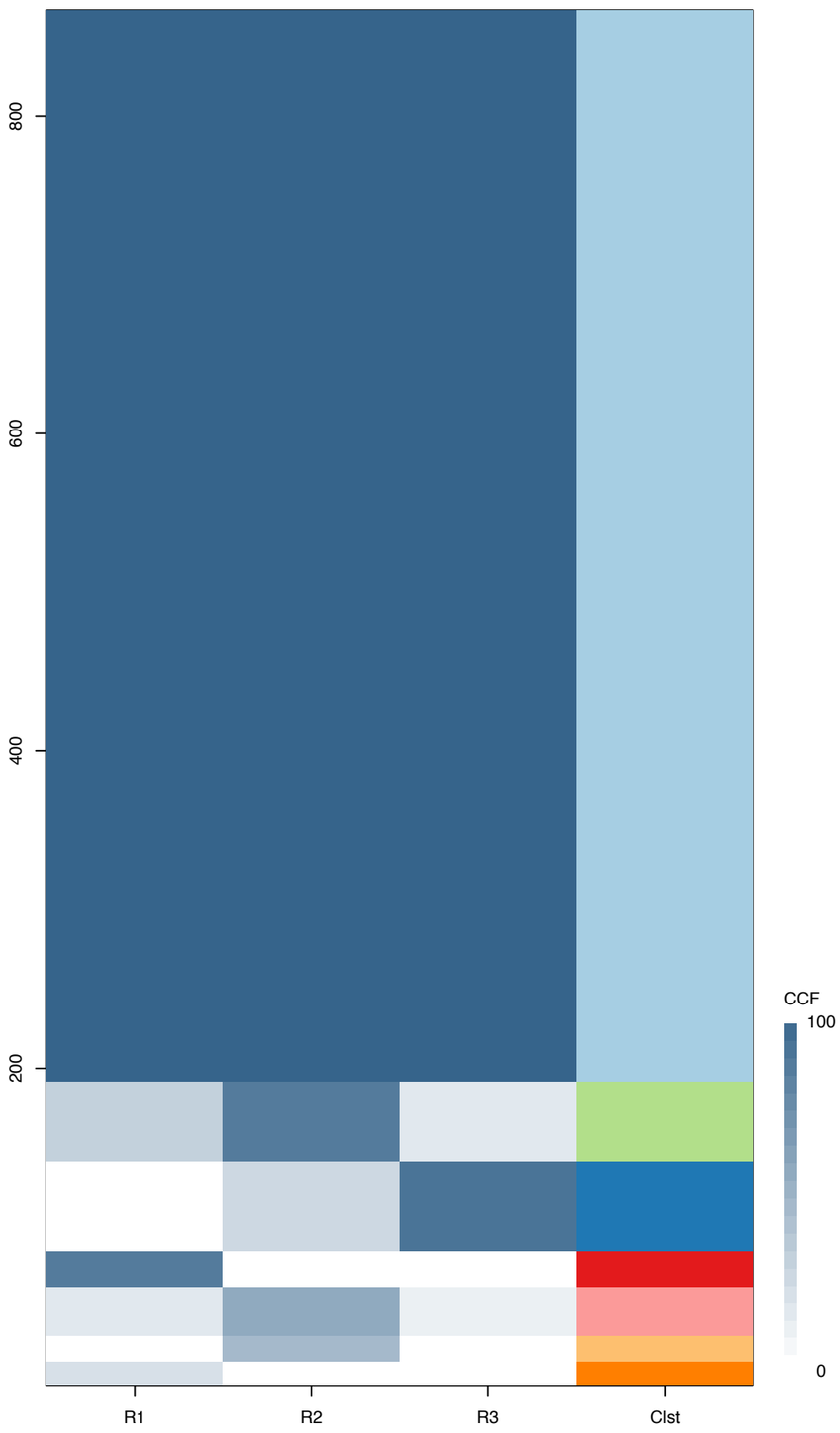
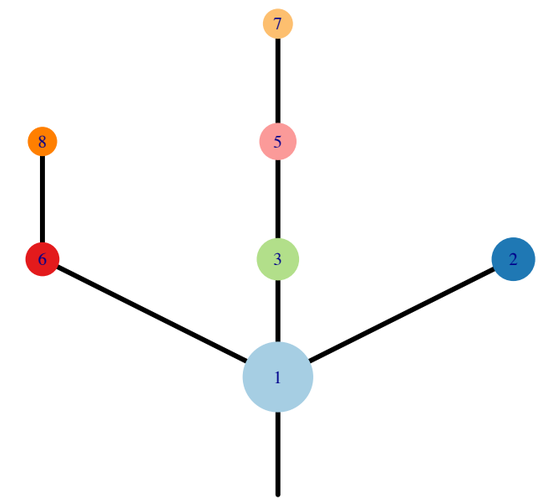


Fig.S12AE



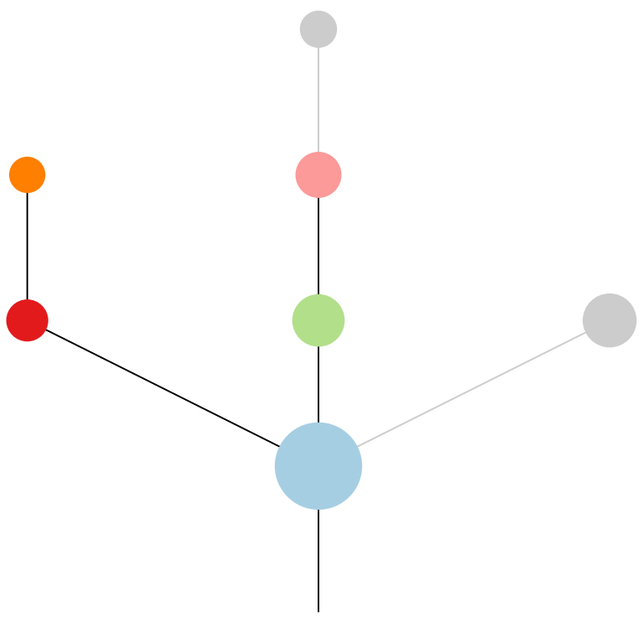
CRUK0031_A



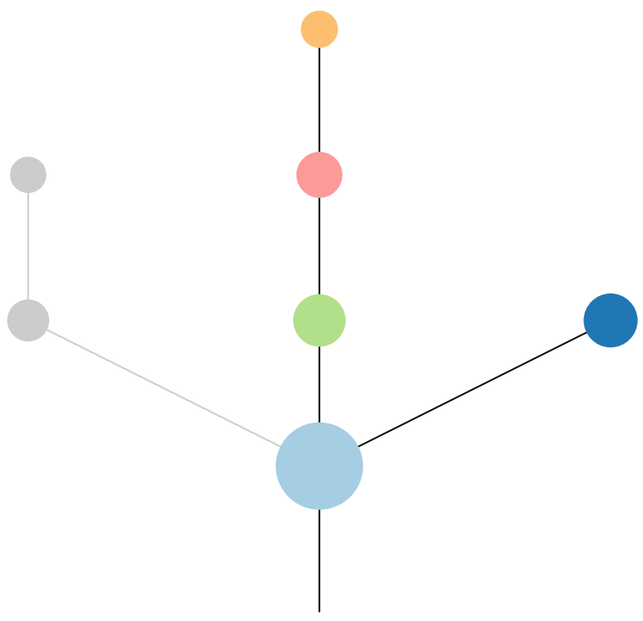
Histology:Adeno, Age:61, PackYears:30, Size:40
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
NCOA2	1	8q13.3	Amp
HEY1	1	8q21.13	Amp
CDKN2A	1	9p21.3	SNV
PRF1	1	10q22.1	SNV
ETNK1	1	12p12.1	Amp
PPFIBP1	1	12p11.23	Amp
PTPRB	1	12q15	SNV
LCP1	1	13q14.13	Amp
KEAP1	1	19p13.2	SNV
BCR	1	22q11.23	Amp
NF1	6	17q11.2	SNV
CASP8	?	2q33.1	SNV

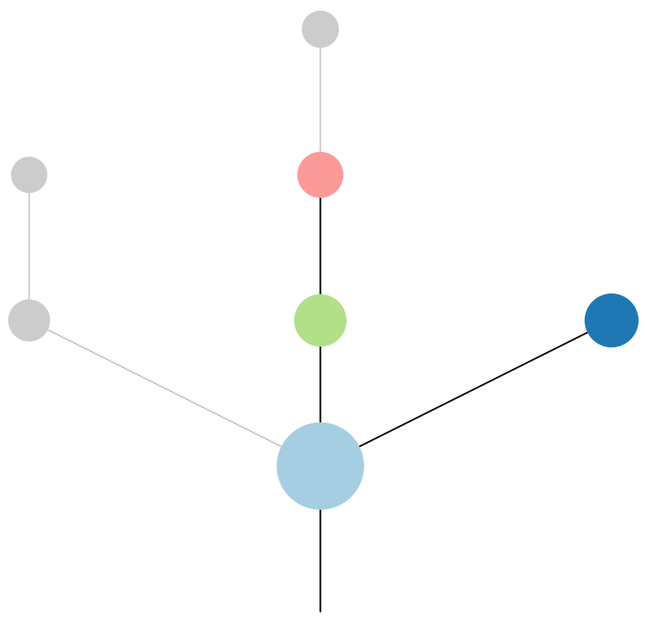
R1

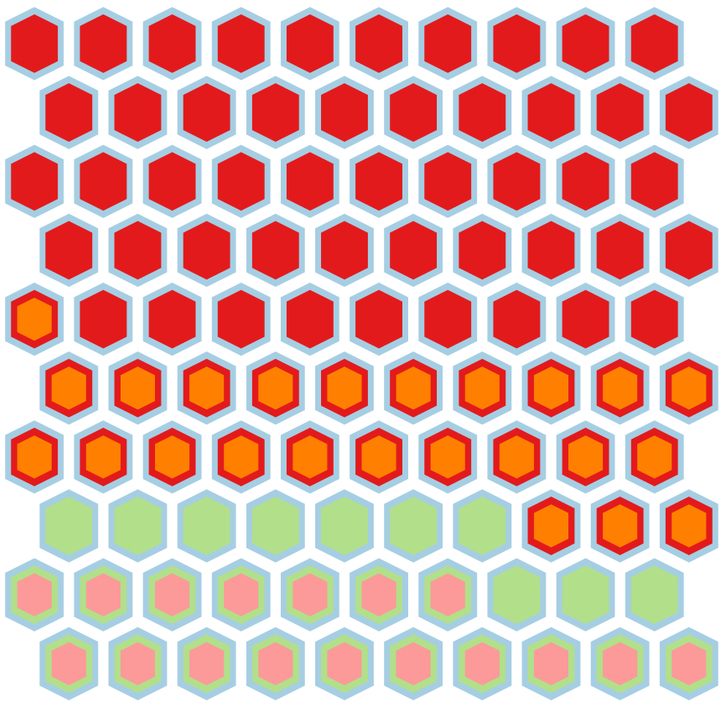


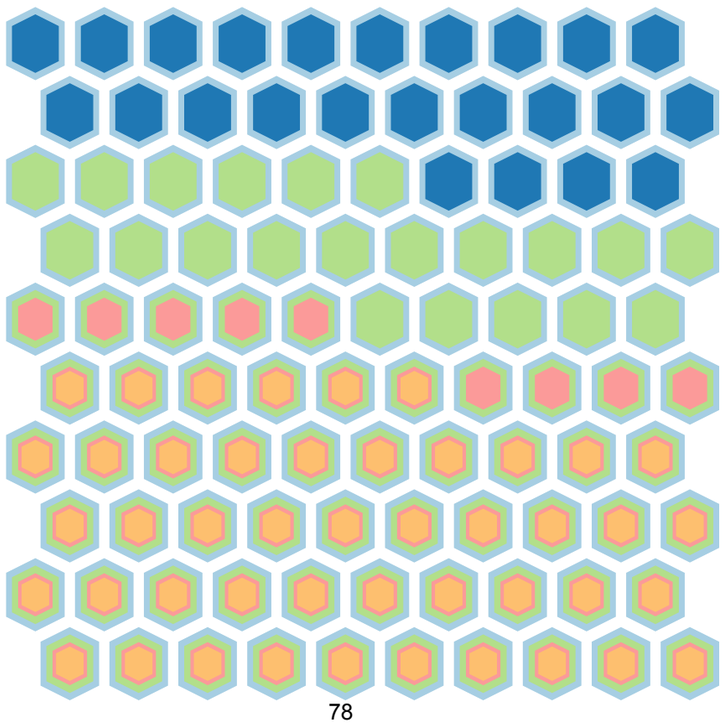
R2



R3







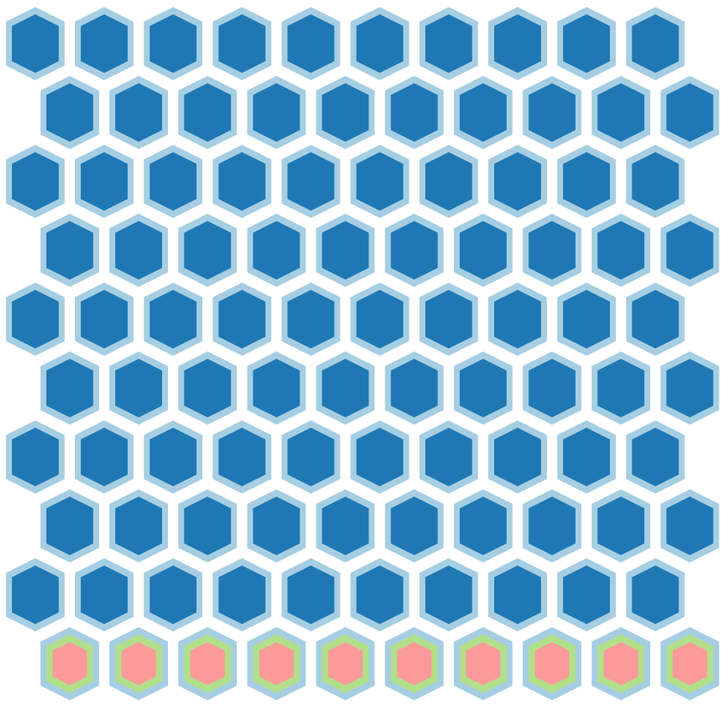
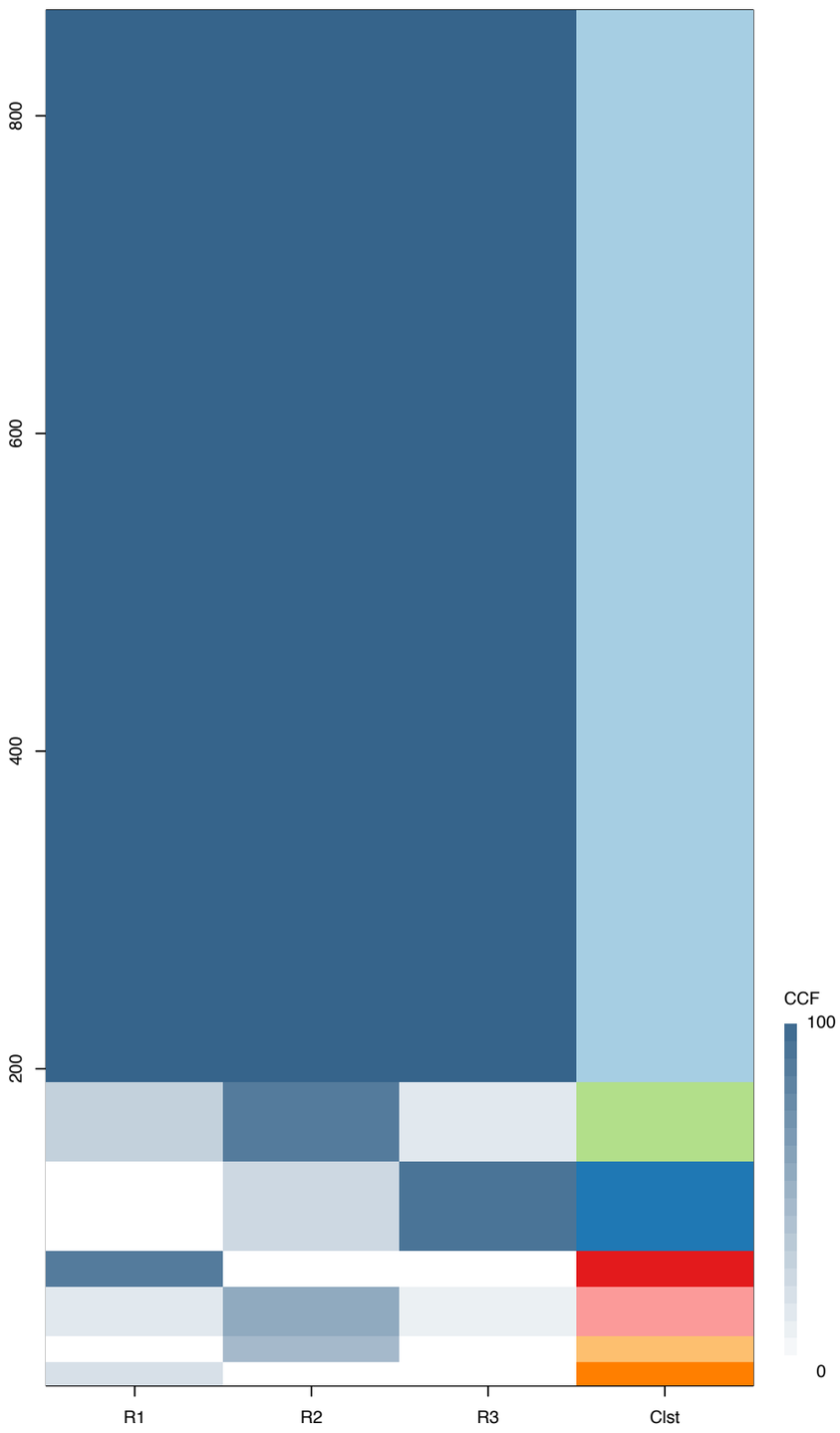
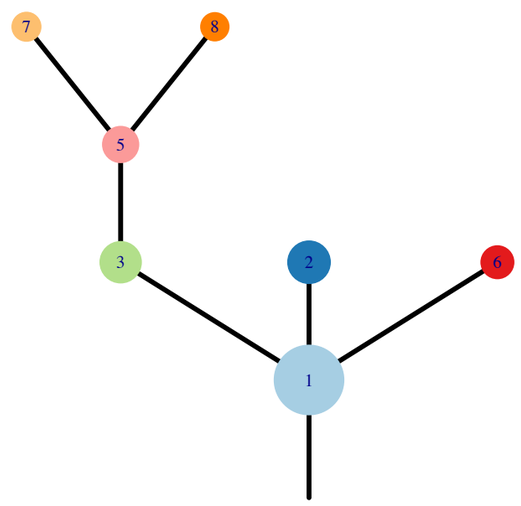


Fig.S12AE



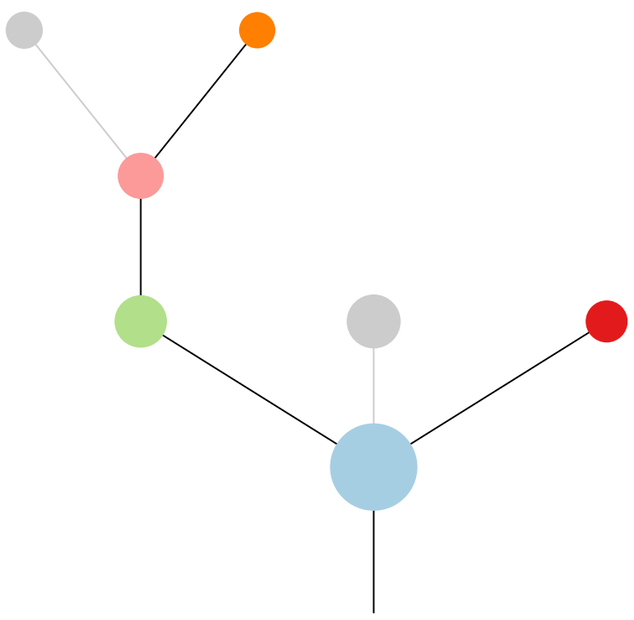
CRUK0031_B

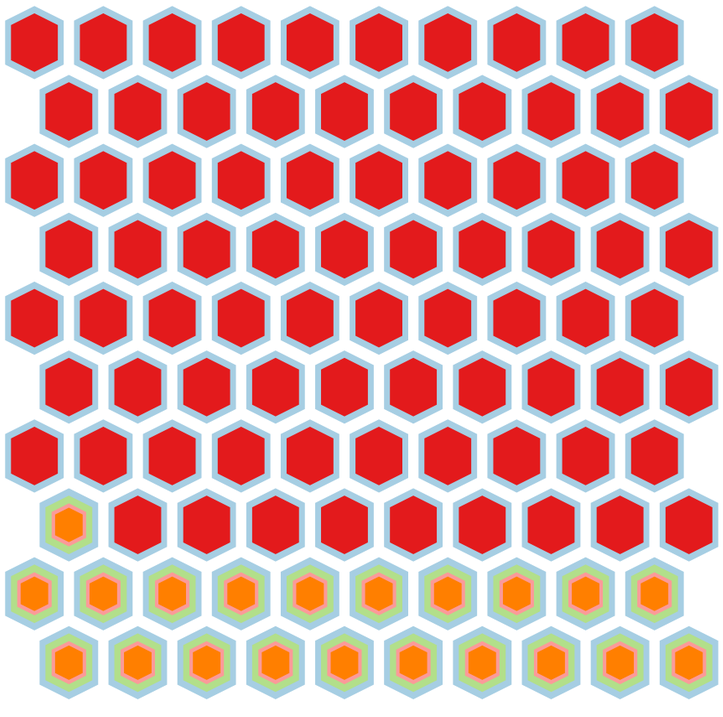


Histology:Adeno, Age:61, PackYears:30, Size:40
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

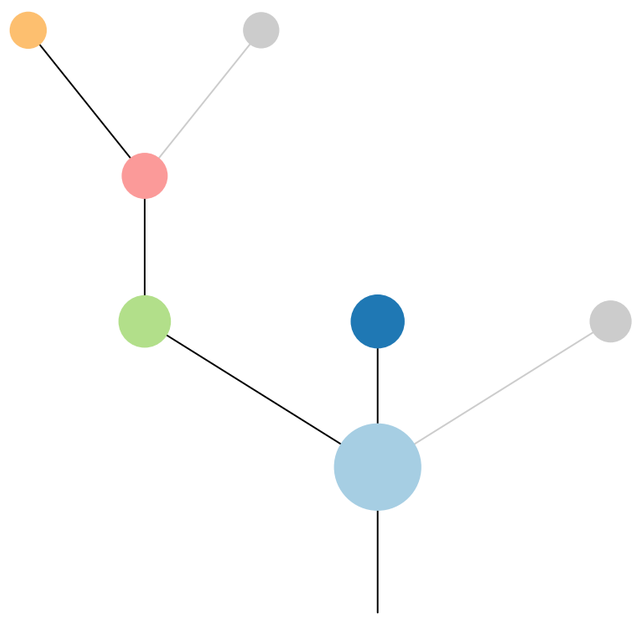
Gene	Cluster	Cytoband	Type
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
NCOA2	1	8q13.3	Amp
HEY1	1	8q21.13	Amp
CDKN2A	1	9p21.3	SNV
PRF1	1	10q22.1	SNV
ETNK1	1	12p12.1	Amp
PPFIBP1	1	12p11.23	Amp
PTPRB	1	12q15	SNV
LCP1	1	13q14.13	Amp
KEAP1	1	19p13.2	SNV
BCR	1	22q11.23	Amp
NF1	6	17q11.2	SNV
CASP8	?	2q33.1	SNV

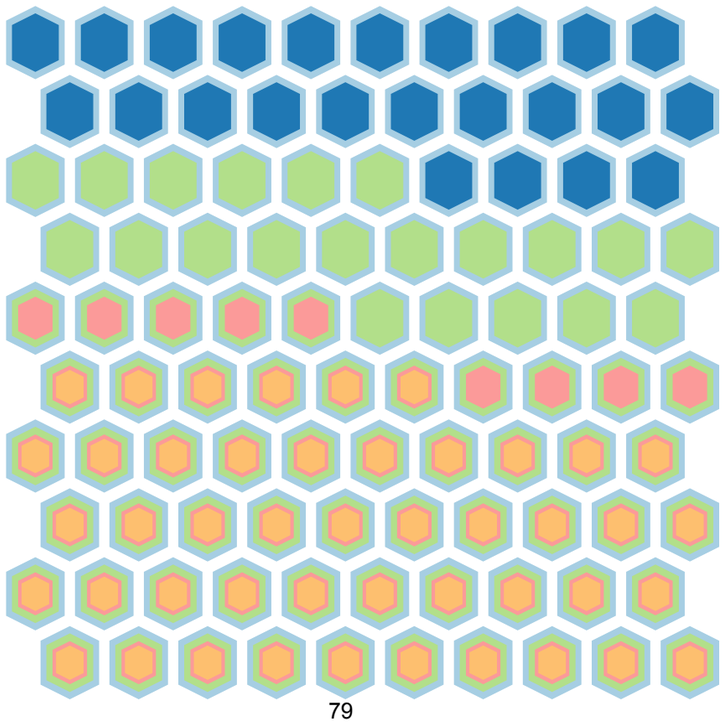
R1



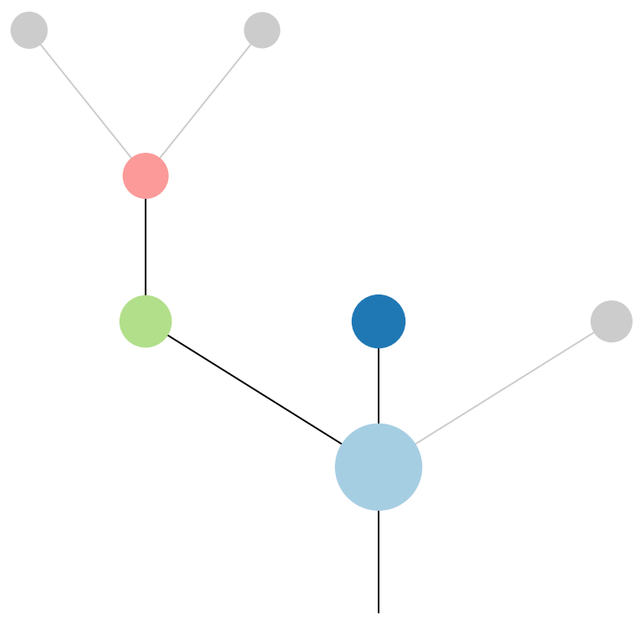


R2





R3



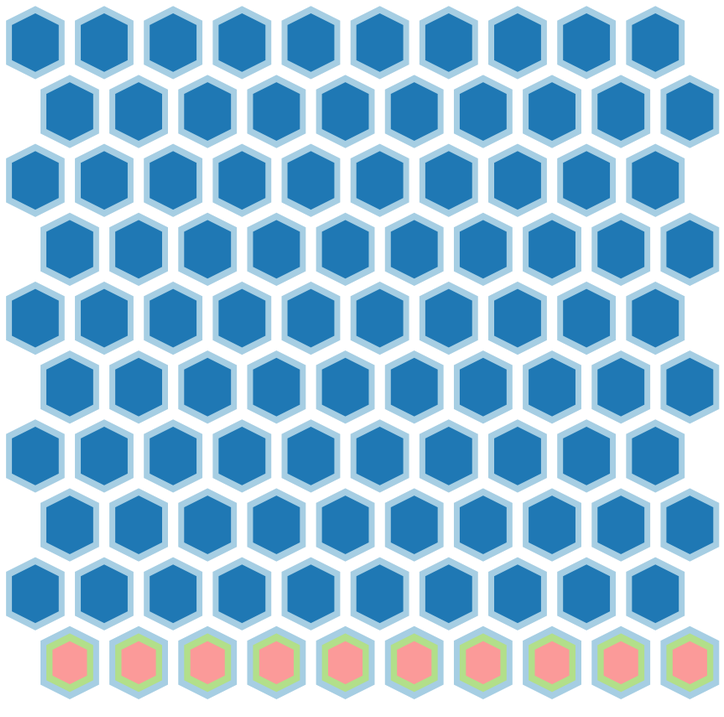
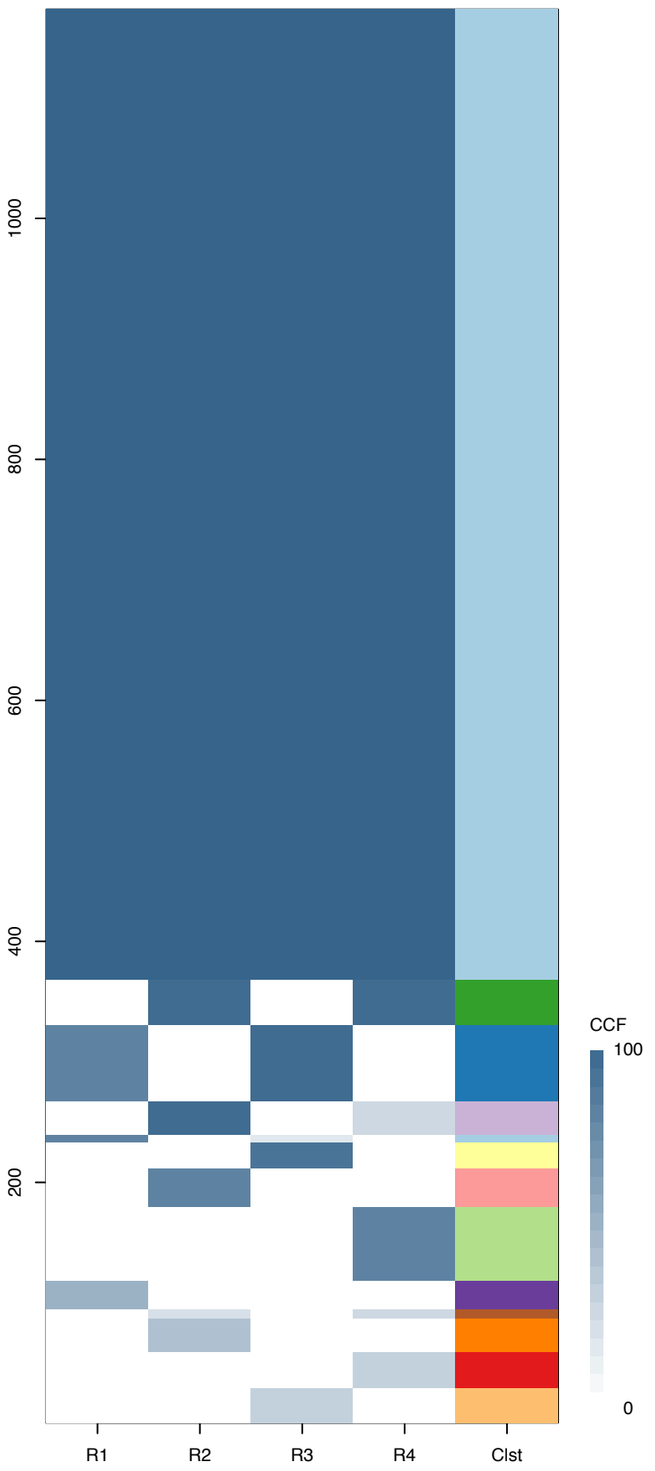
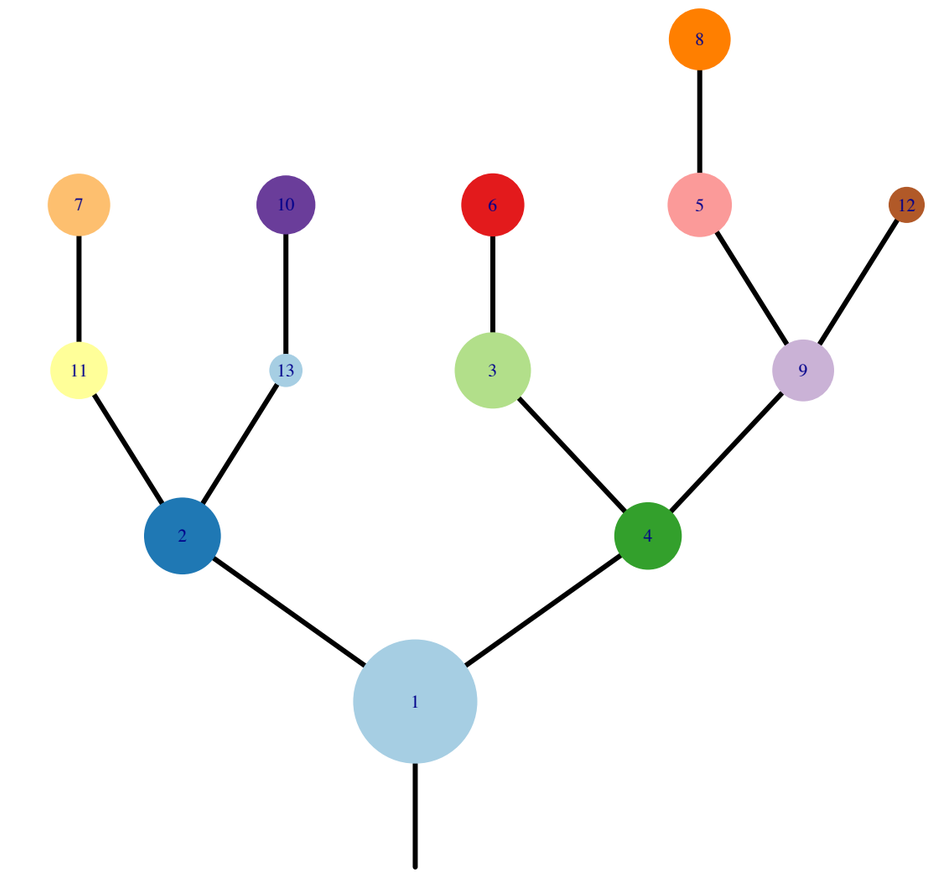


Fig.S12AF



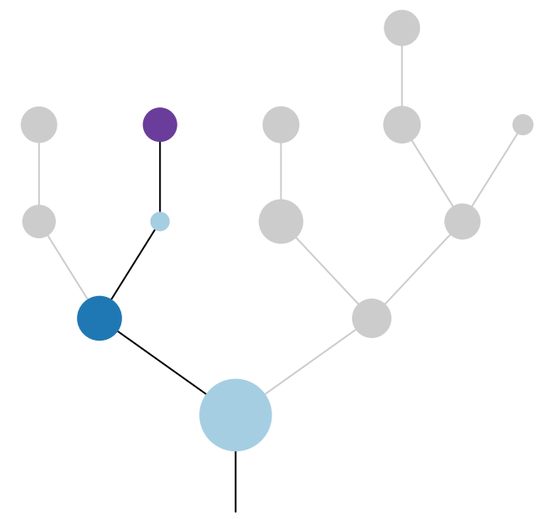
CRUK0032



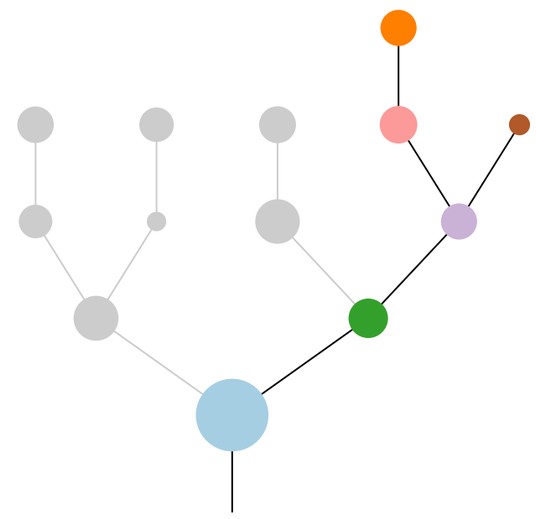
Histology:Adeno, Age:74, PackYears:58, Size:34
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
RAD21	1	8q24.11	SNV
ATM	1	11q22.3	SNV
NKX2−1	1	14q13.3	Amp
FOXA1	1	14q21.1	Amp
COL1A1	1	17q21.33	Amp
HLF	1	17q22	Amp
MSI2	1	17q22	Amp
RNF43	1	17q22	SNV
CLTC	1	17q23.1	Amp
CD79B	1	17q23.3	Amp
RNF213	1	17q25.3	Amp
U2AF1	1	21q22.3	SNV
CCND1	4	11q13.3	Amp
ARID1B	6	6q25.3	SNV
DDX5	7	17q23.3	Amp
KDM5C	?		SNV

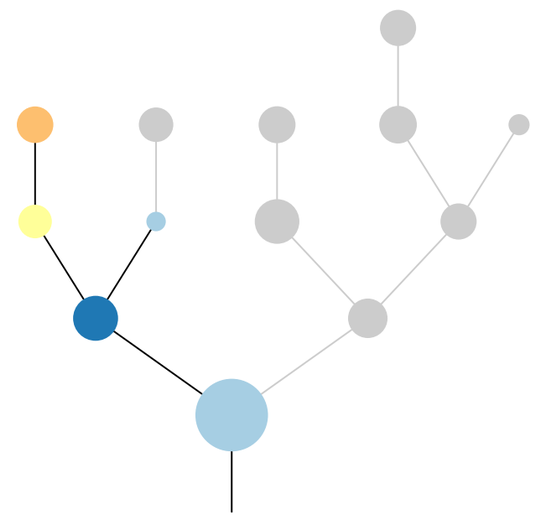
R1



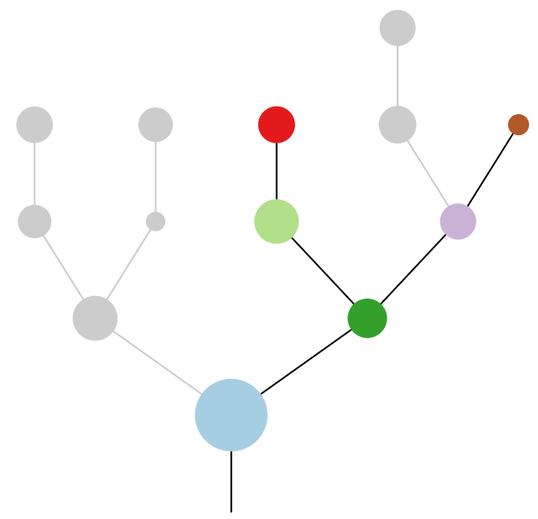
R2



R3



R4



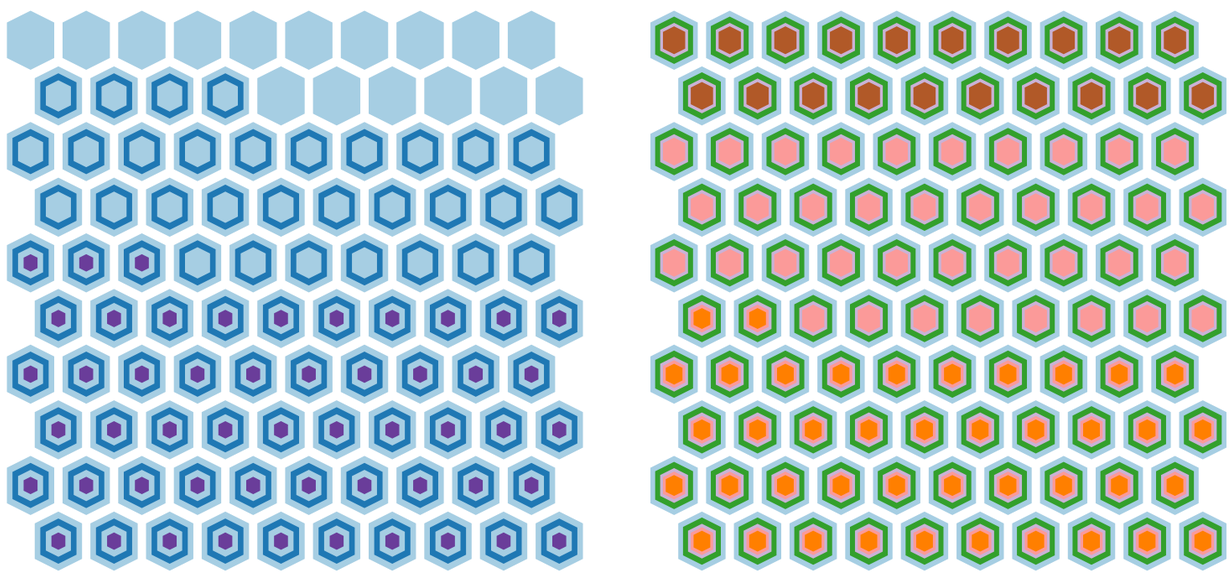
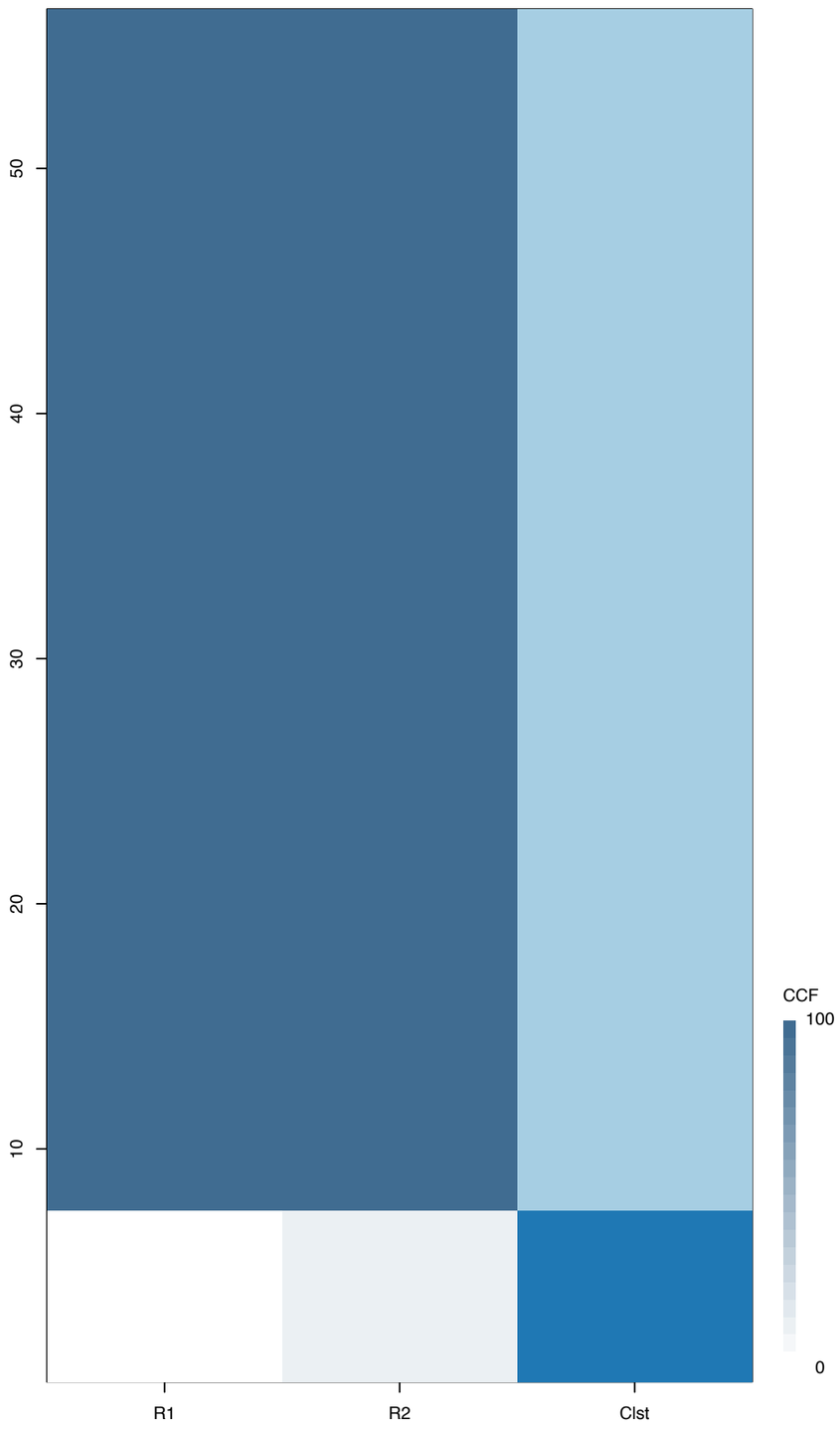
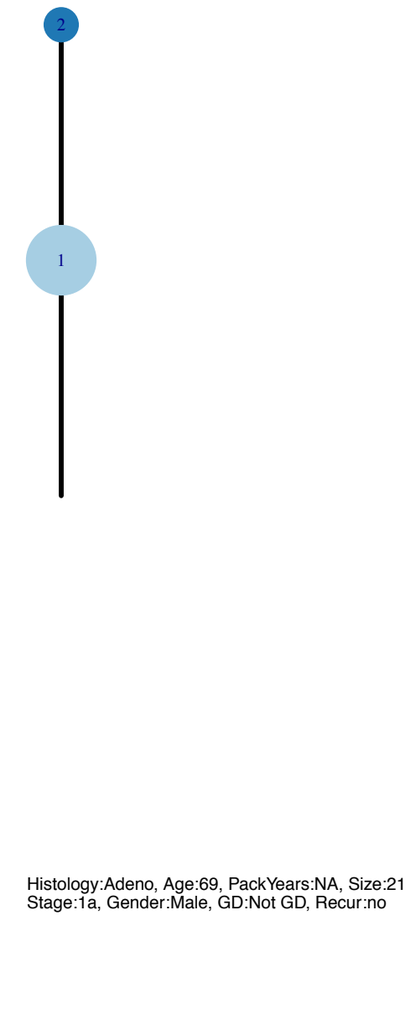
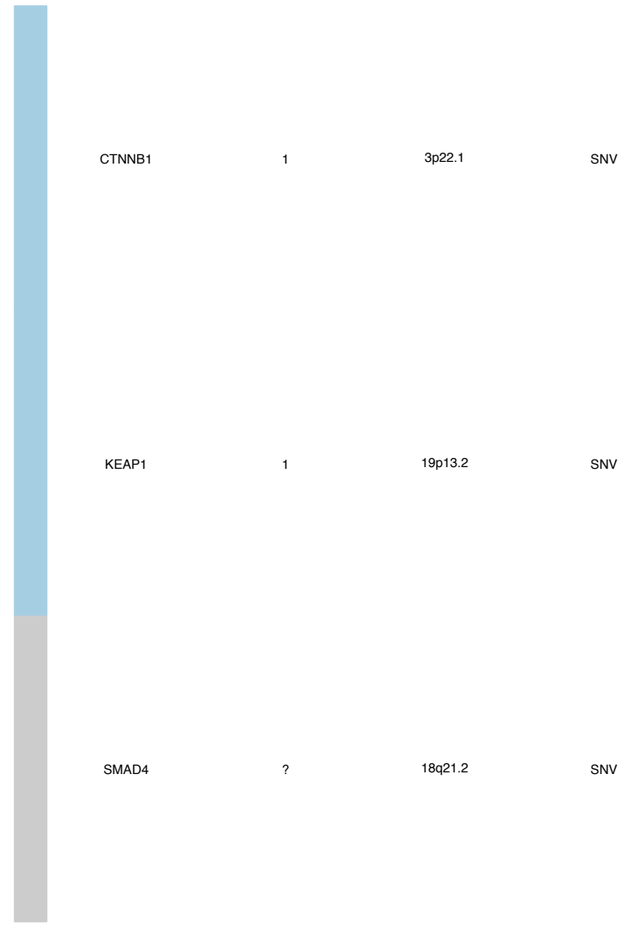


Fig.S12AG



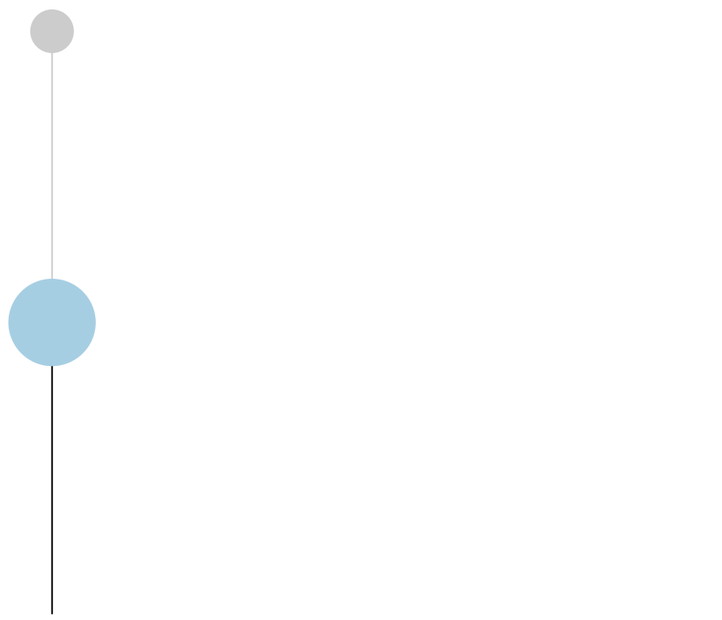
CRUK0033



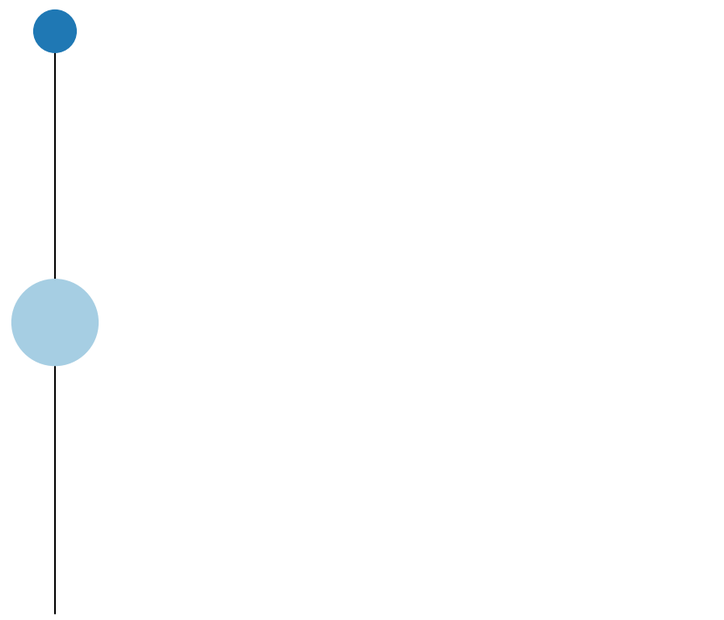


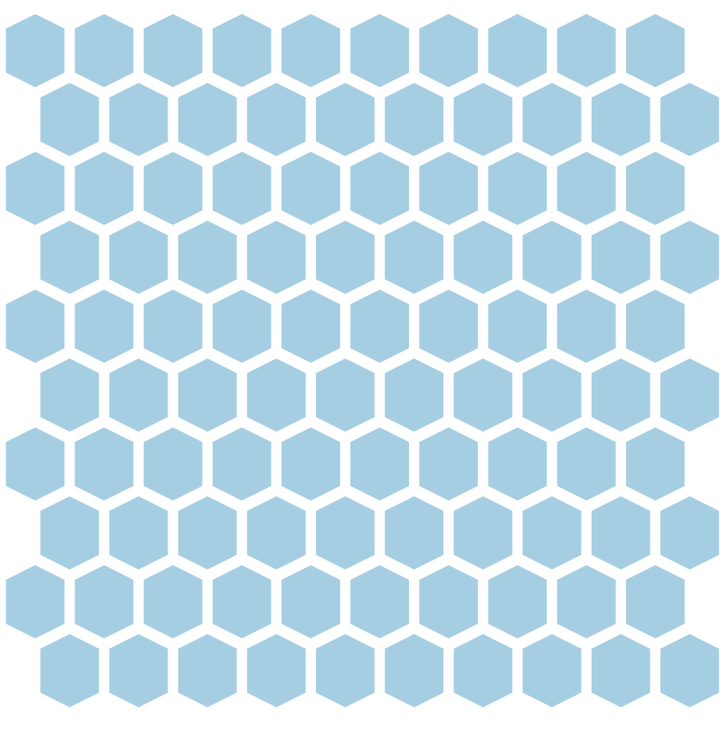
Gene	Cluster	Cytoband	Type
CTNNB1	1	3p22.1	SNV
KEAP1	1	19p13.2	SNV
SMAD4	?	18q21.2	SNV

R1



R2





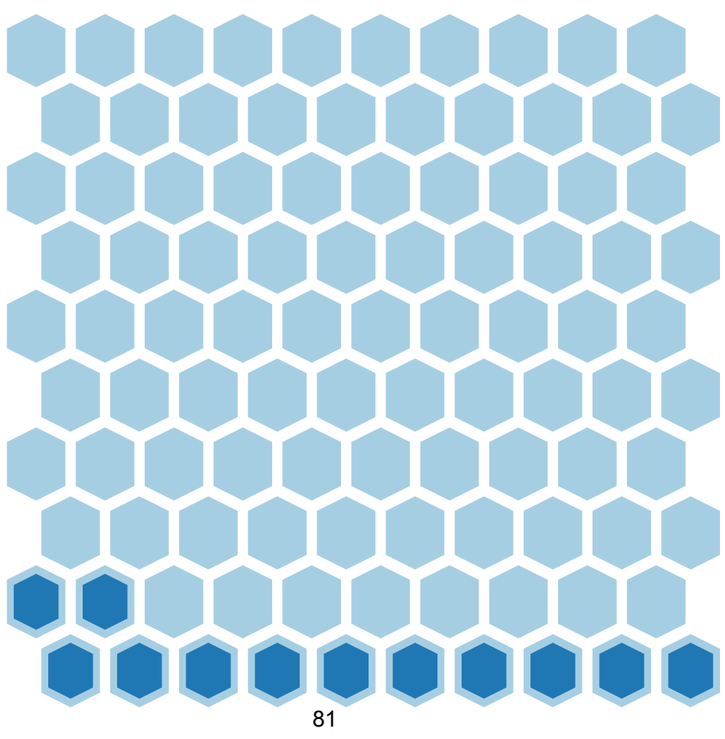
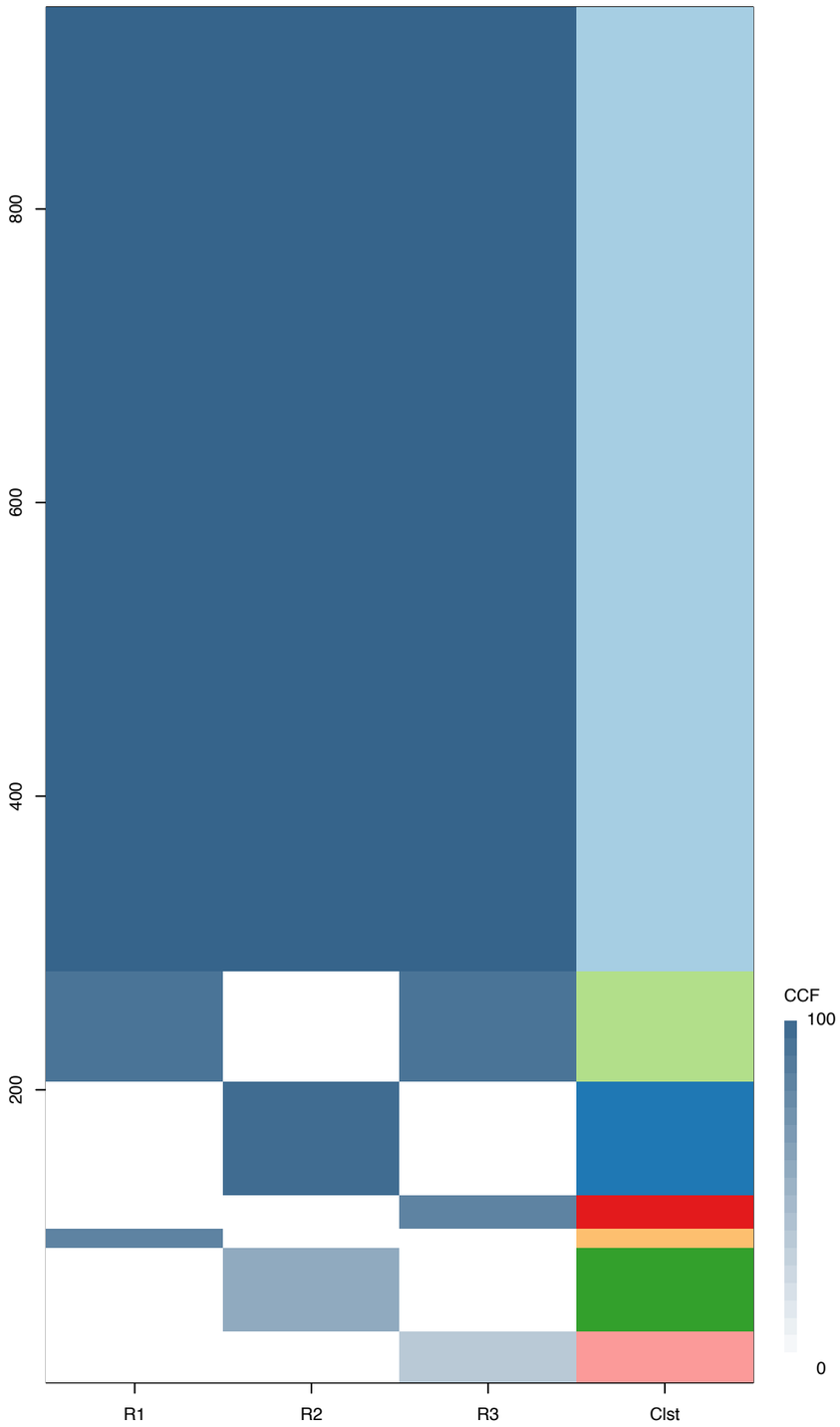
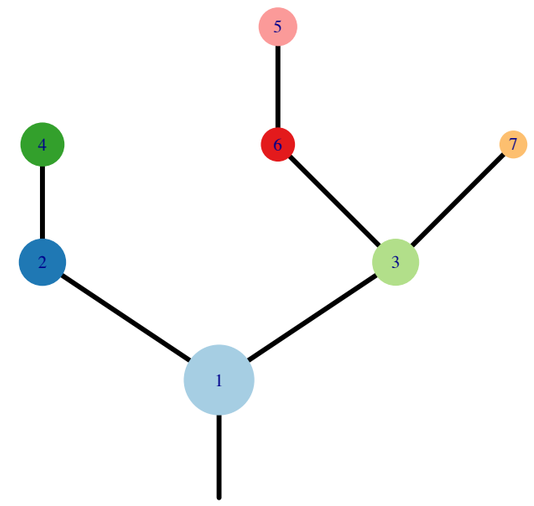


Fig.S12AH **CRUK0034**





Histology:Adeno, Age:70, PackYears:33, Size:40
Stage:1b, Gender:Female, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
PTCH1	1	9q22.32	SNV
ATM	1	11q22.3	SNV
KRAS	1	12p12.1	SNV
FOXA1	1	14q21.1	Amp
PLXNB2	1	22q13.33	SNV
MGA	4	15q15.1	SNV
NKX2-1	?	14q13.3	Amp
CIC	?	19q13.2	SNV

R1 **R2** **R3**

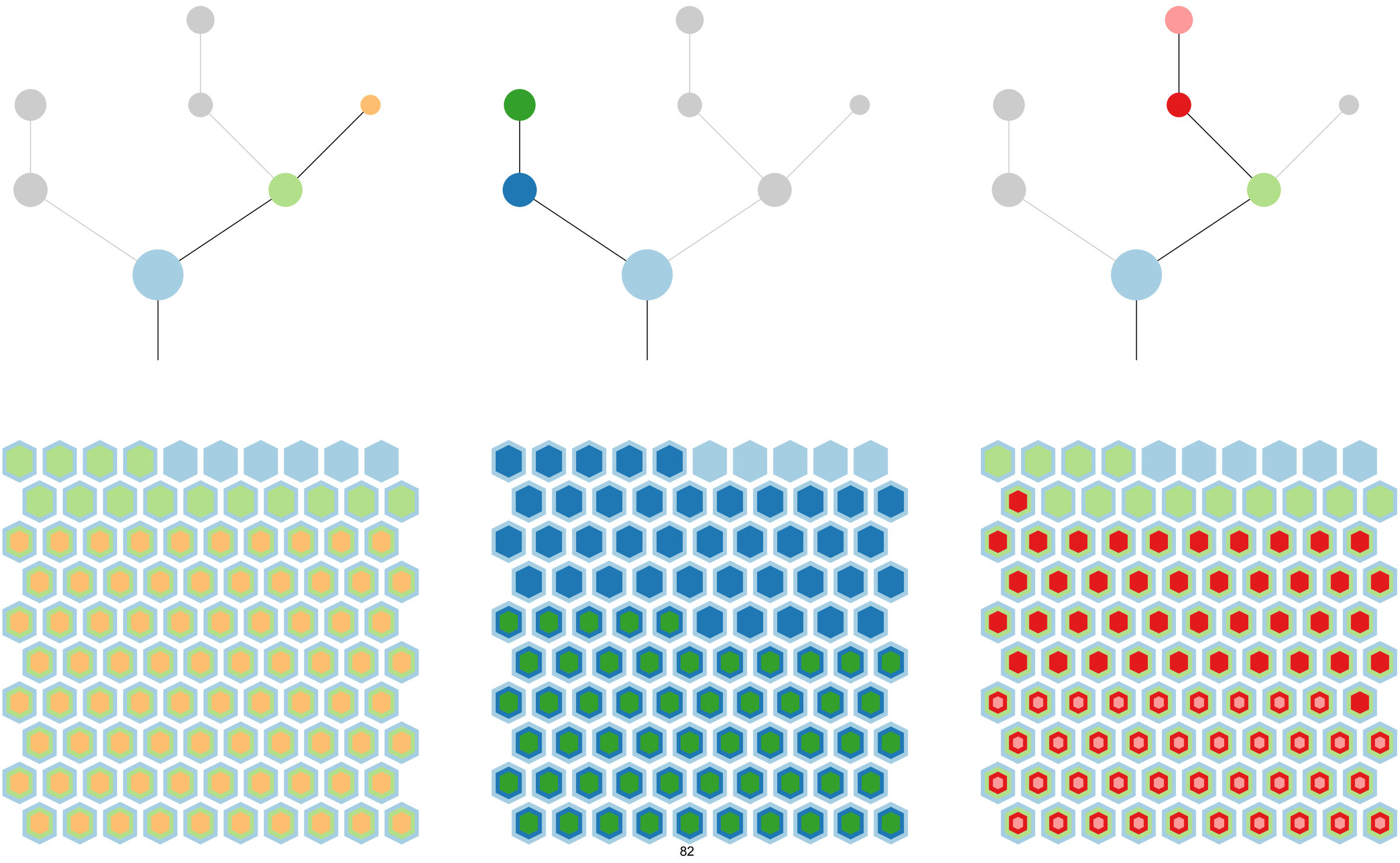
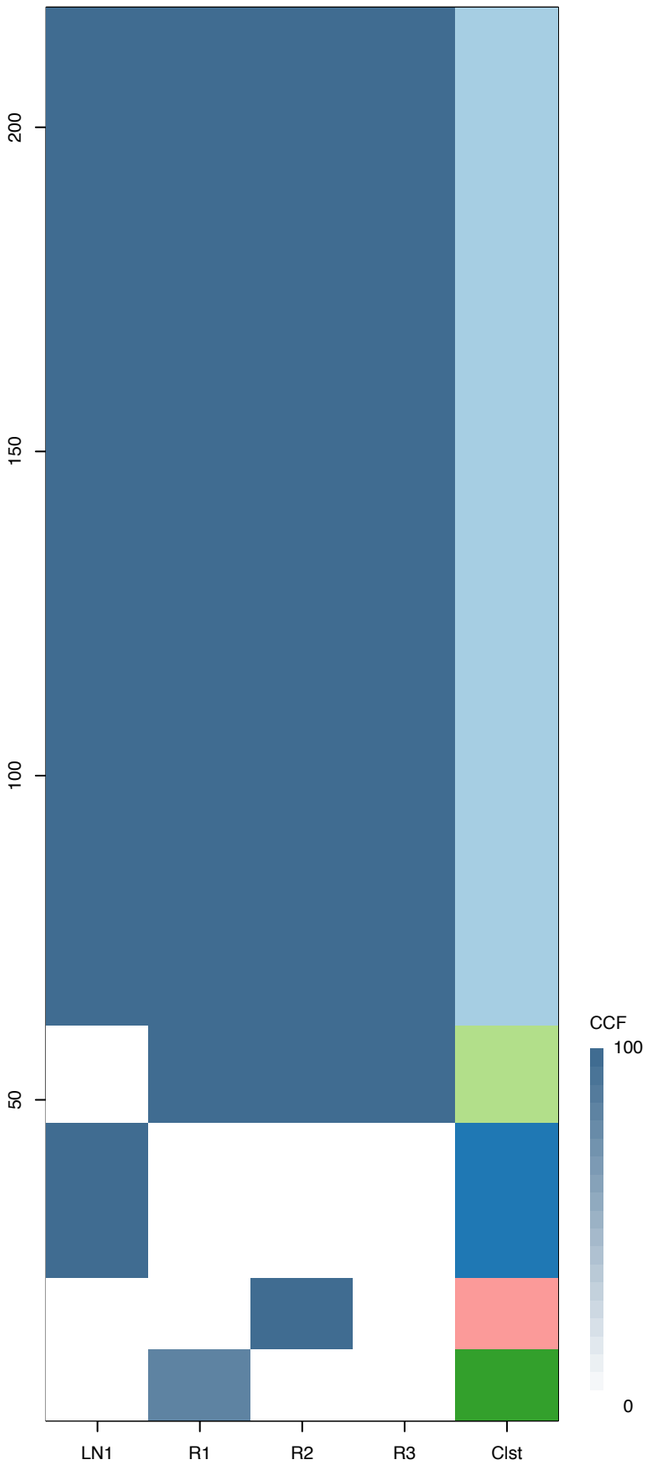
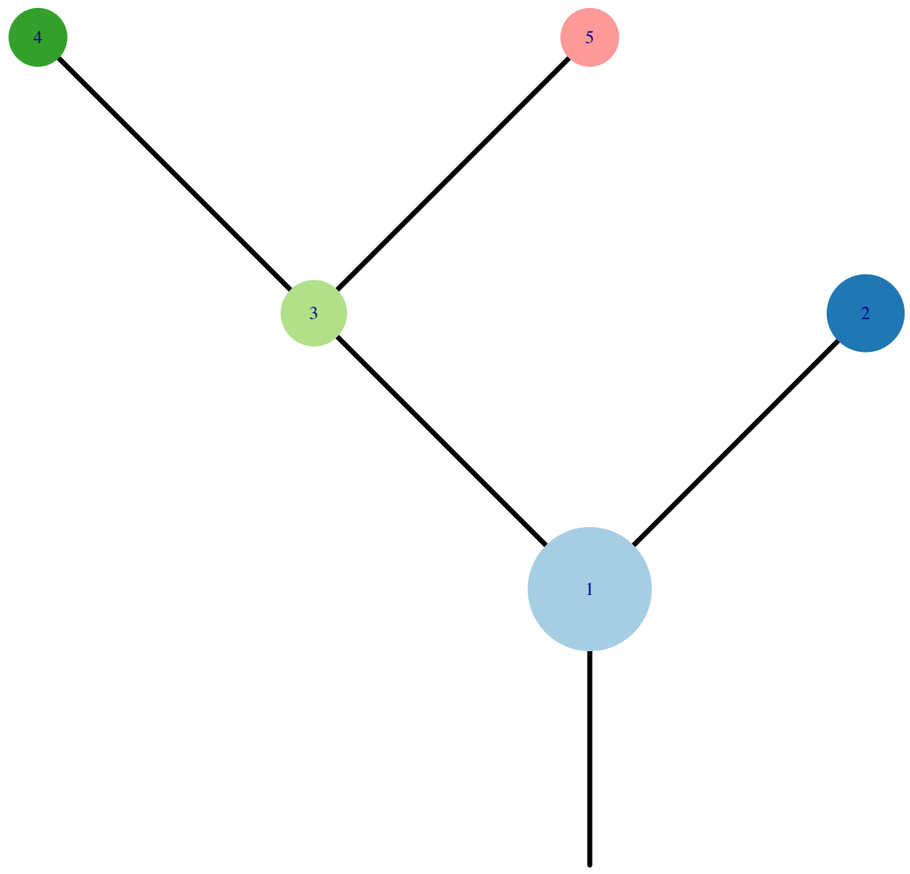


Fig.S12AI



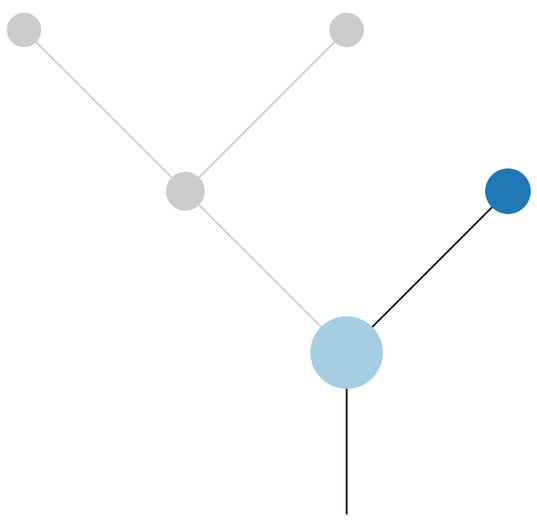
CRUK0035



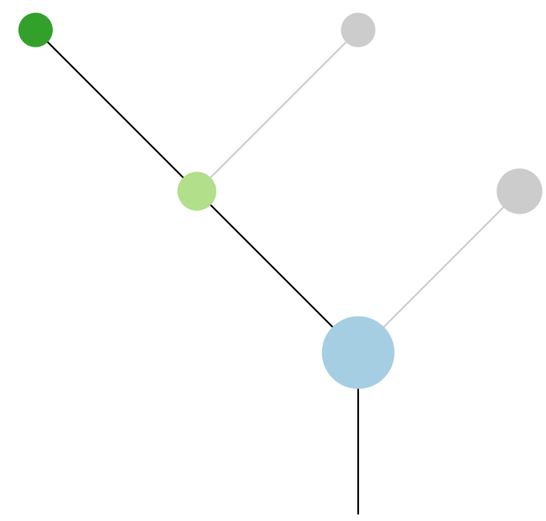
Histology:Adeno, Age:77, PackYears:28, Size:28
Stage:3b, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
ETV1	1	7p21.2	Amp
HNRNPA2B1	1	7p15.2	Amp
HOXA9	1	7p15.2	Amp
HOXA11	1	7p15.2	Amp
HOXA13	1	7p15.2	Amp
JAZF1	1	7p15.2	Amp
FAS	1	10q23.31	SNV
TP53	1	17p13.1	SNV
ZNF180	1	19q13.31	Amp
BCL3	1	19q13.32	Amp
SS18L1	1	20q13.33	Amp
FLT4	4	5q35.3	SNV
EML4	?	2p21	Amp

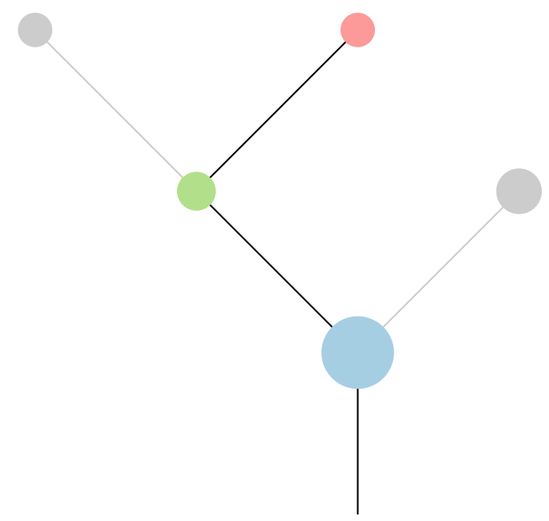
LN1



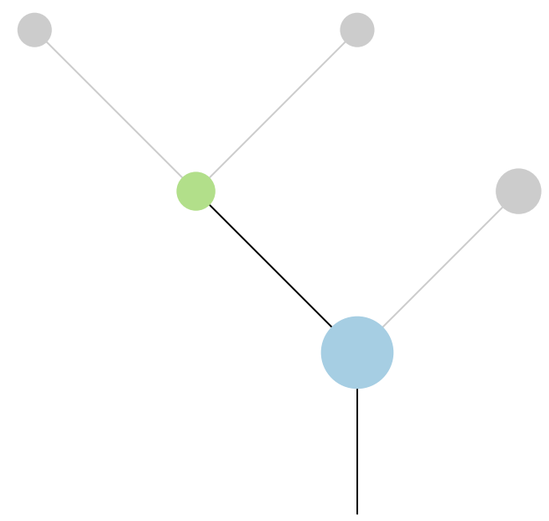
R1



R2



R3



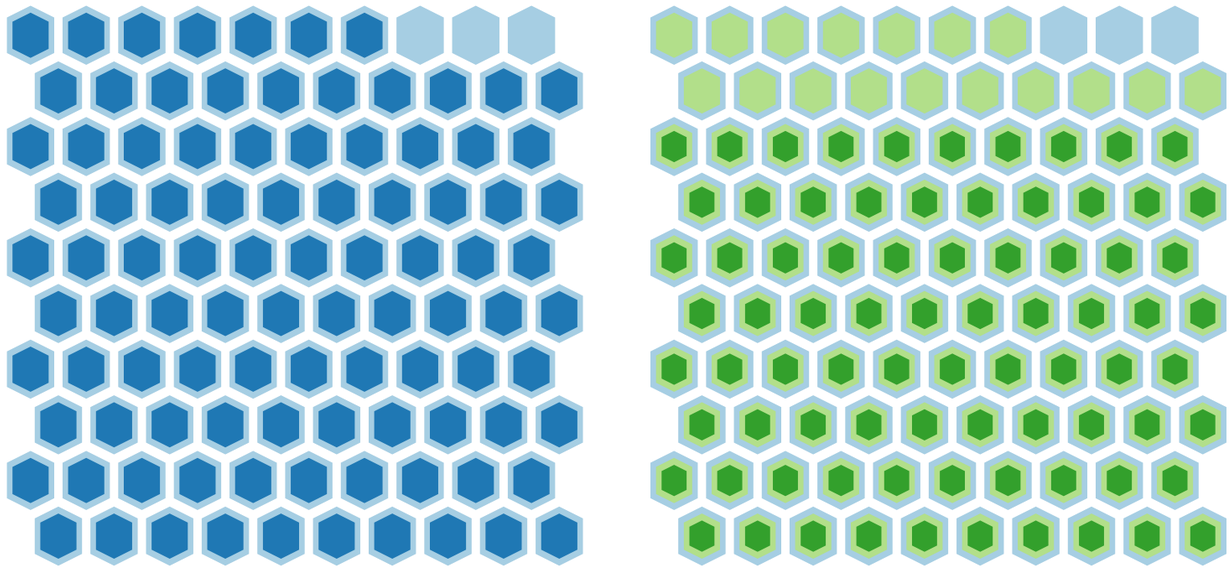
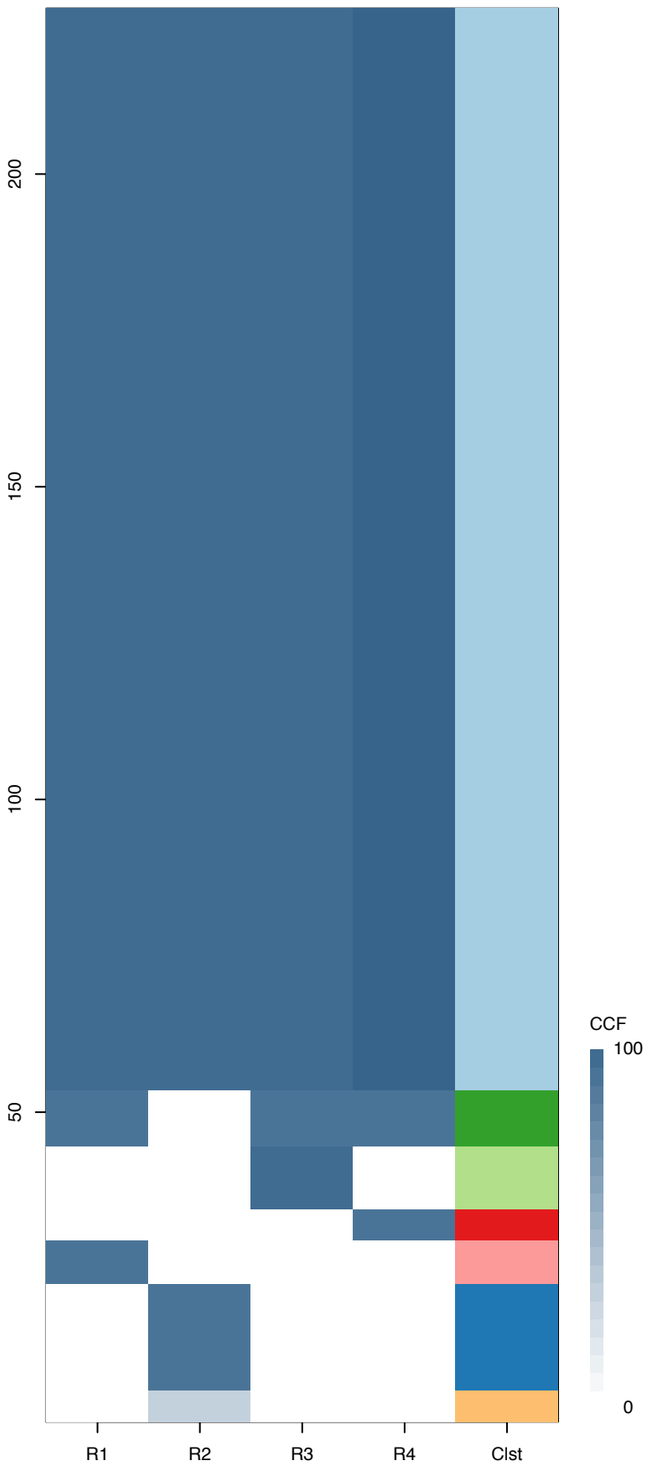
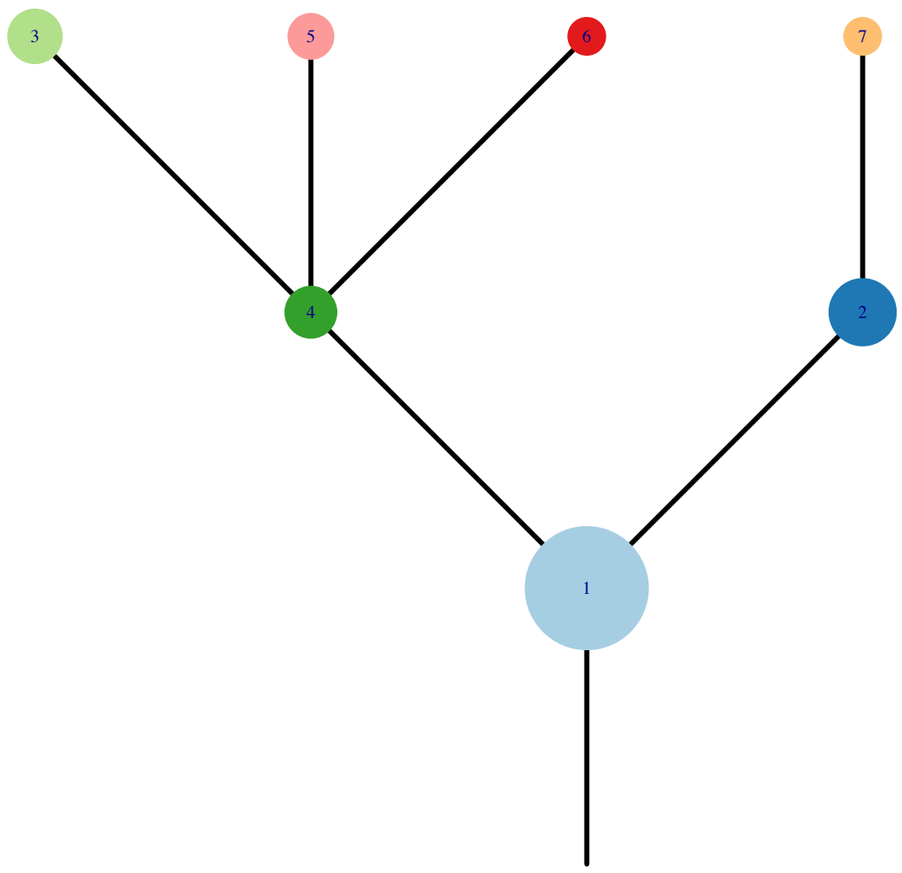


Fig.S12AJ



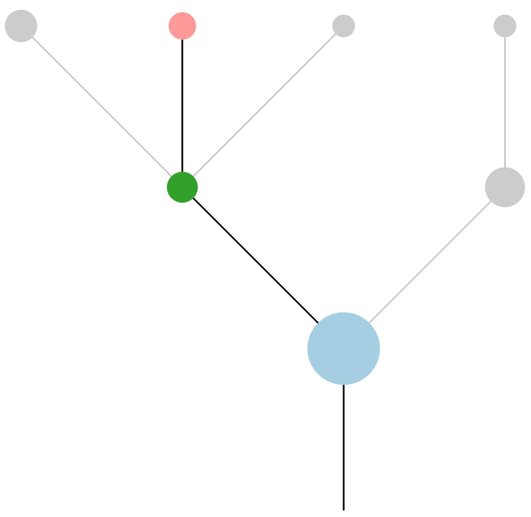
CRUK0036

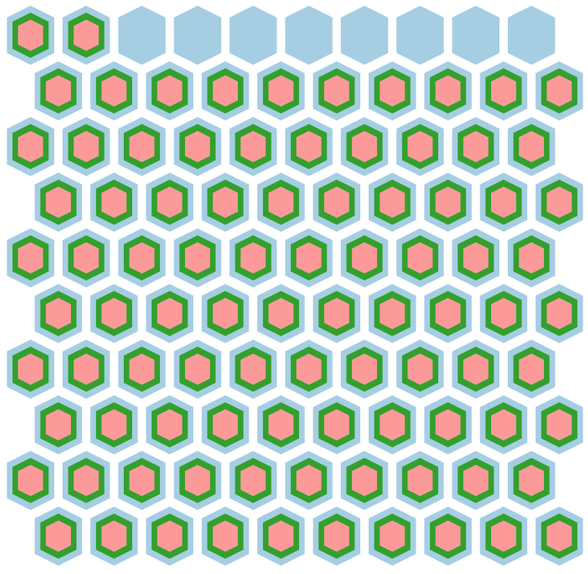


Histology:Adeno, Age:65, PackYears:33.75, Size:40
Stage:1b, Gender:Female, GD:Clonal GD, Recur:yes

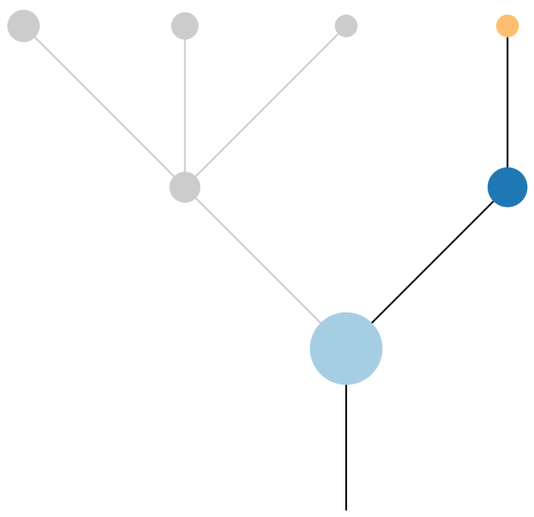
Gene	Cluster	Cytoband	Type
BCL9	1	1q21.2	Amp
ARNT	1	1q21.3	Amp
SETDB1	1	1q21.3	Amp
MLLT11	1	1q21.3	Amp
TPM3	1	1q21.3	Amp
MUC1	1	1q22	Amp
RIT1	1	1q22	Amp
LMNA	1	1q22	Amp
PRCC	1	1q23.1	Amp
NTRK1	1	1q23.1	Amp
FCRL4	1	1q23.1	Amp
FCGR2B	1	1q23.3	Amp
PBX1	1	1q23.3	Amp
PIK3CA	1	3q26.32	SNV
TERT	1	5p15.33	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
VTI1A	1	10q25.2	Amp
TCF7L2	1	10q25.2	Amp
ETNK1	1	12p12.1	Amp
KRAS	1	12p12.1	Amp
KRAS	1	12p12.1	SNV
PPFIBP1	1	12p11.23	Amp
KEAP1	1	19p13.2	SNV
ARHGAP35	1	19q13.32	SNV
COX6C	2	8q22.2	Amp
SMARCA4	6	19p13.2	SNV

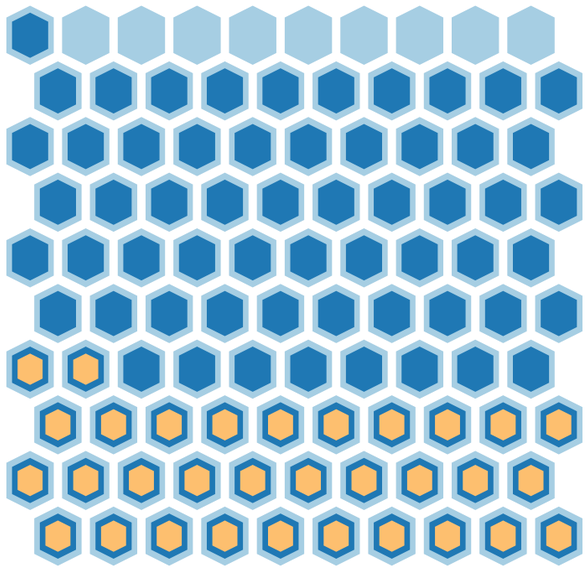
R1



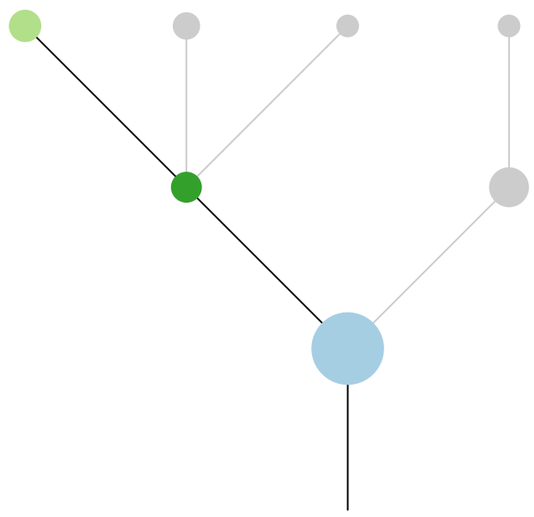


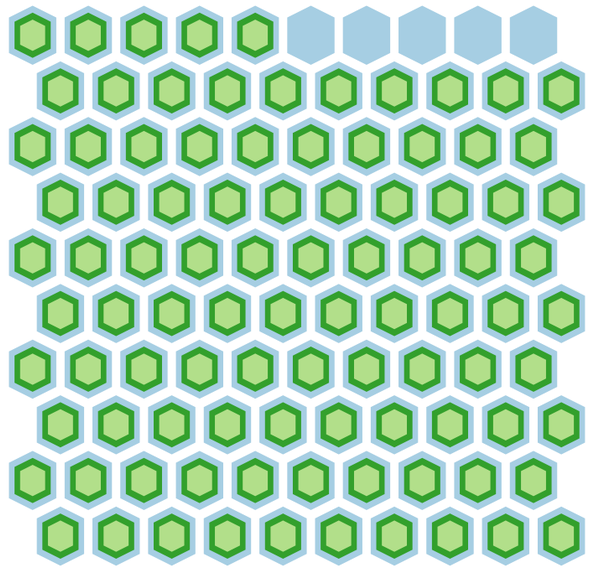
R2



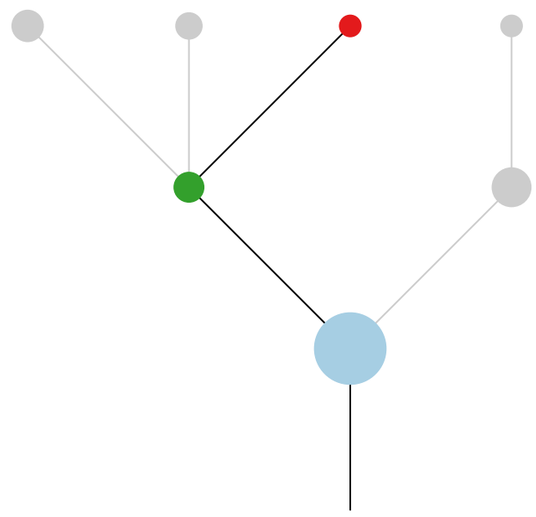


R3





R4



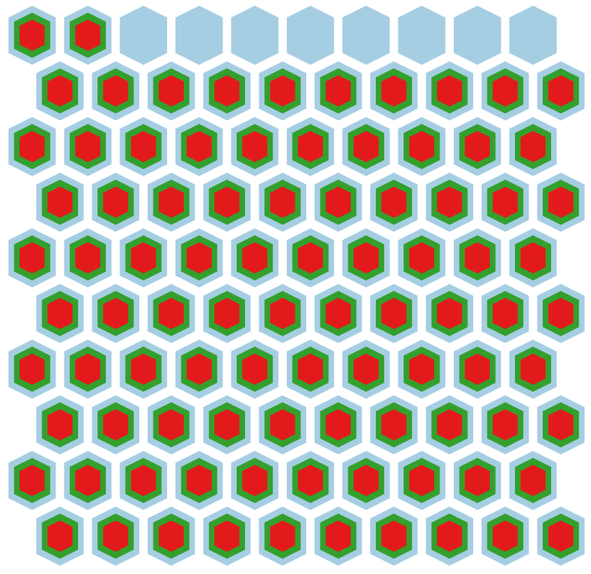
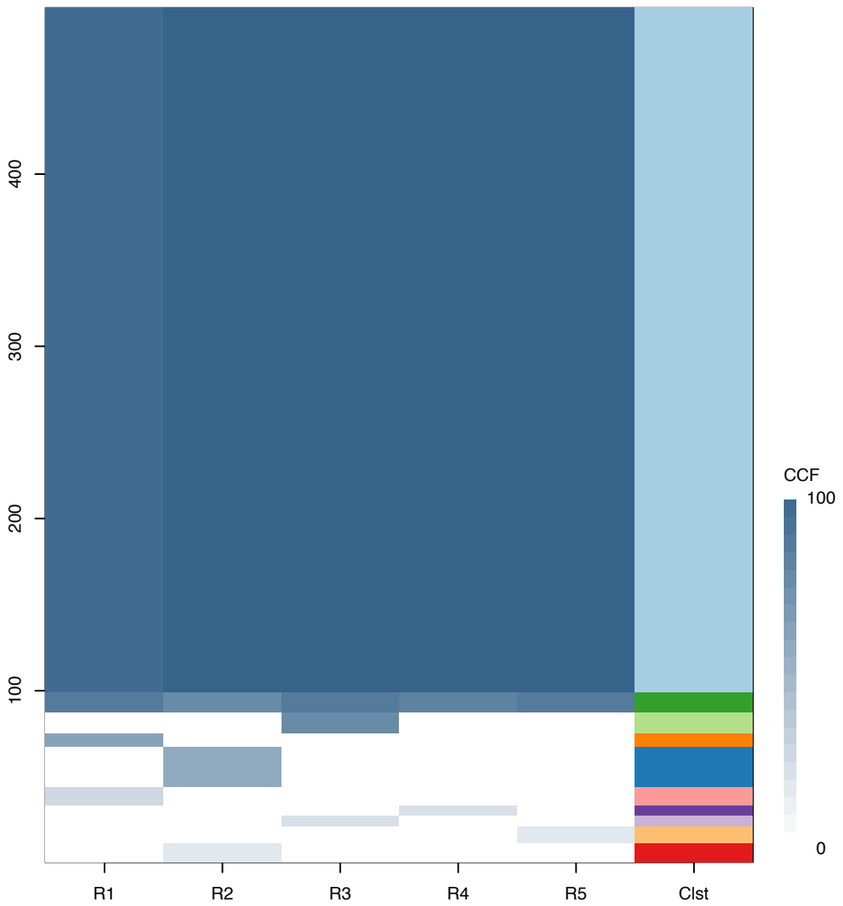
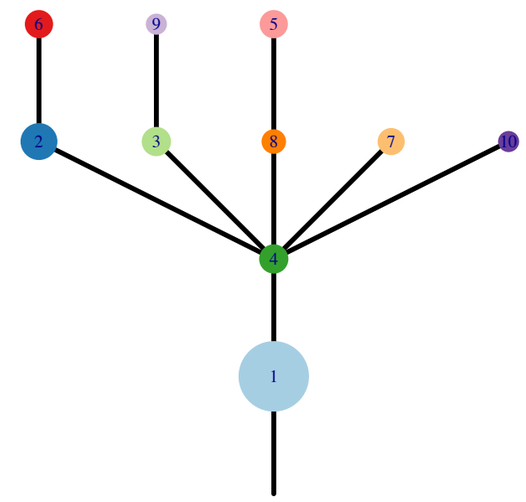


Fig.S12AK



CRUK0037_A



Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

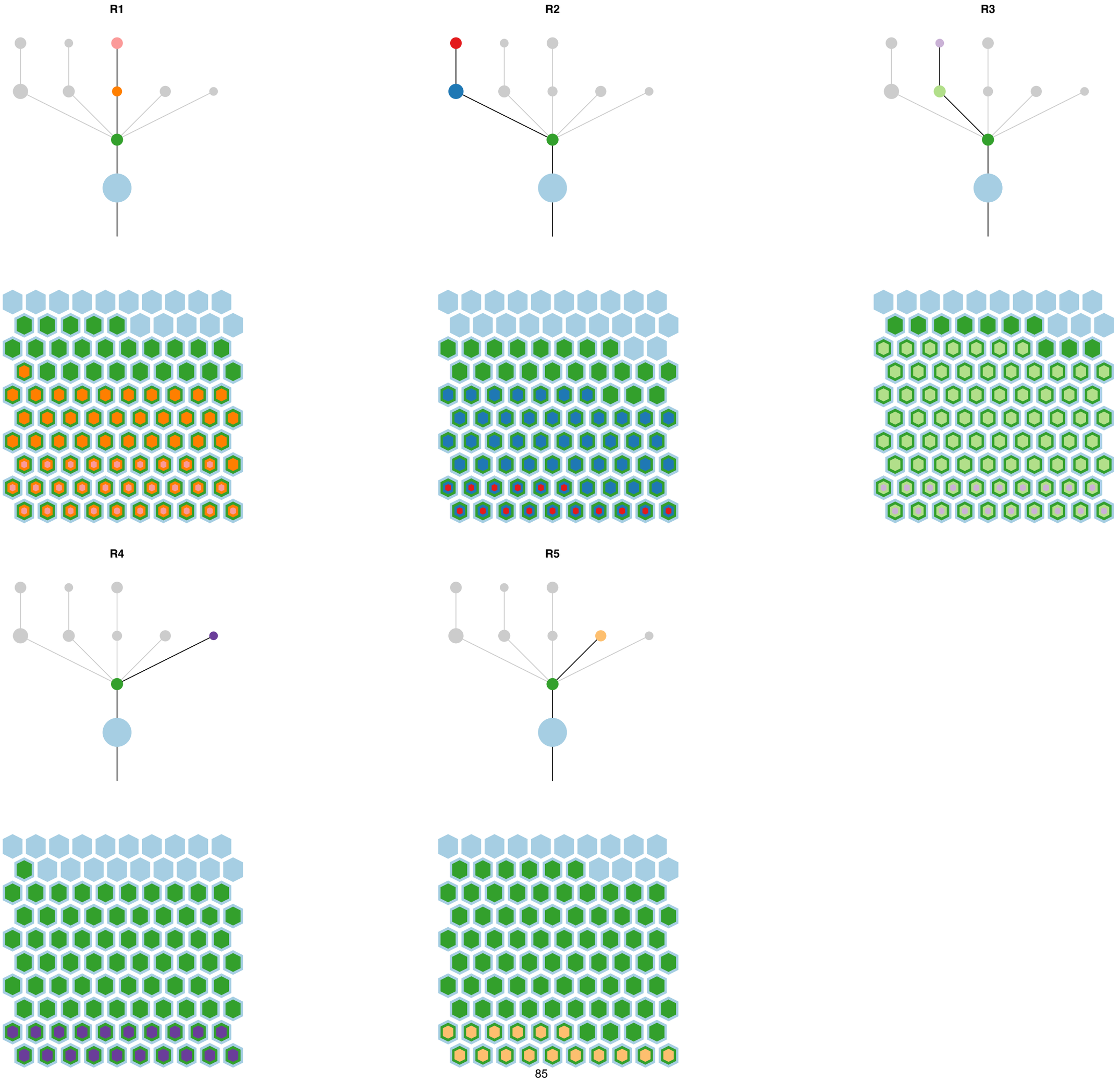
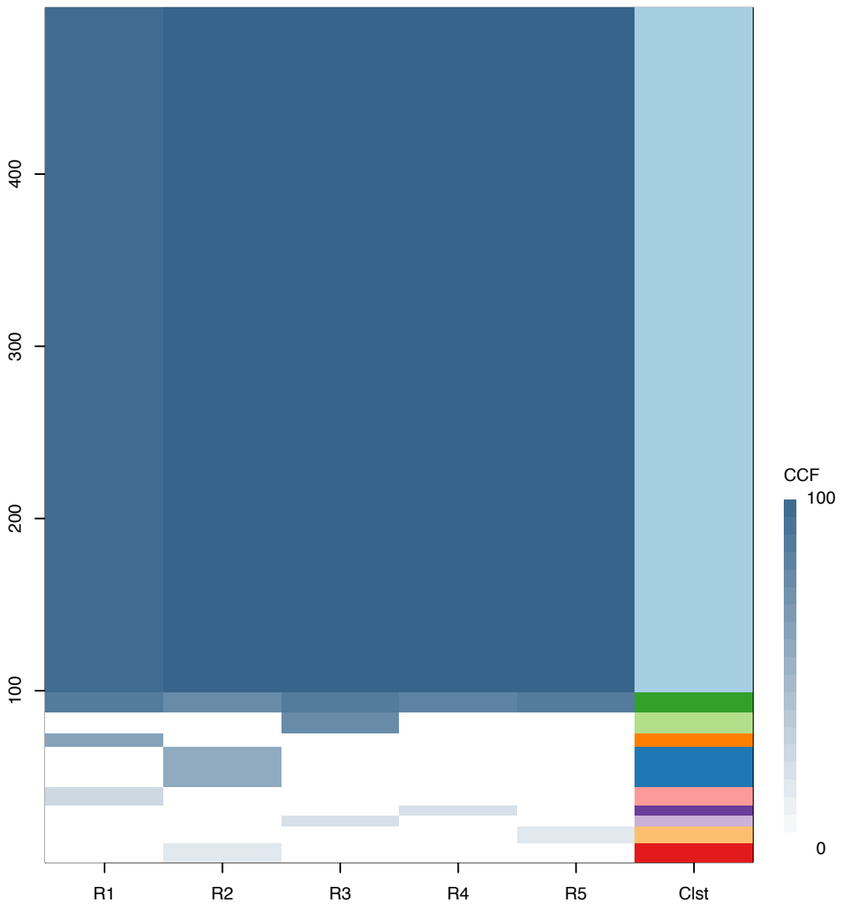
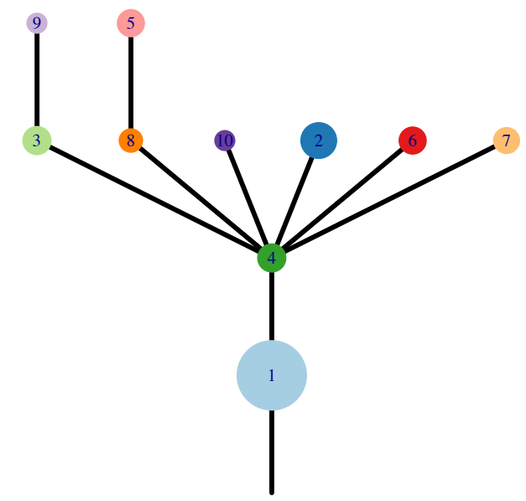


Fig.S12AK



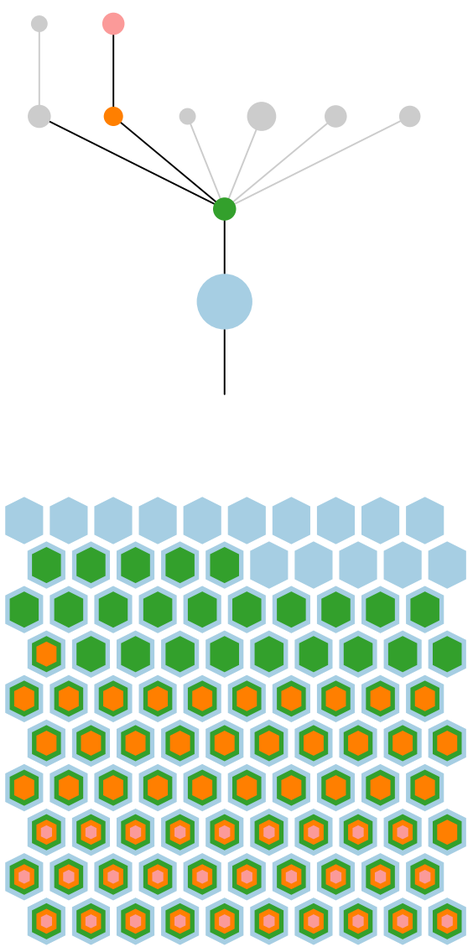
CRUK0037_B



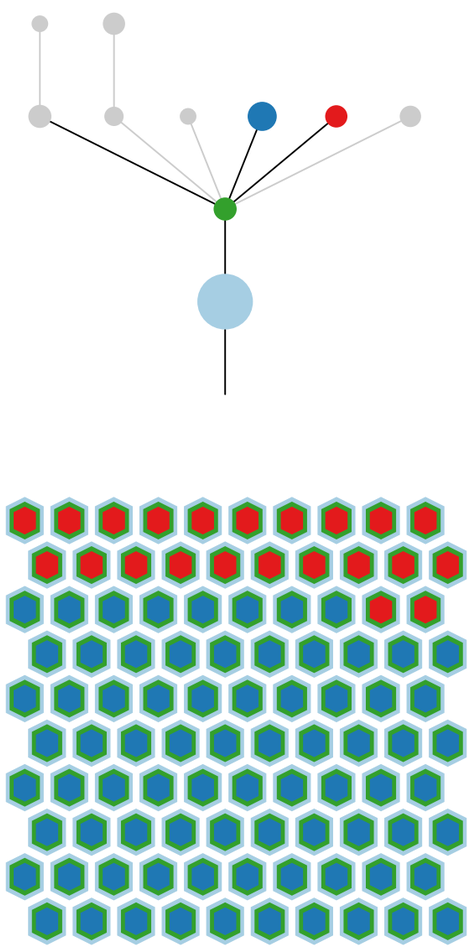
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

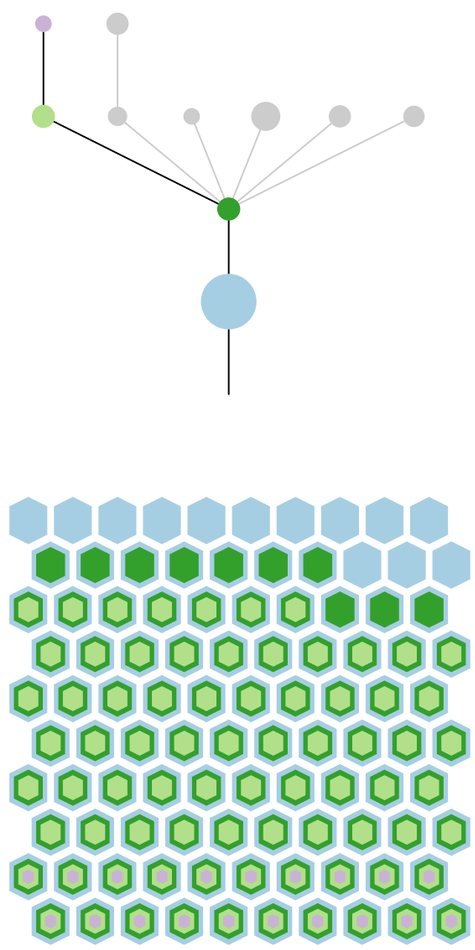
R1



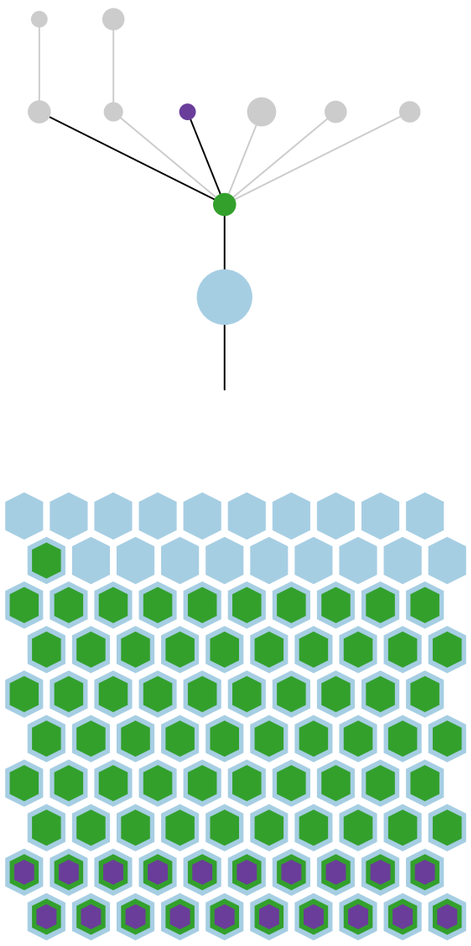
R2



R3



R4



R5

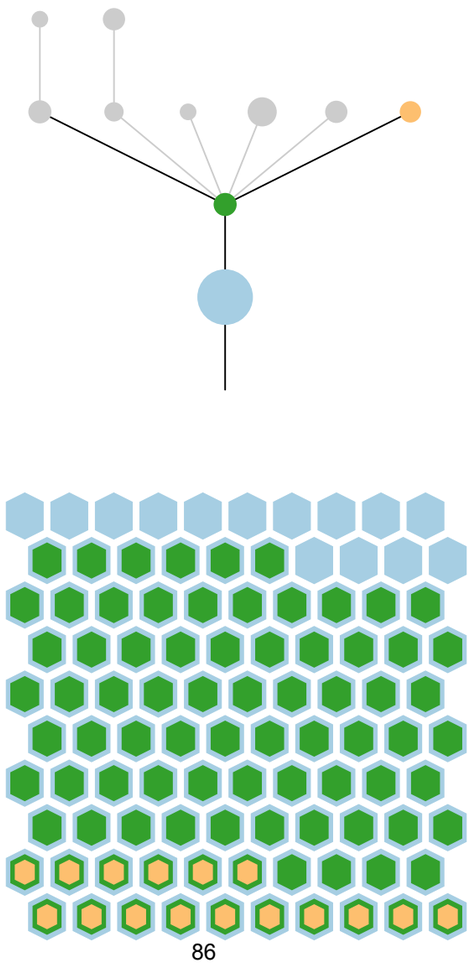
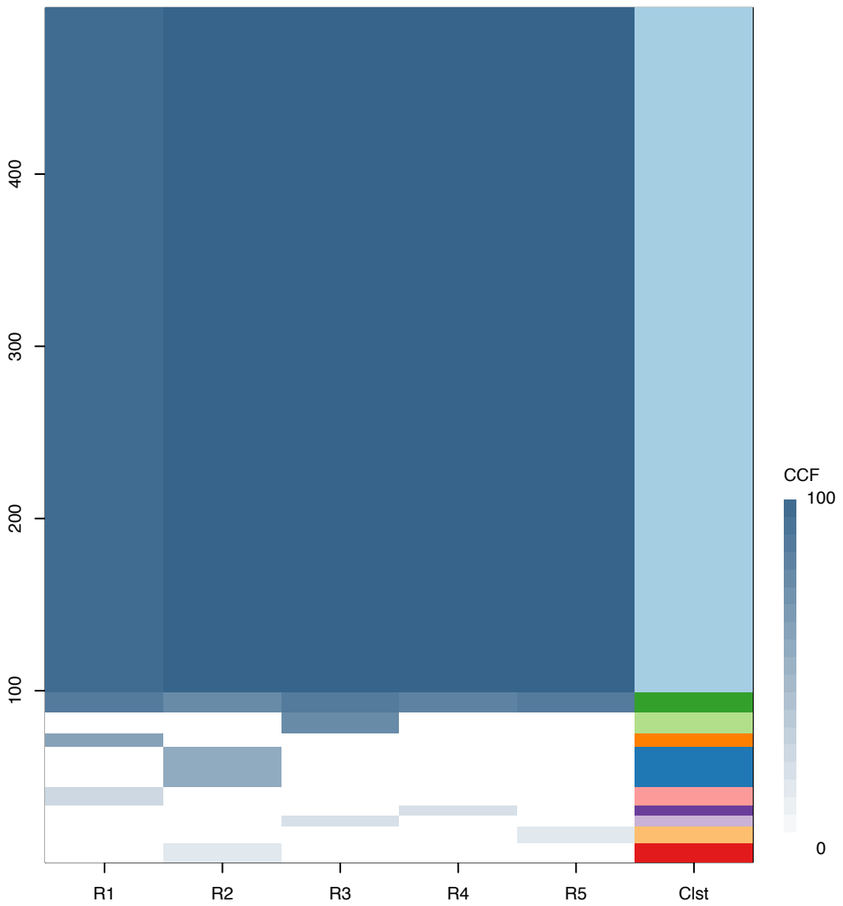
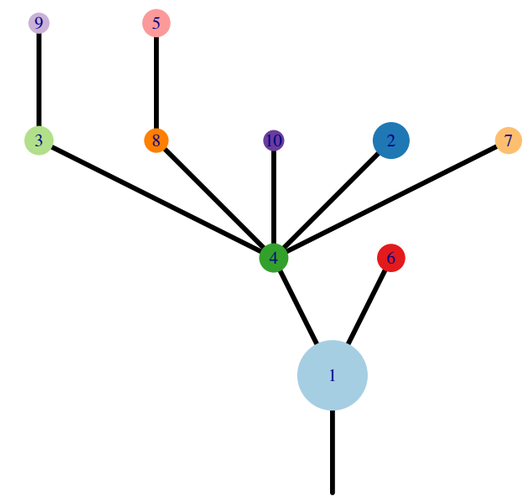


Fig.S12AK



CRUK0037_C



Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

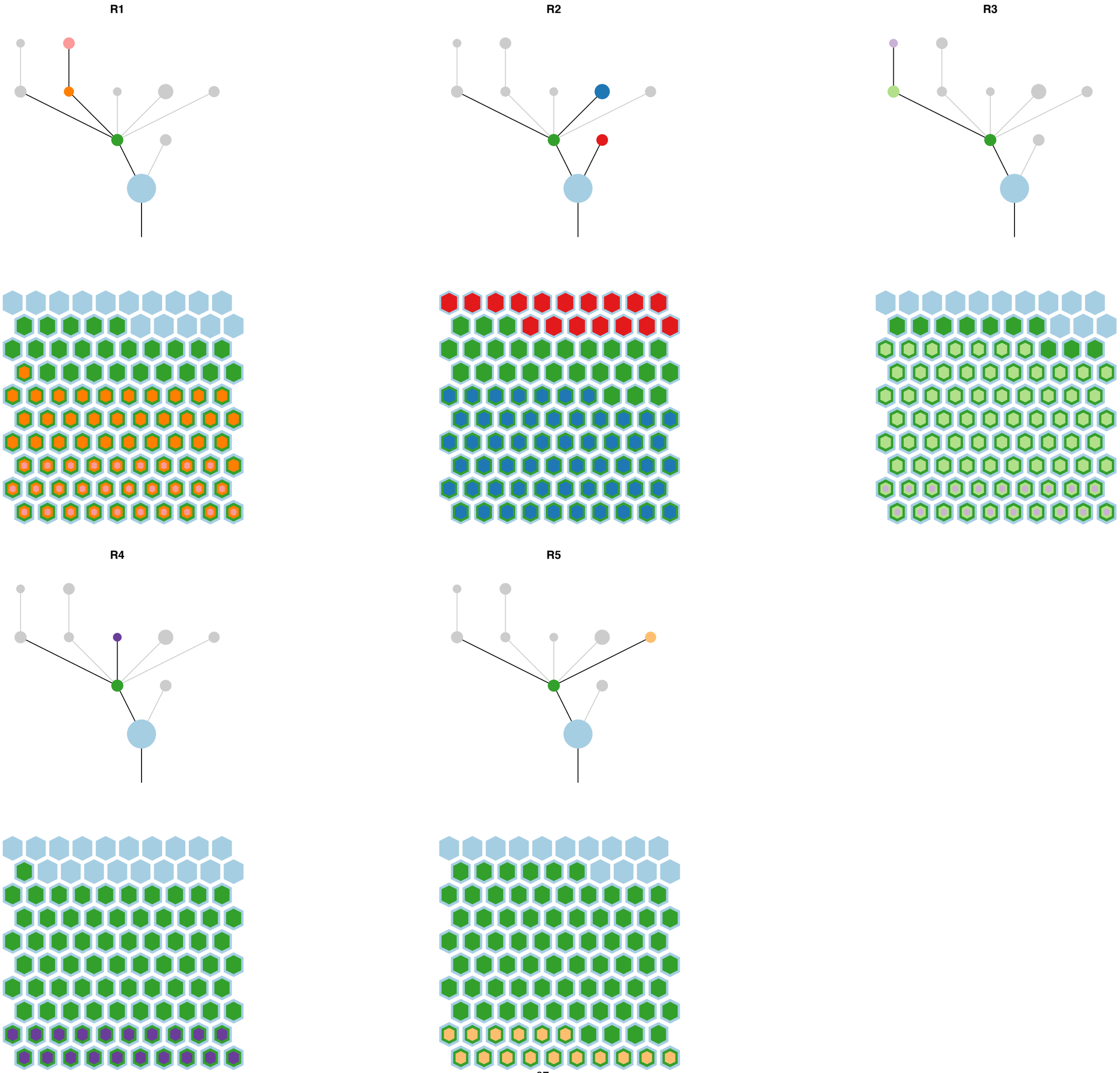
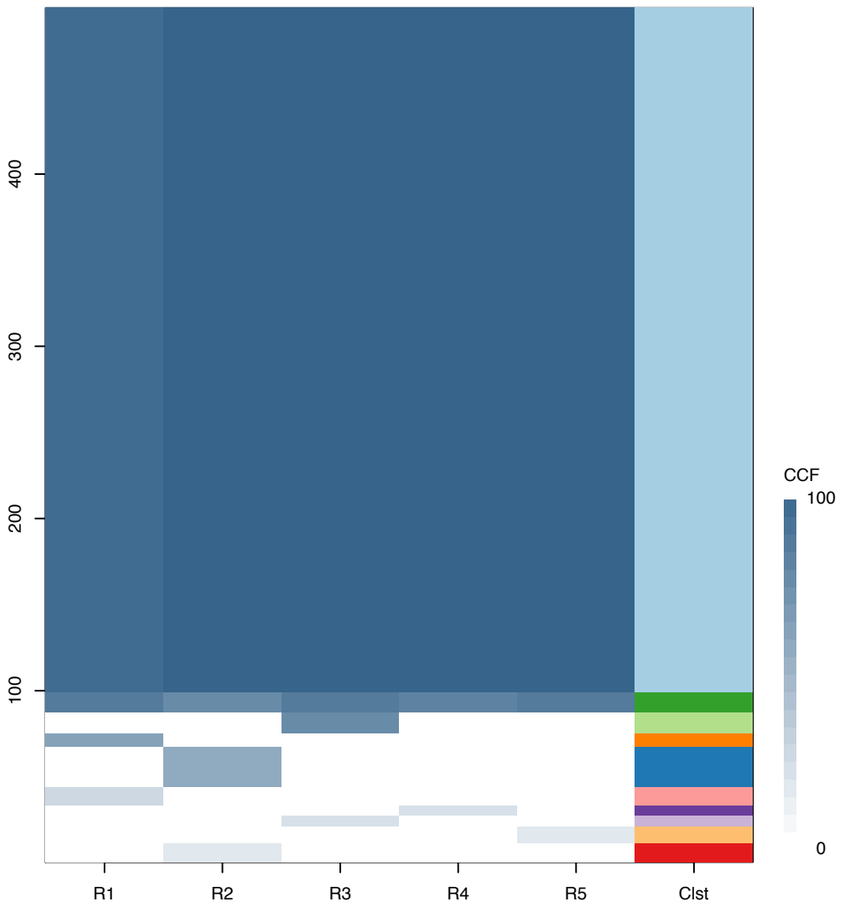
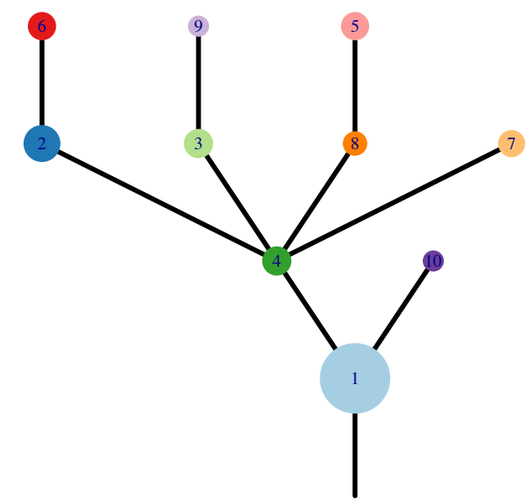


Fig.S12AK



CRUK0037_D



Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

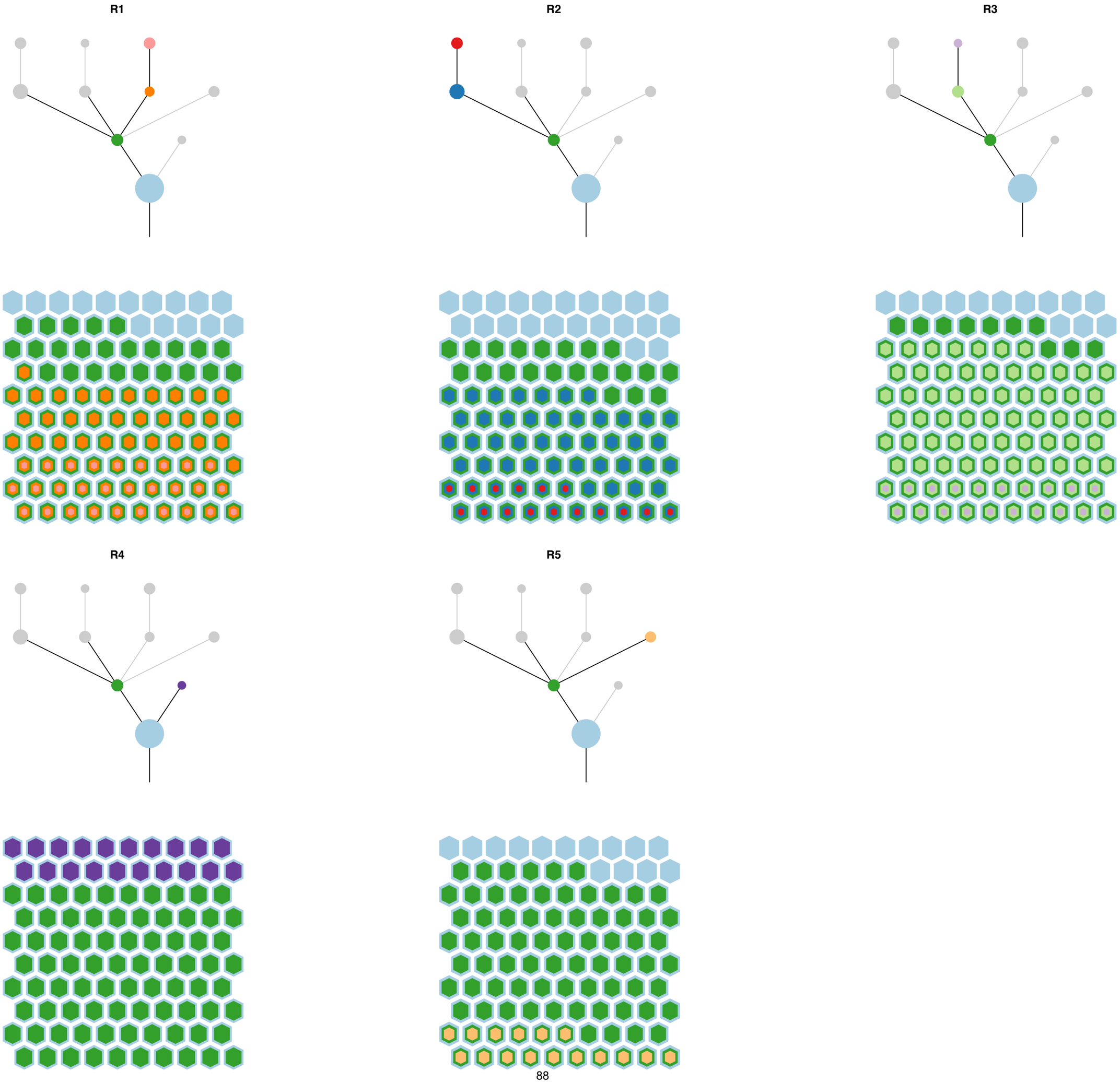
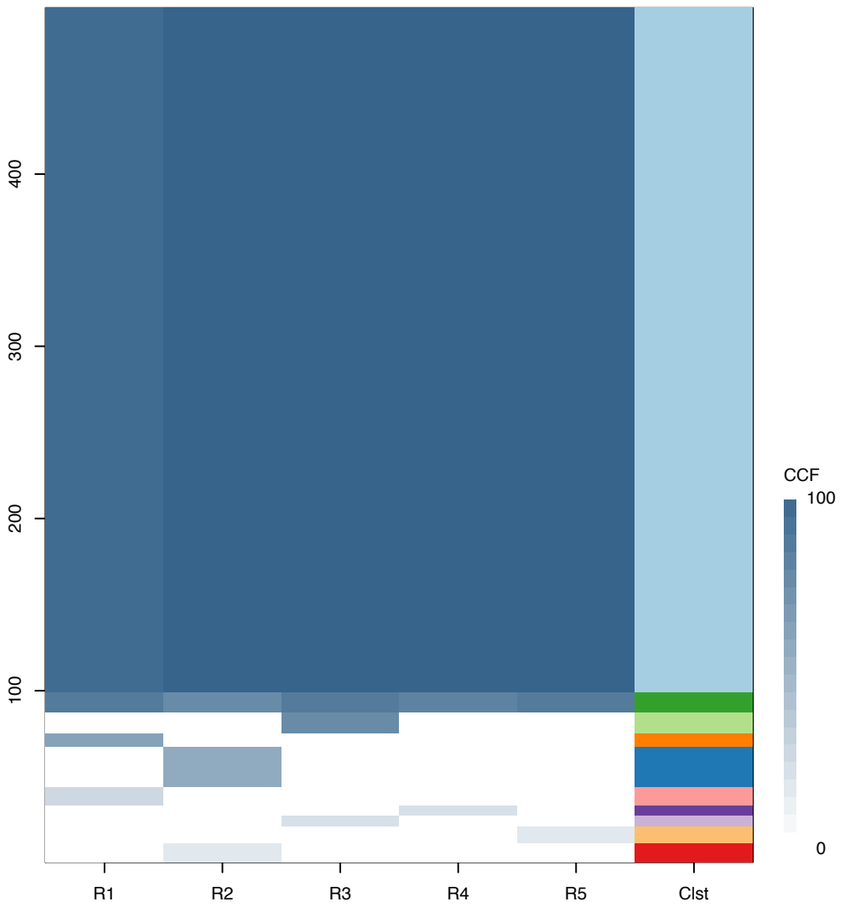
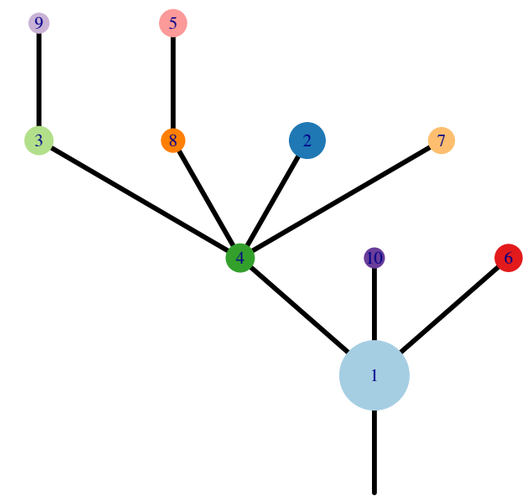


Fig.S12AK



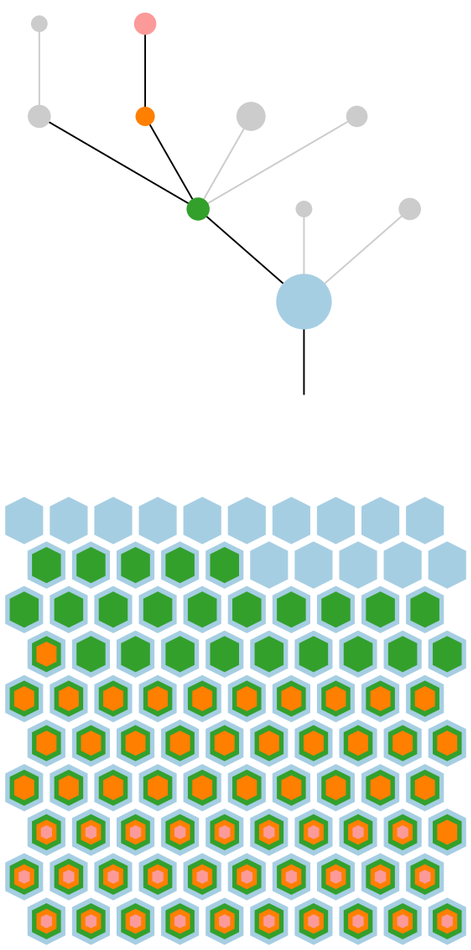
CRUK0037_E



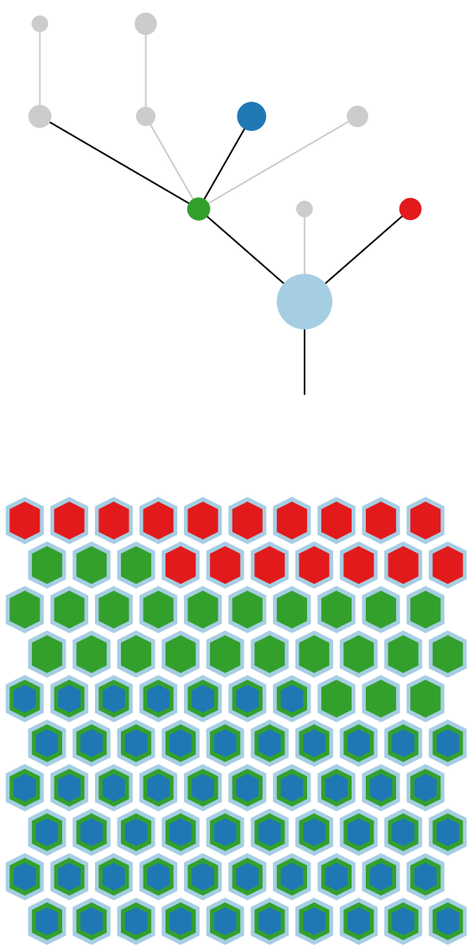
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

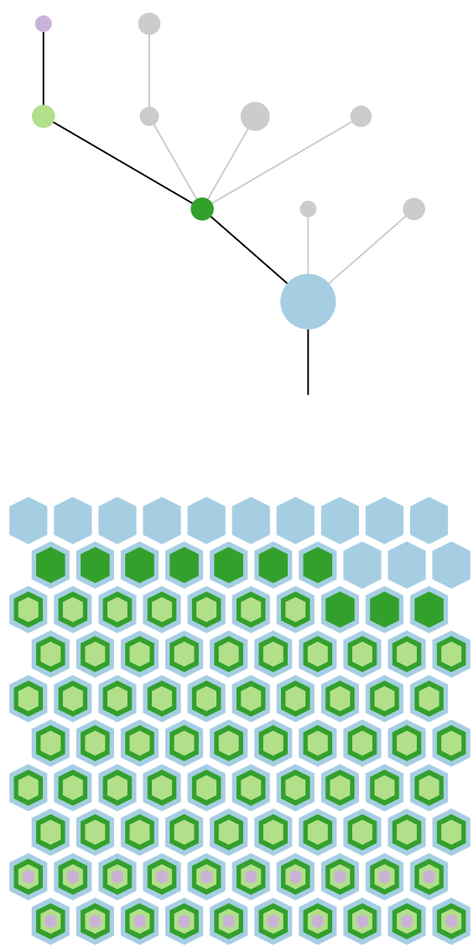
R1



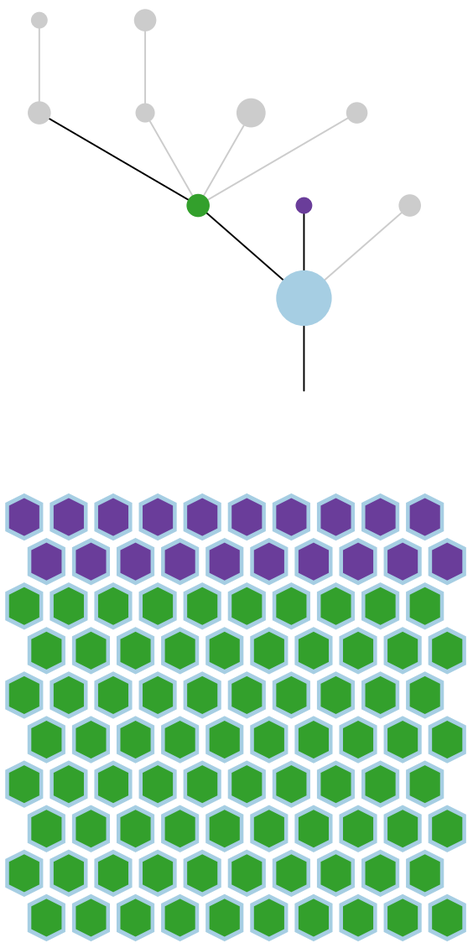
R2



R3



R4



R5

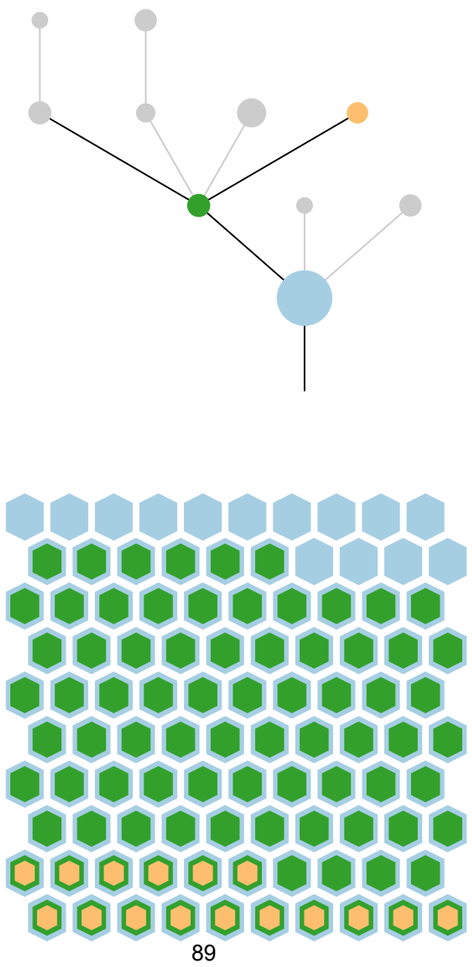
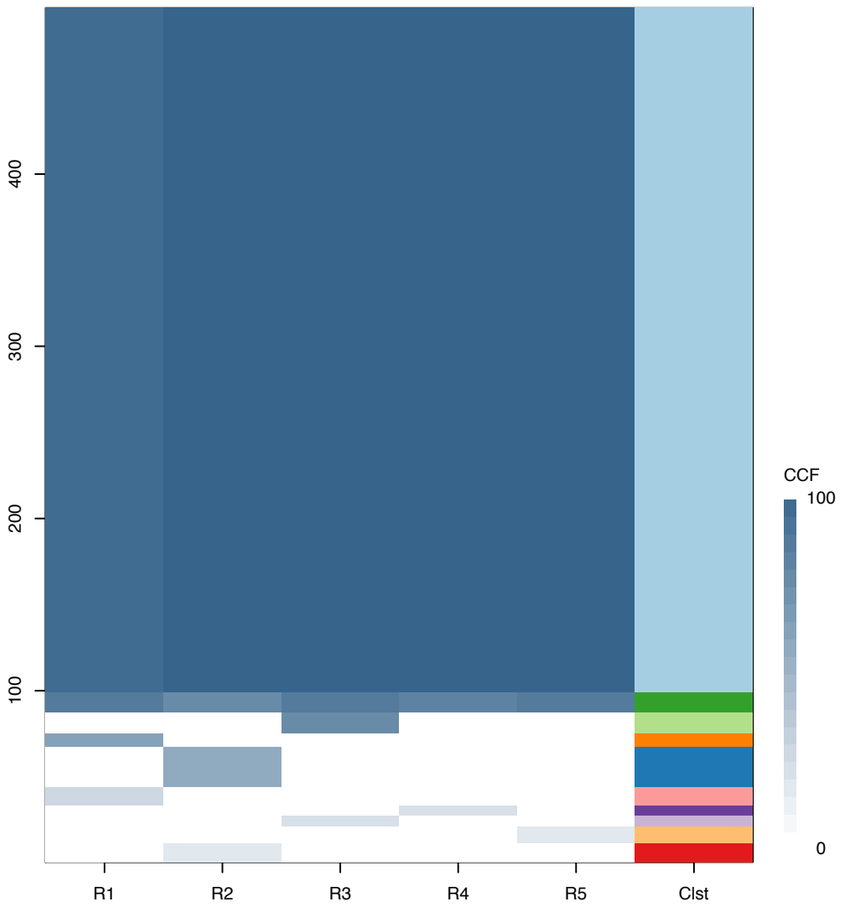
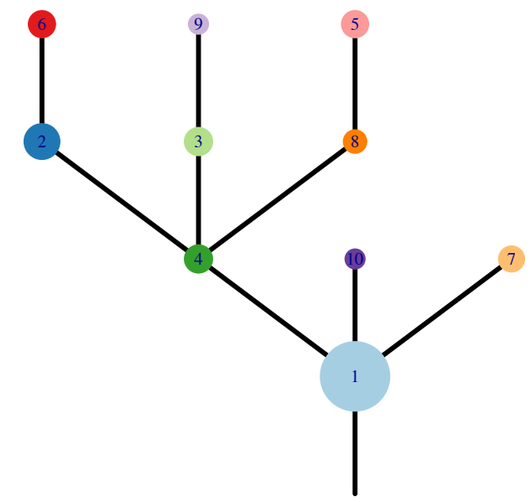


Fig.S12AK



CRUK0037_F



Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

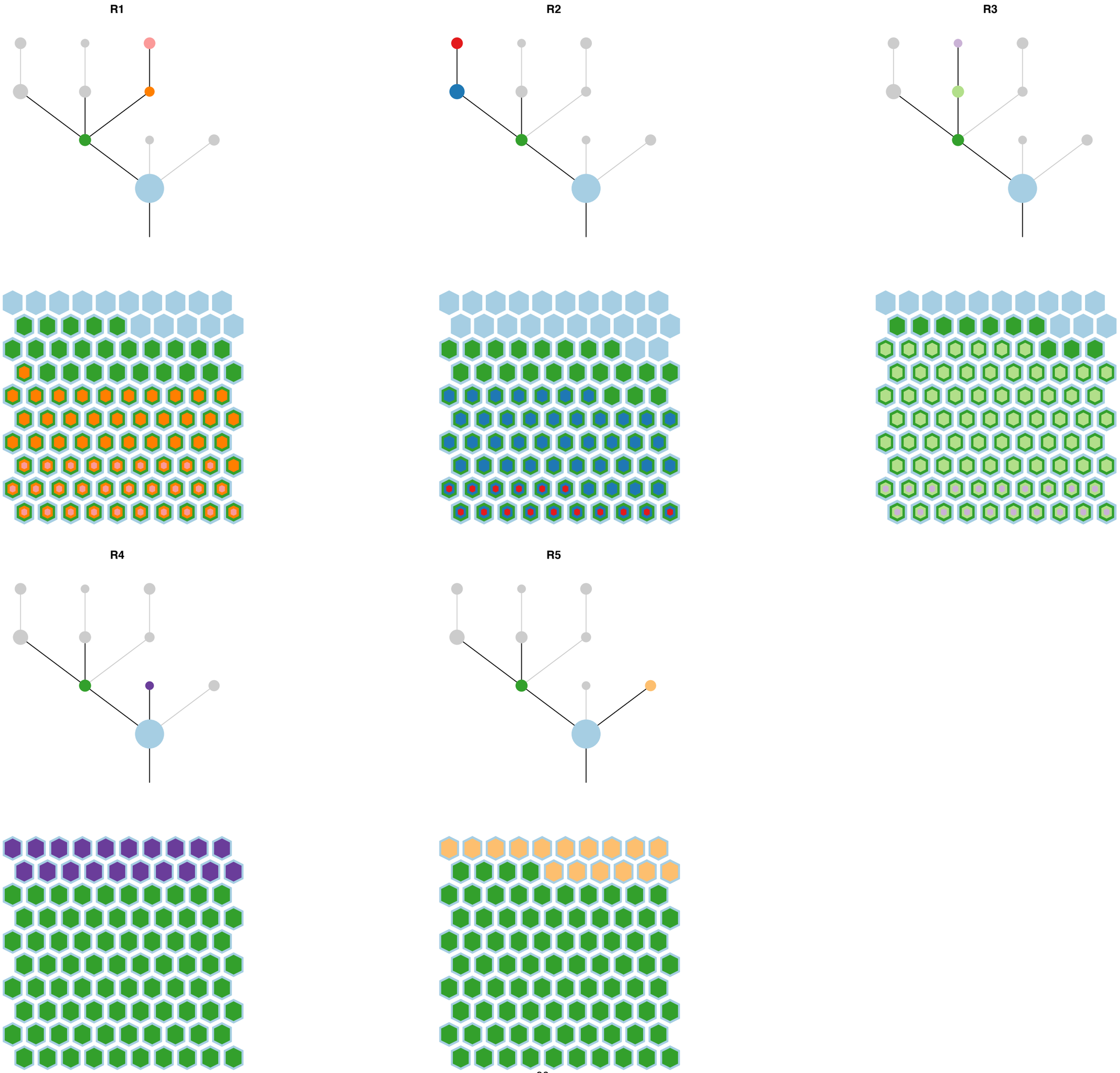
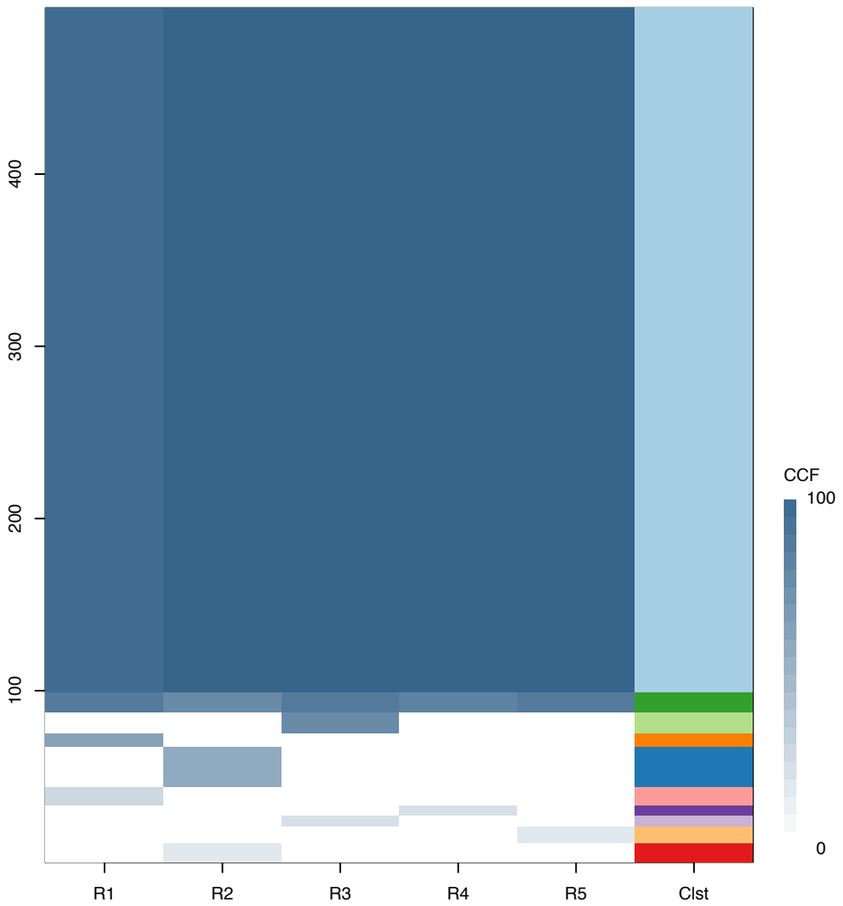
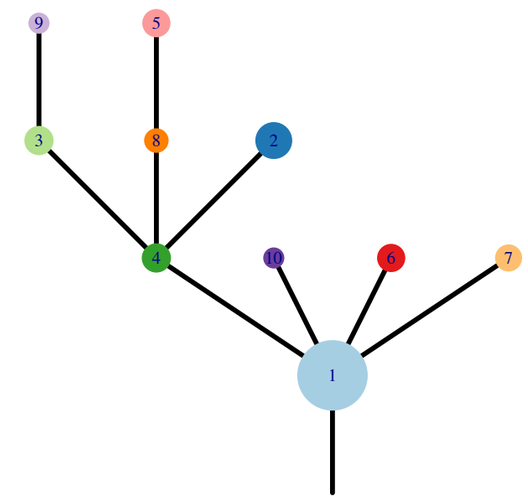


Fig.S12AK



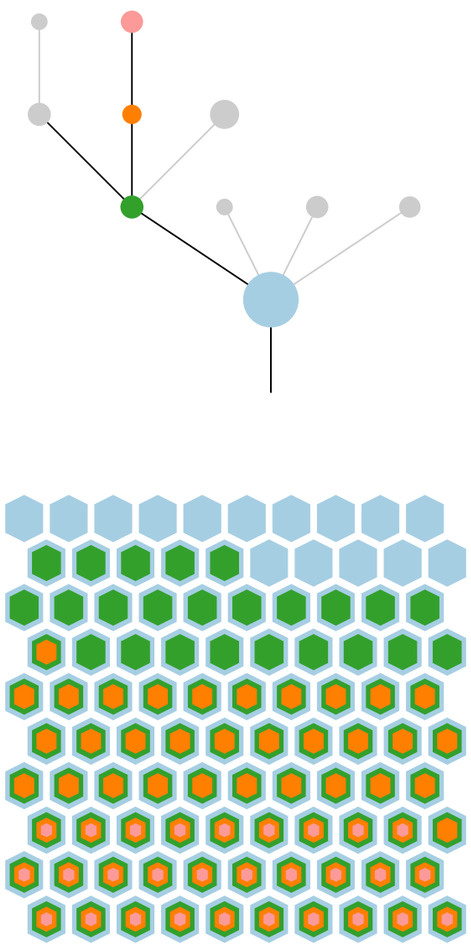
CRUK0037_G



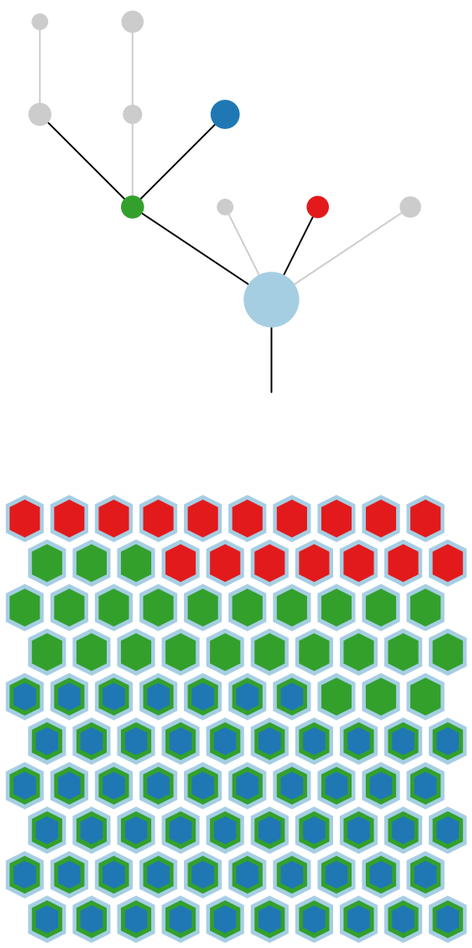
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

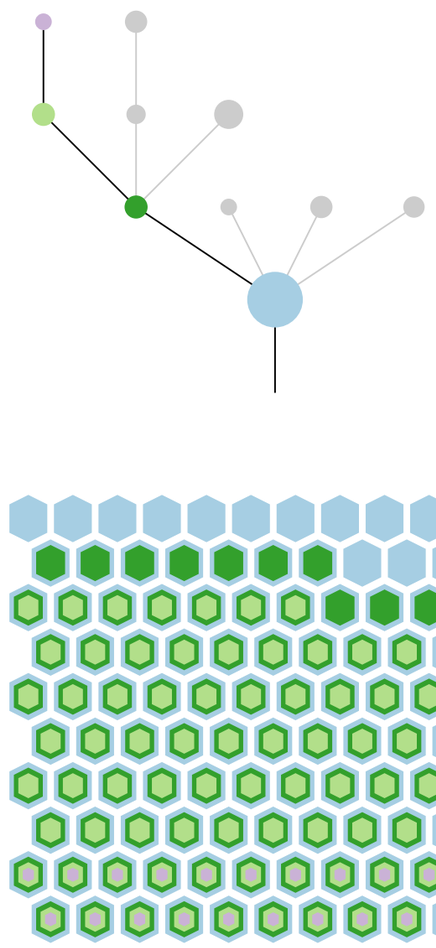
R1



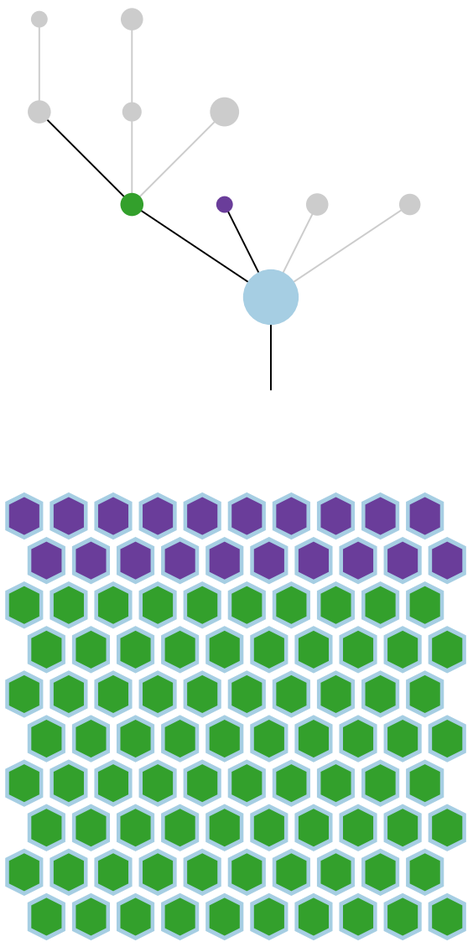
R2



R3



R4



R5

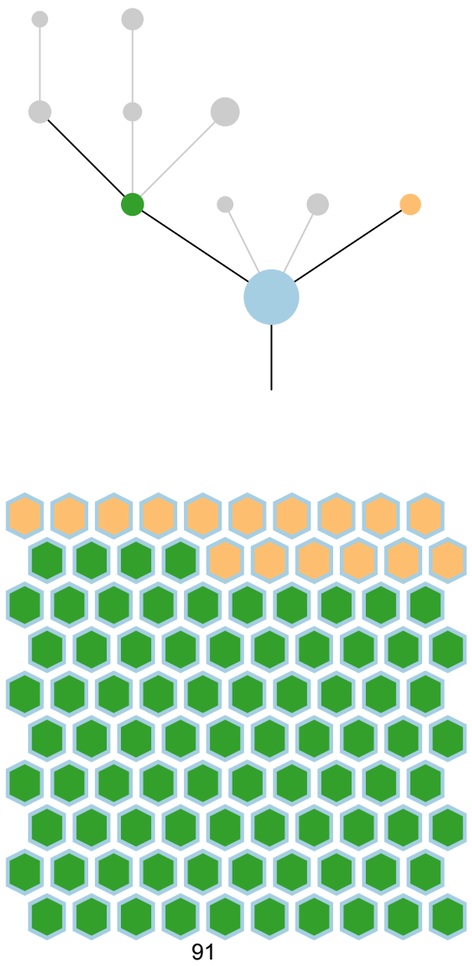
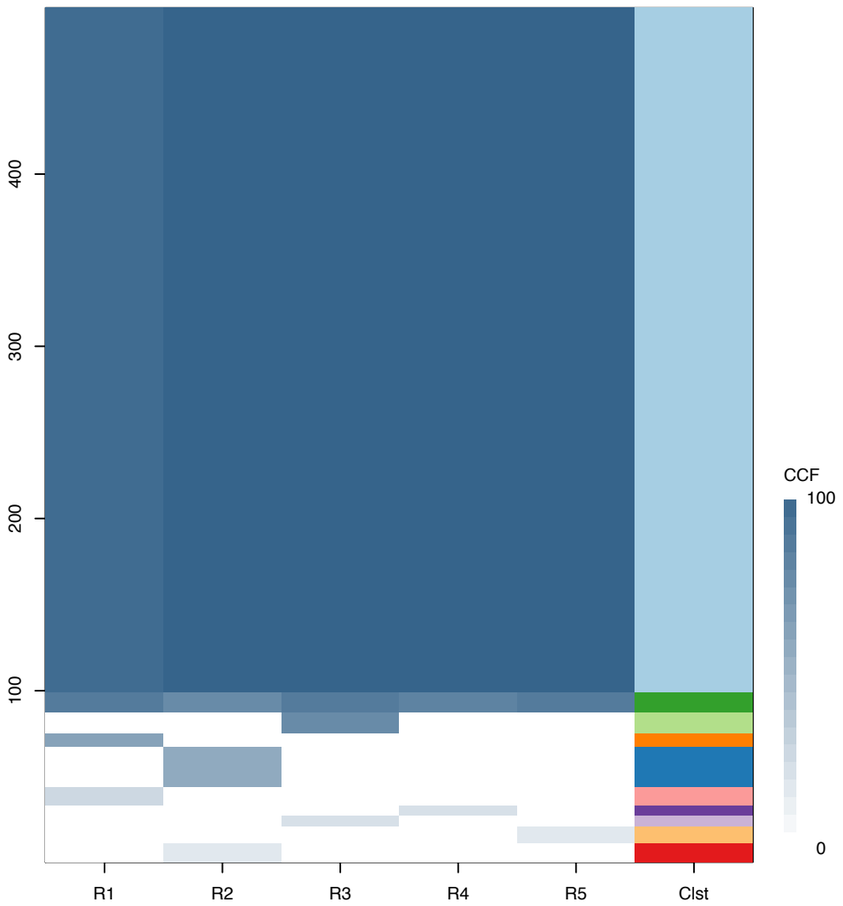
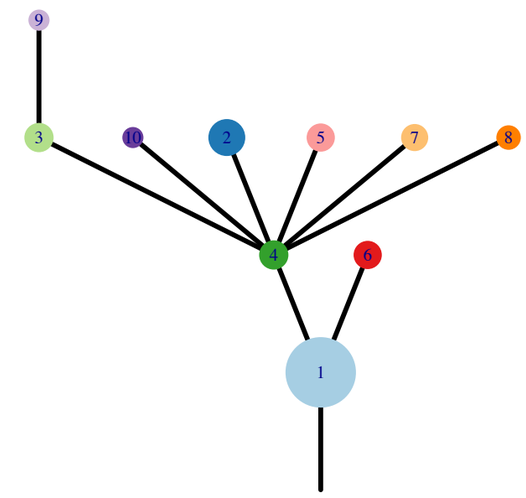


Fig.S12AK



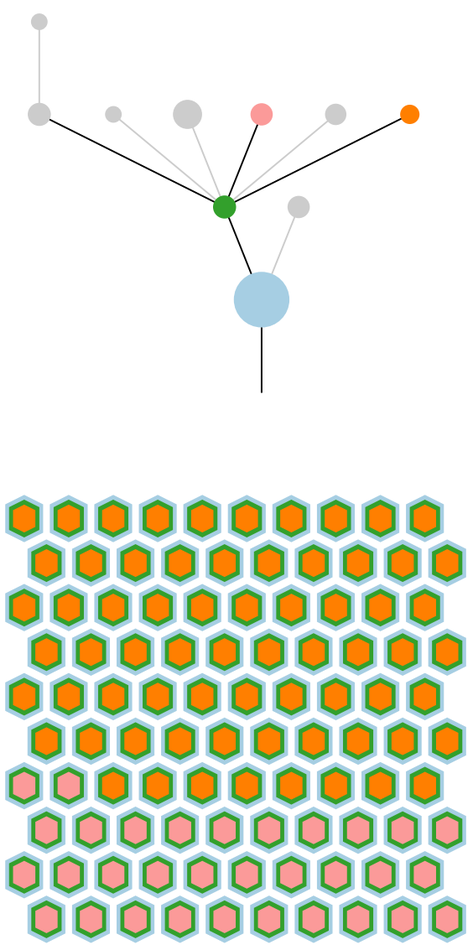
CRUK0037_H



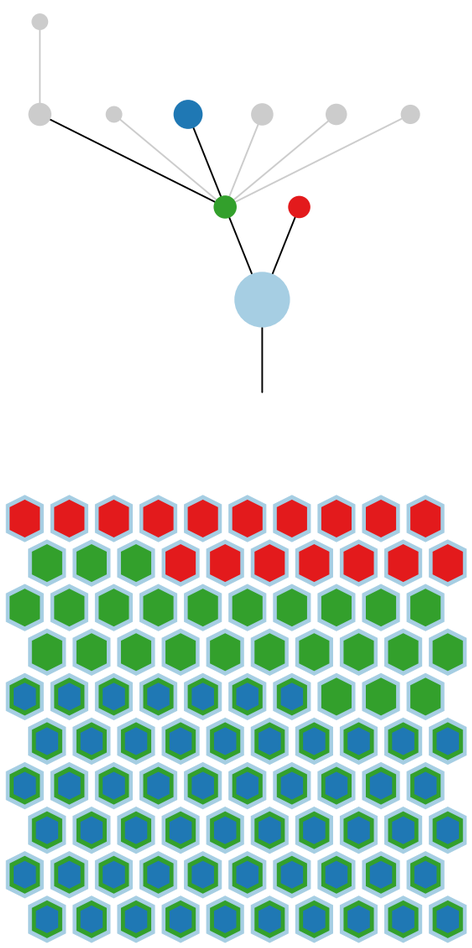
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

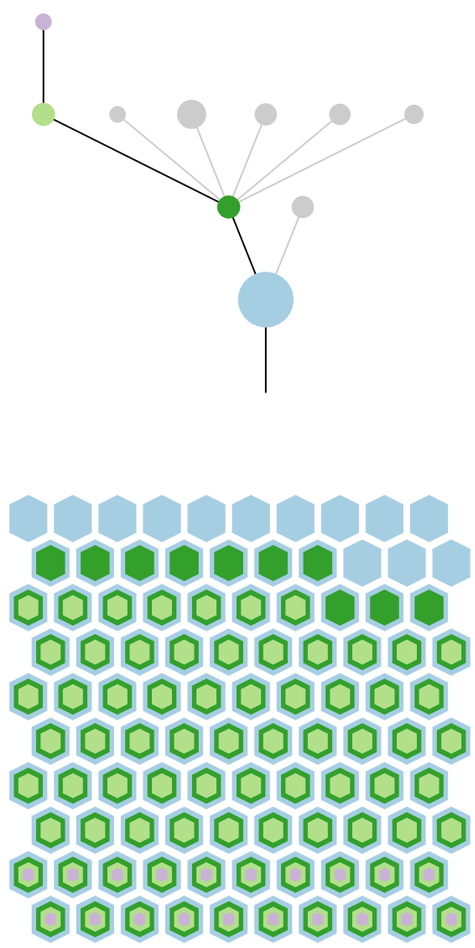
R1



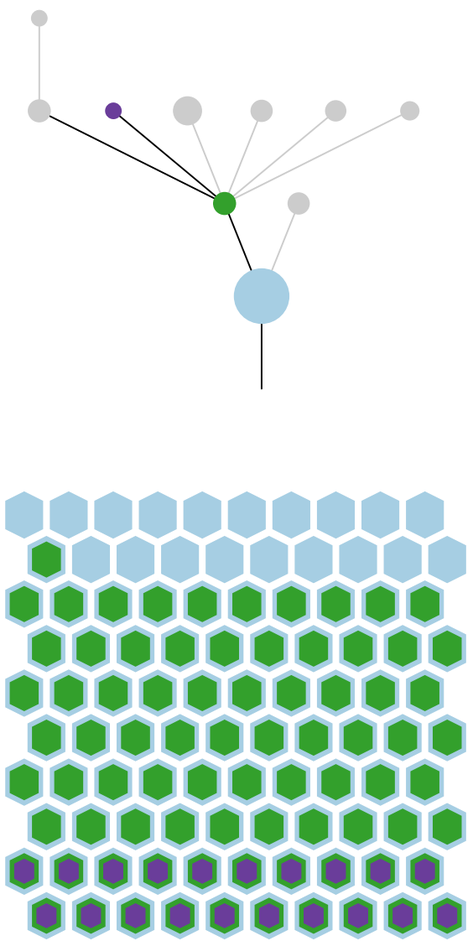
R2



R3



R4



R5

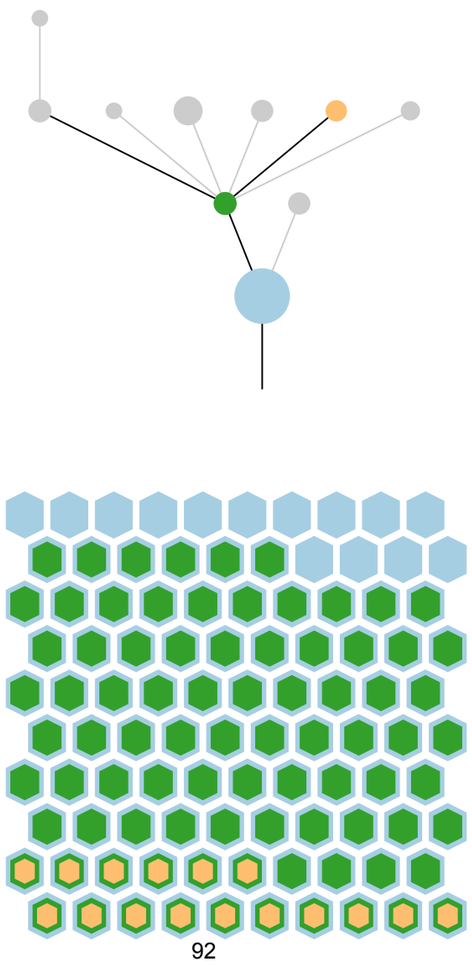
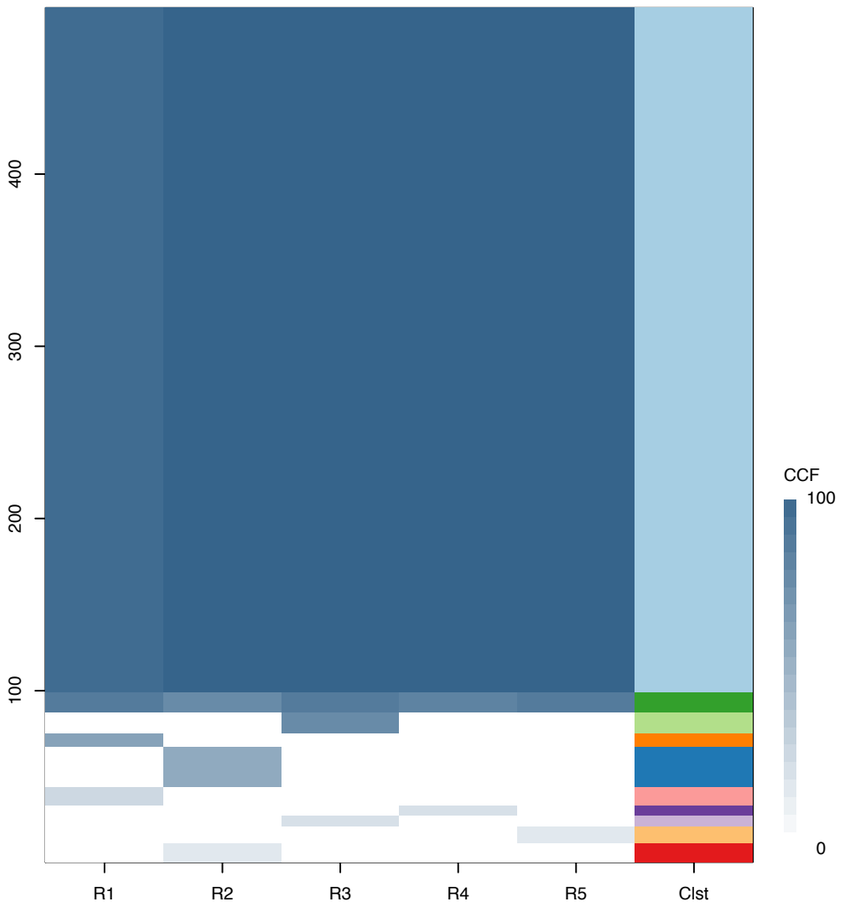
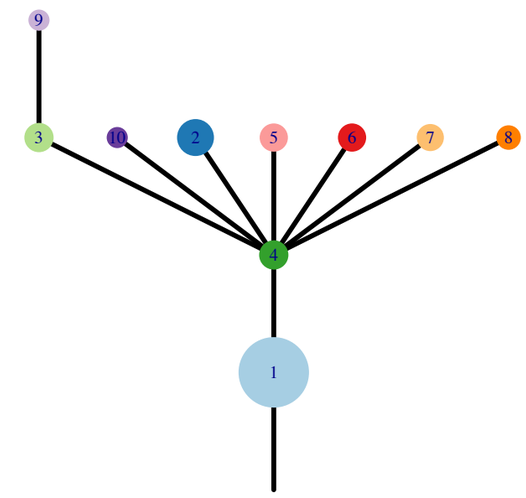


Fig.S12AK



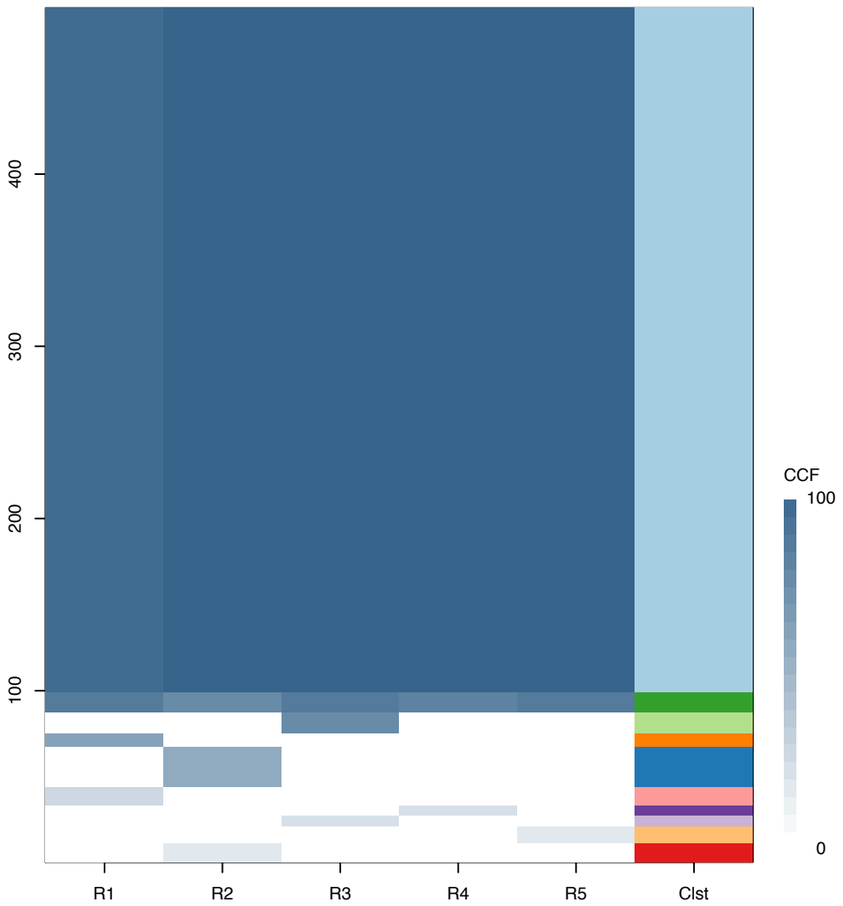
CRUK0037_I



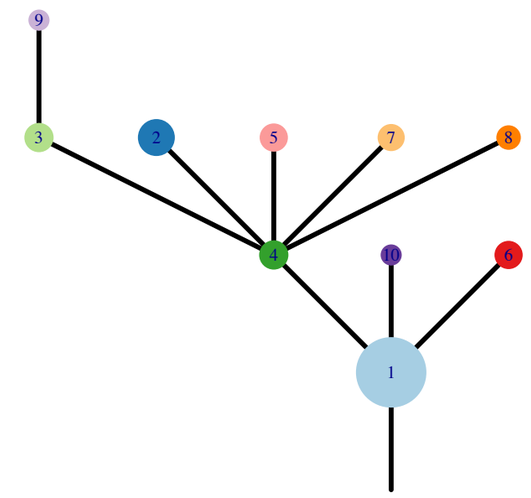
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

Fig.S12AK



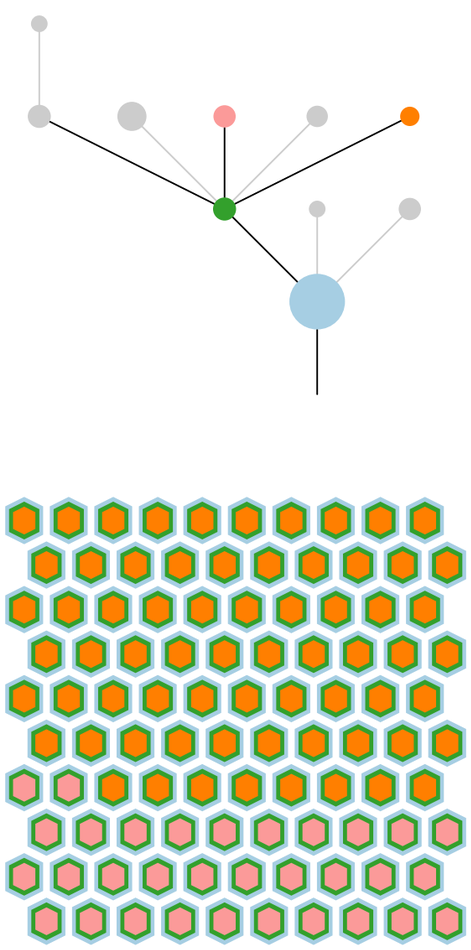
CRUK0037_J



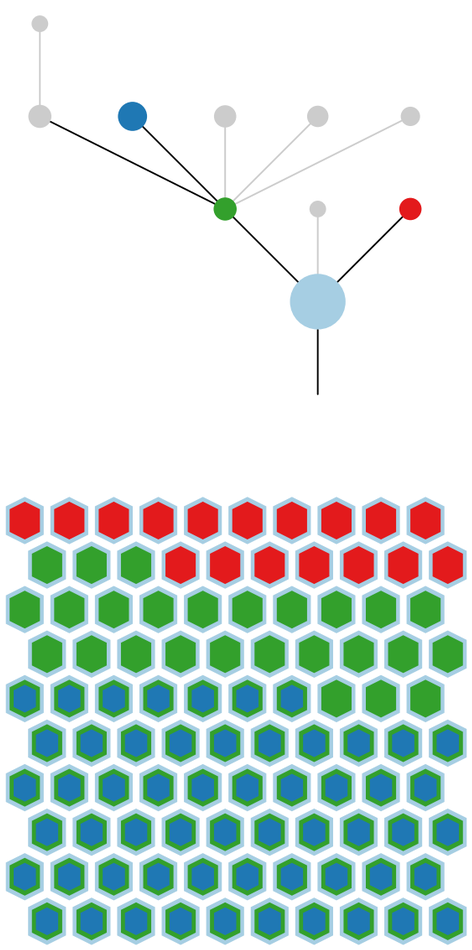
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

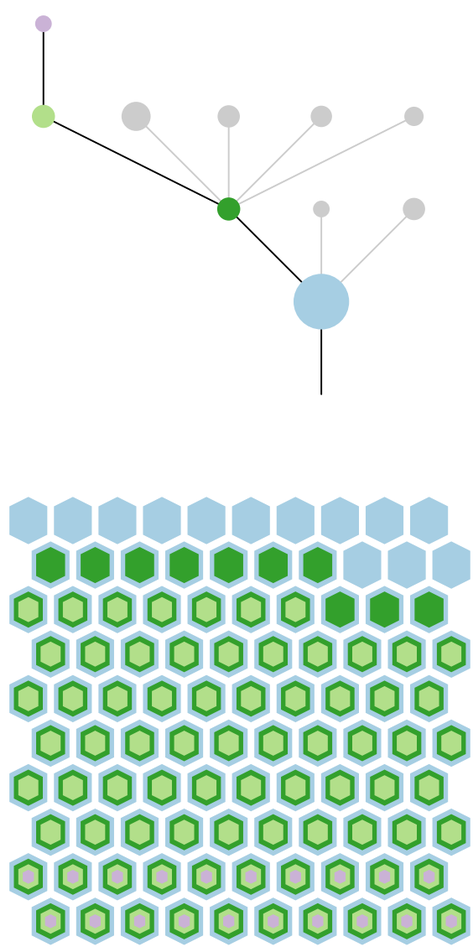
R1



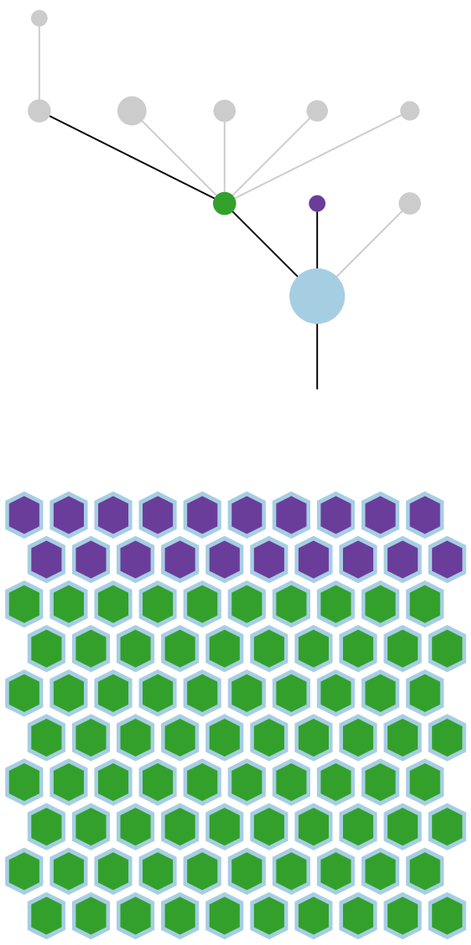
R2



R3



R4



R5

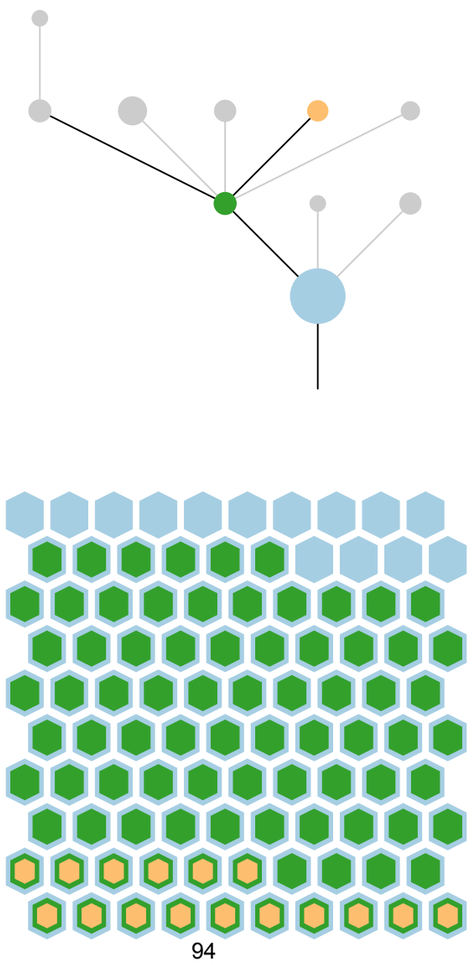
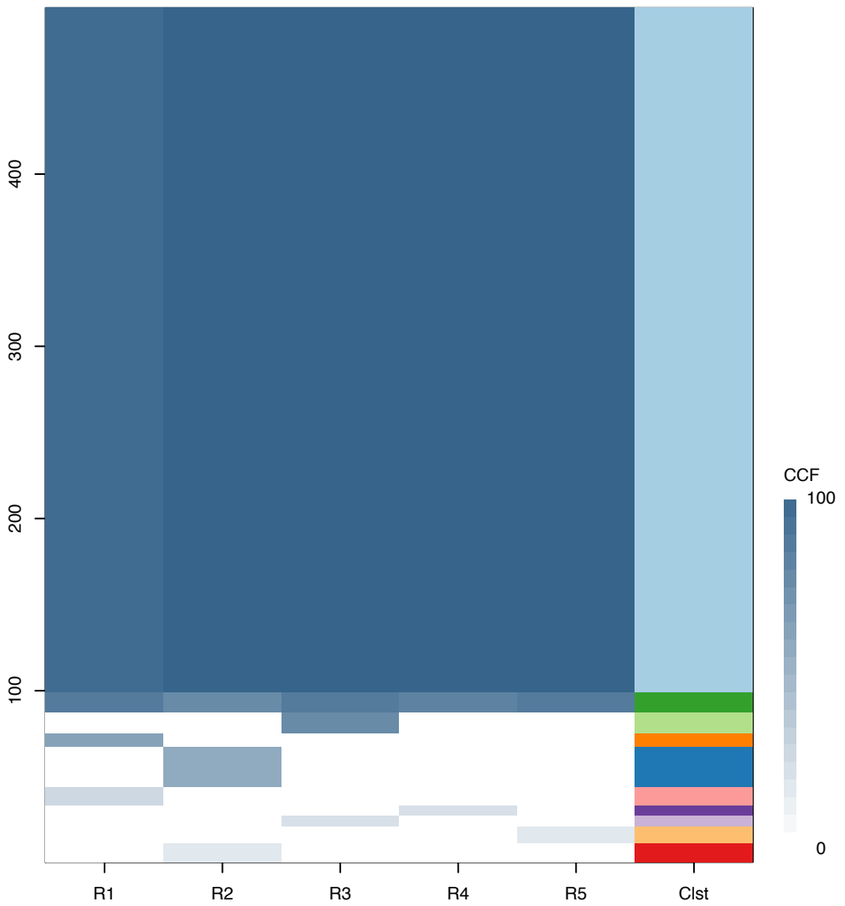
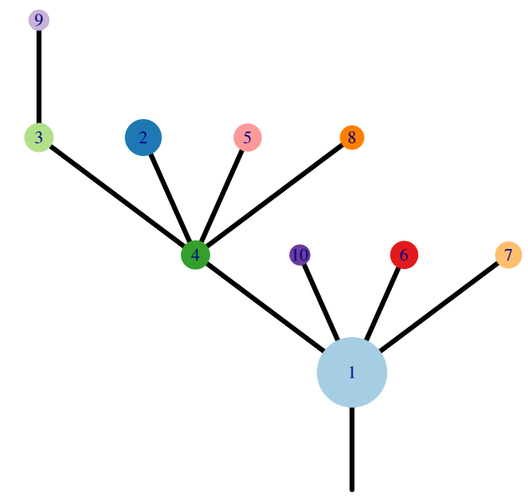


Fig.S12AK



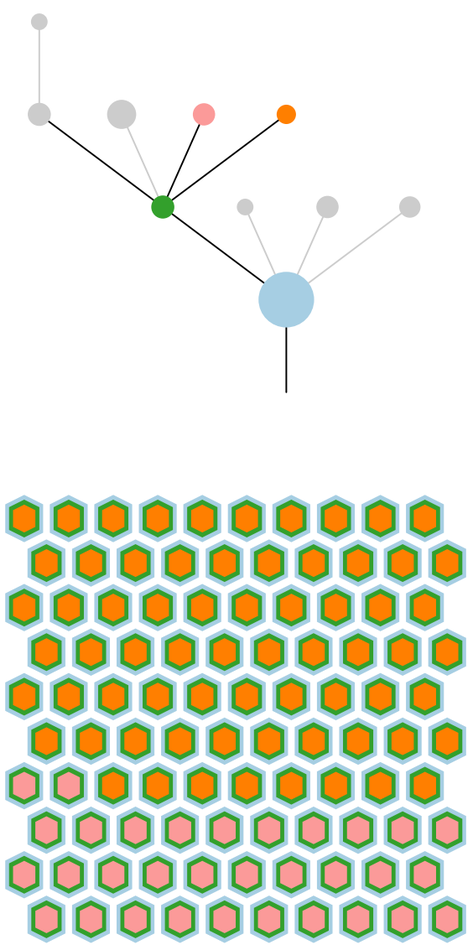
CRUK0037_K



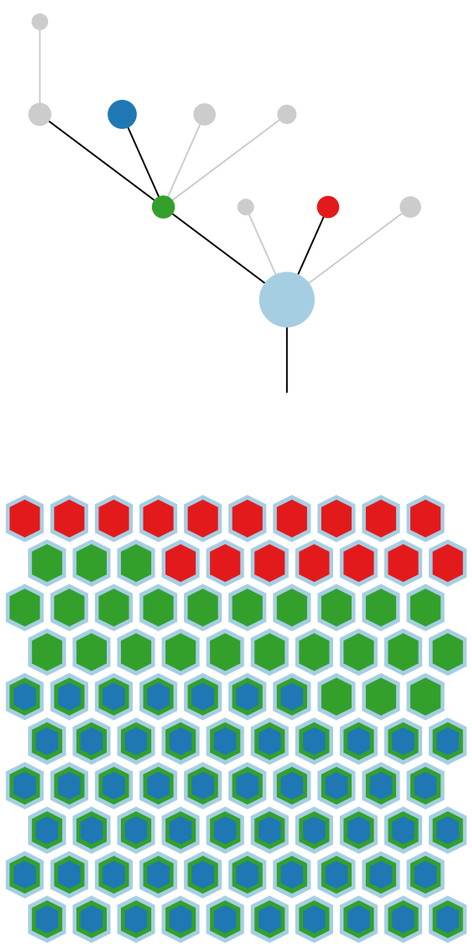
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

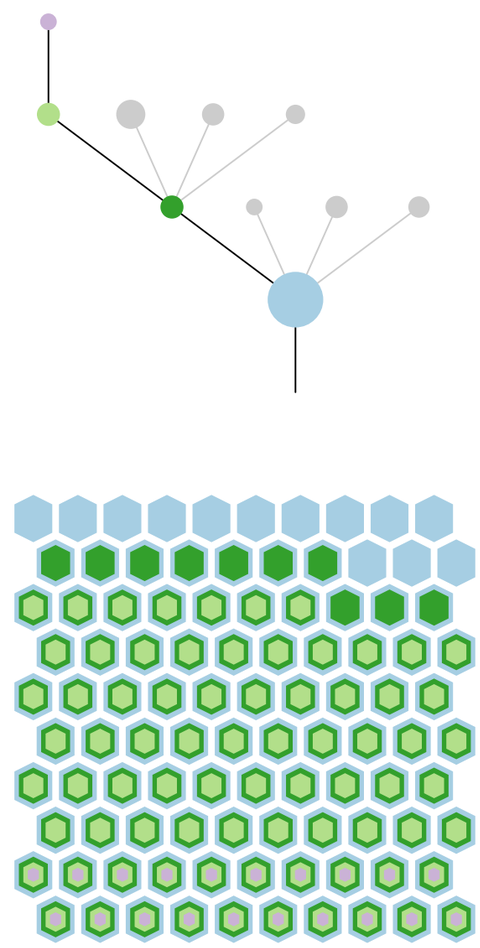
R1



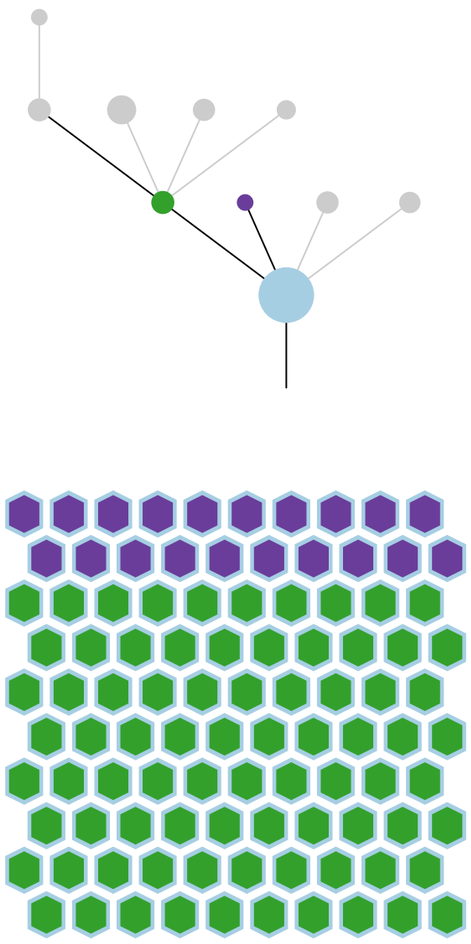
R2



R3



R4



R5

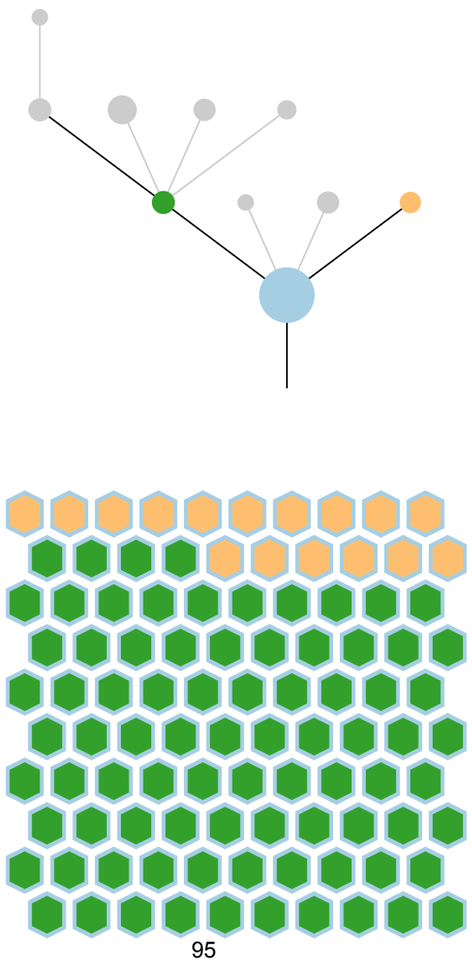
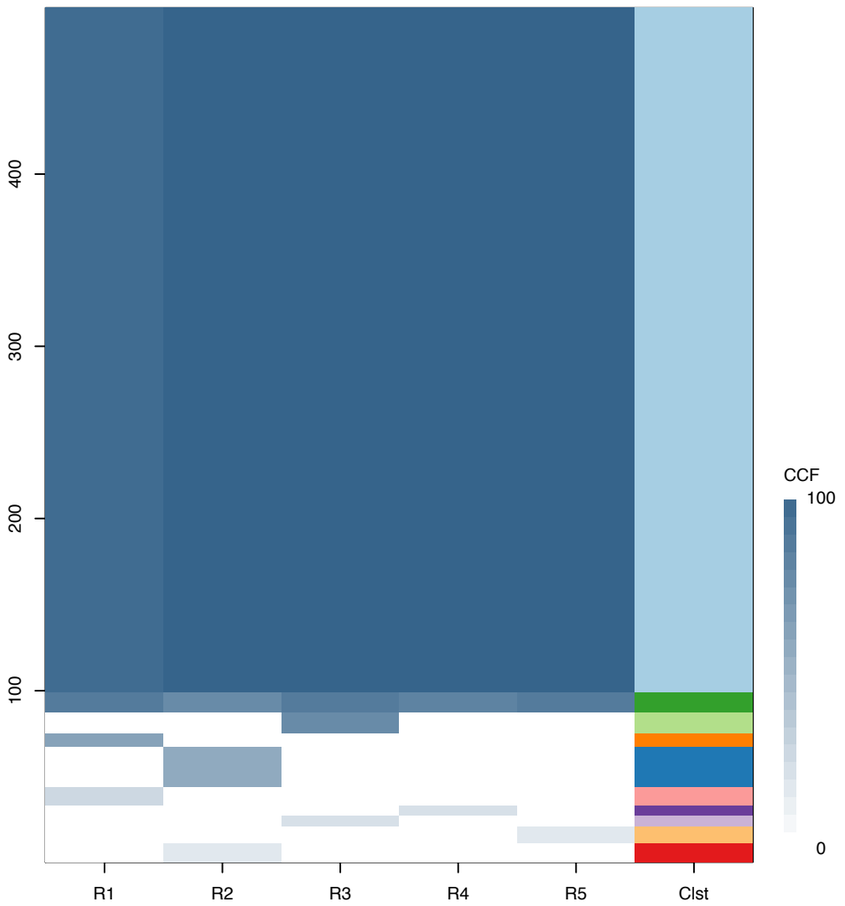
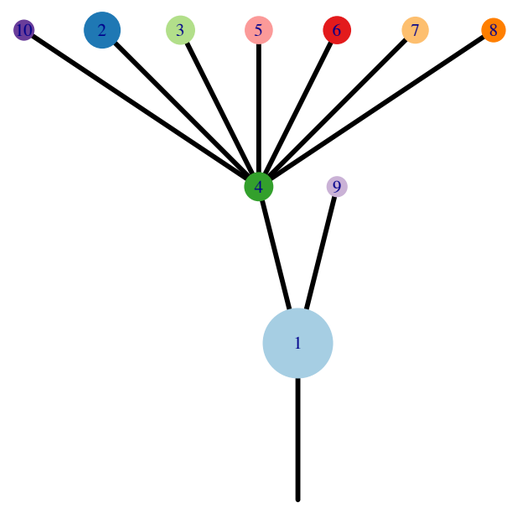


Fig.S12AK

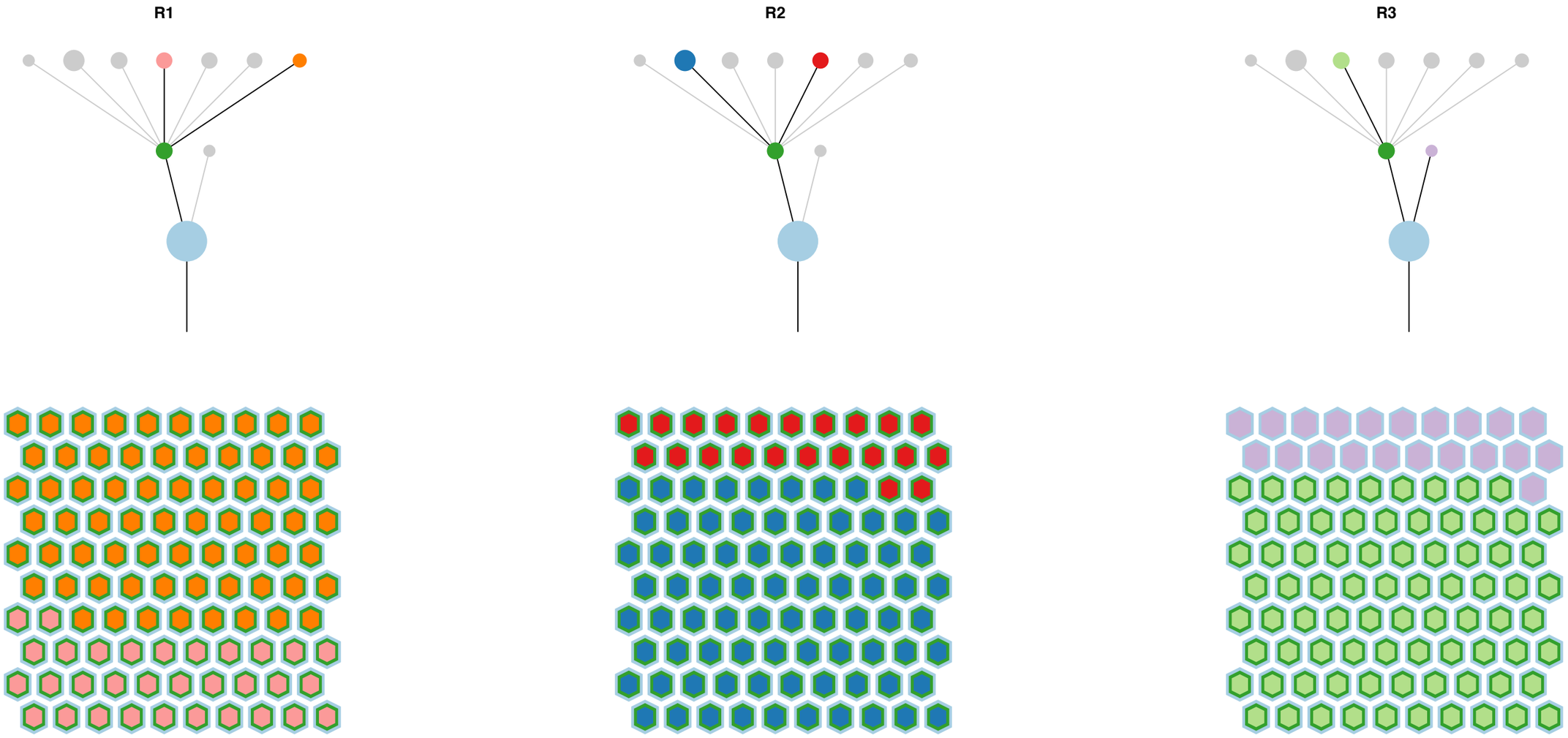


CRUK0037_L



Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV



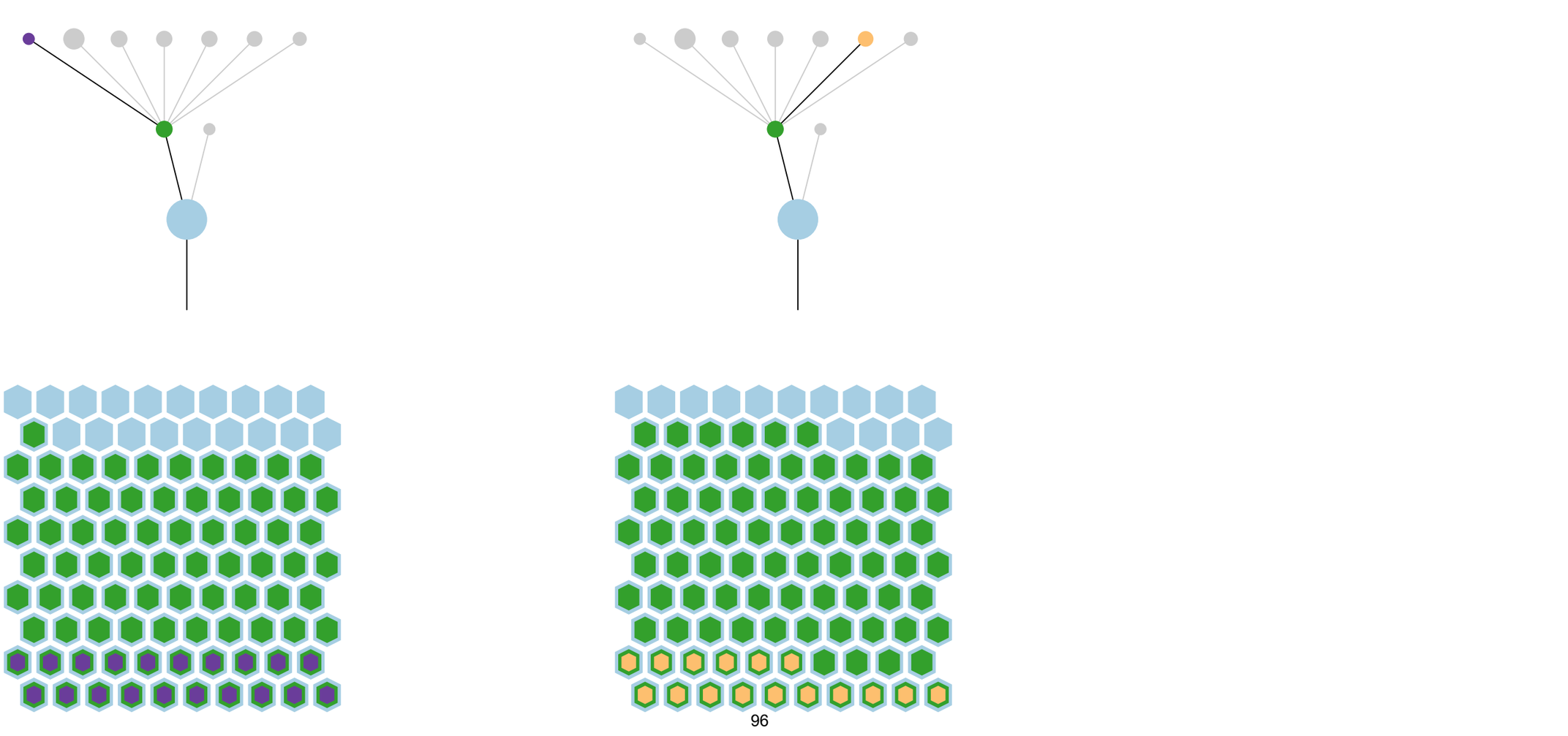
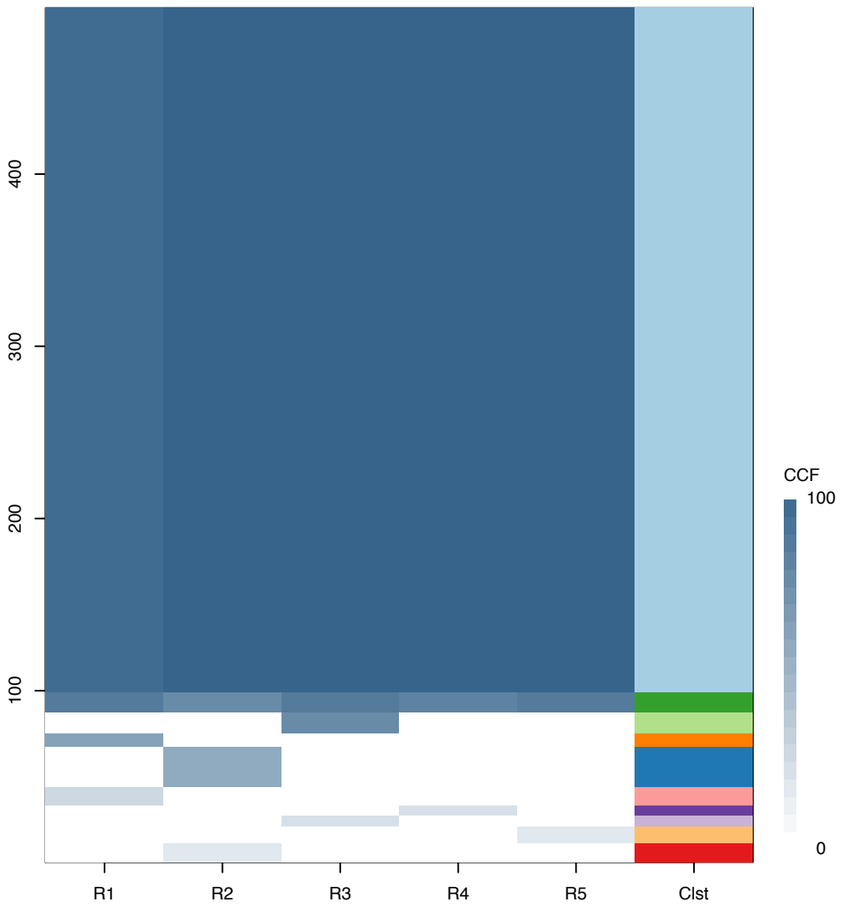
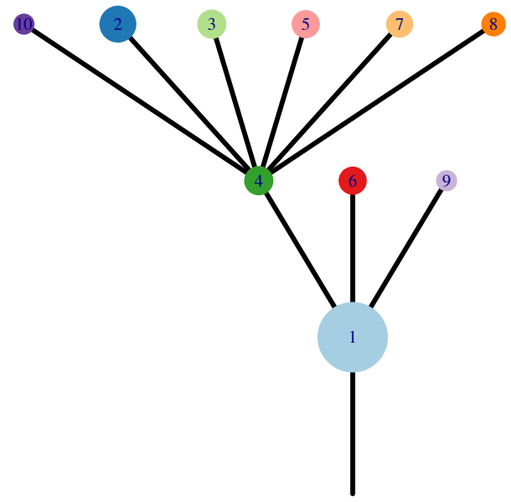


Fig.S12AK



CRUK0037_M



Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

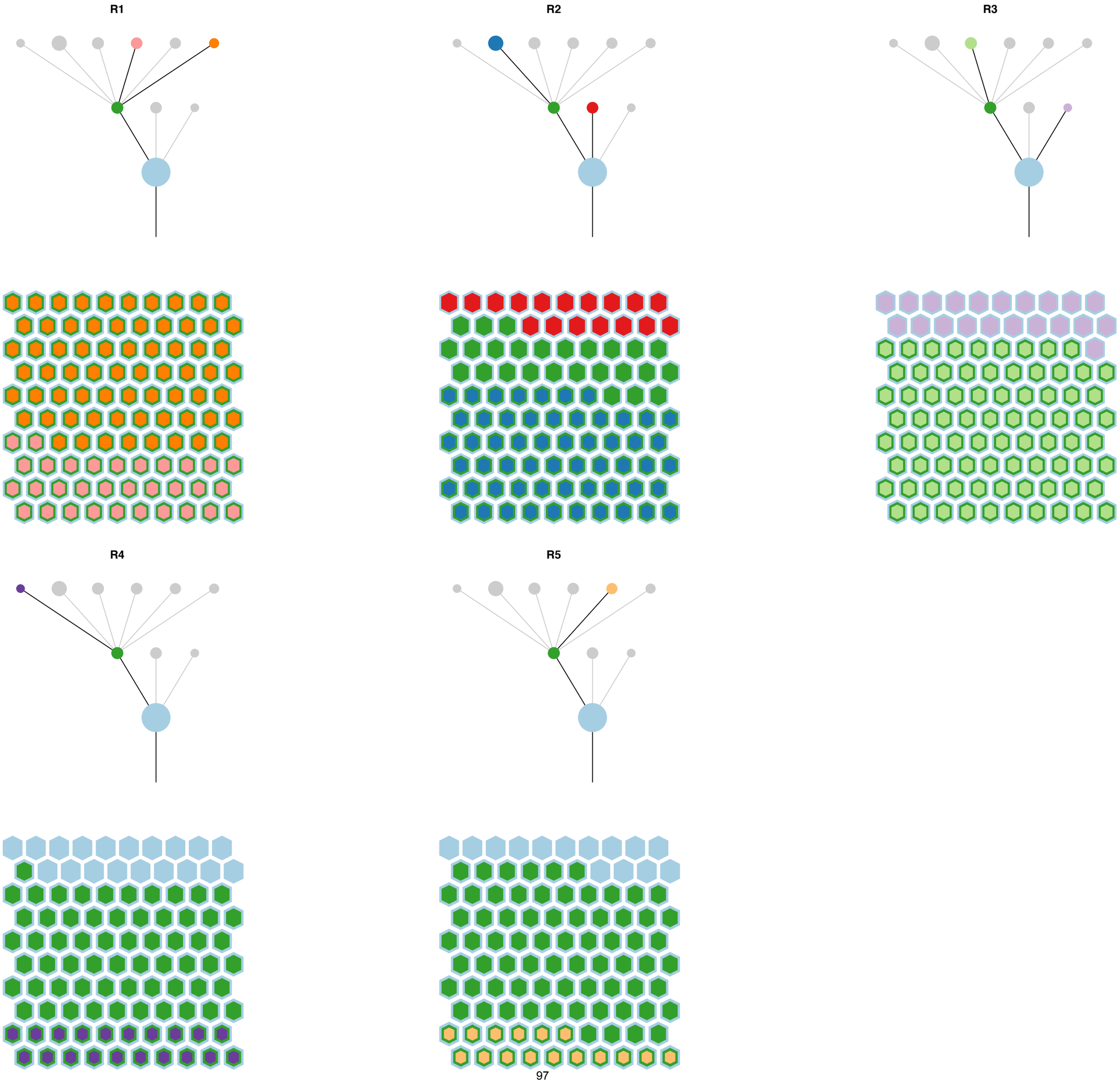
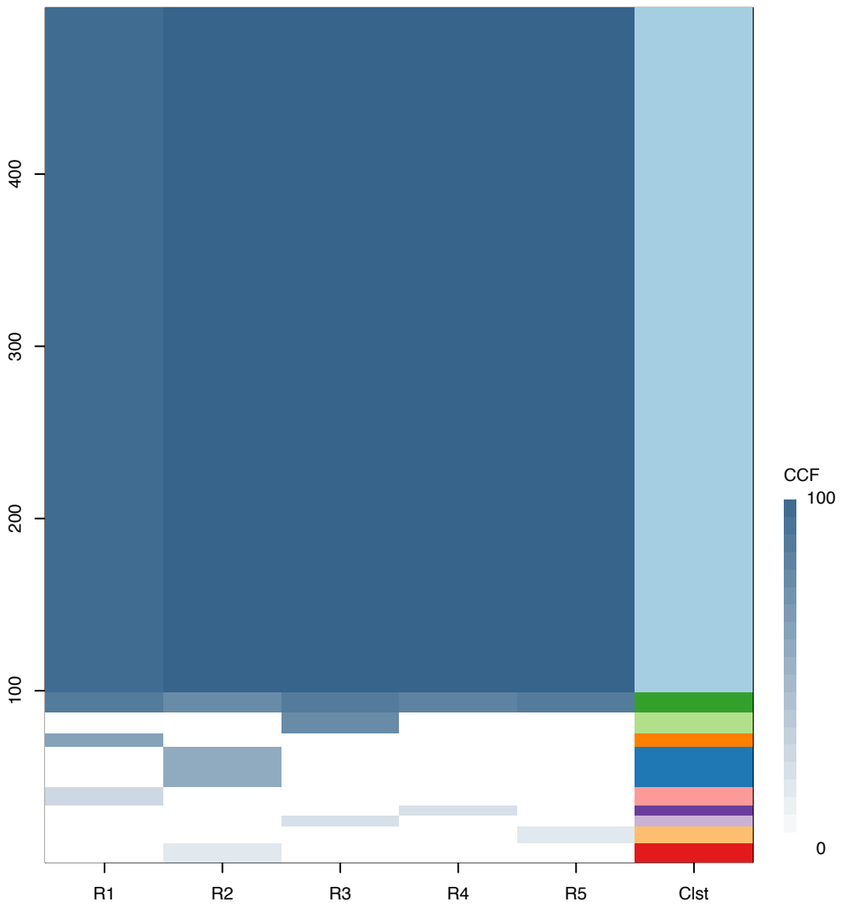
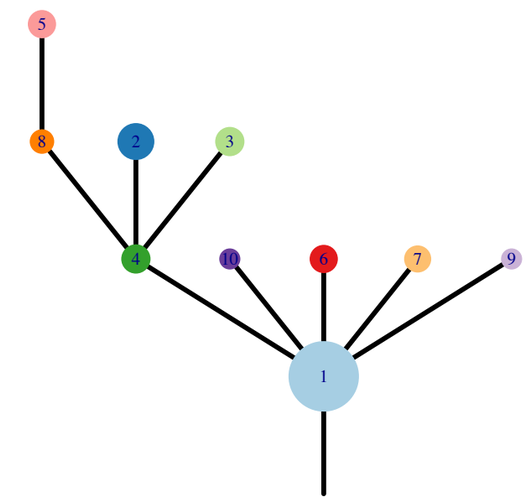


Fig.S12AK



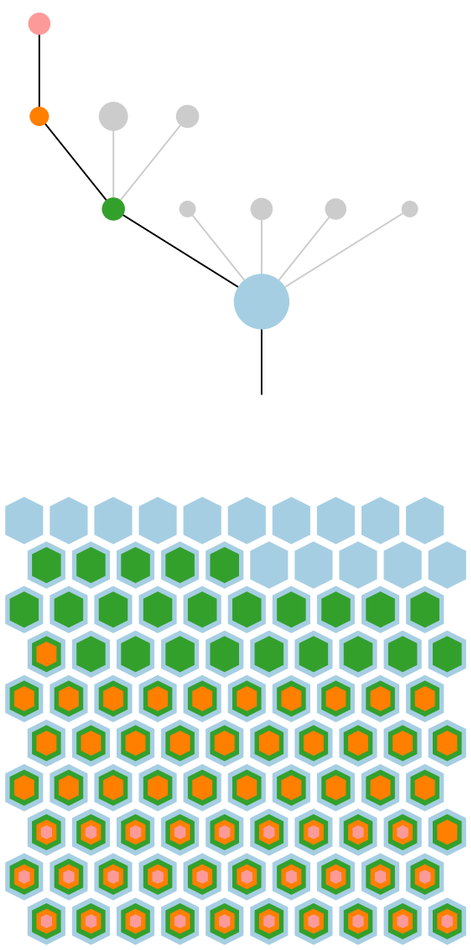
CRUK0037_N



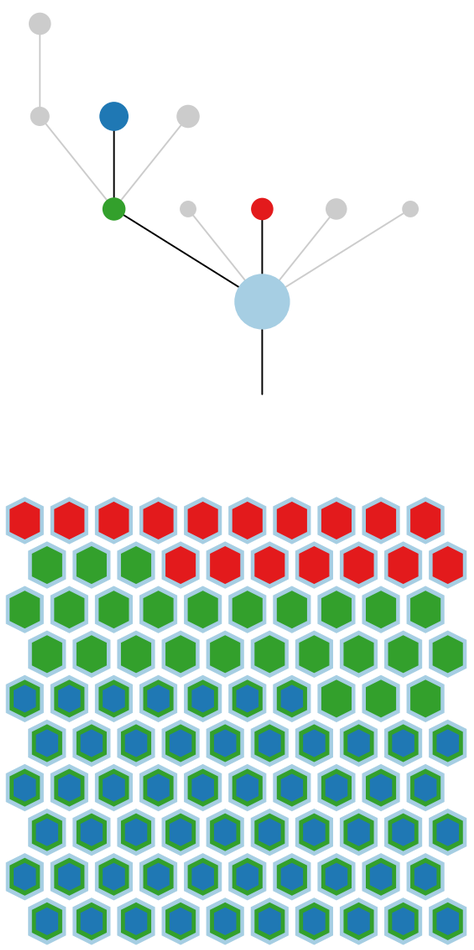
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

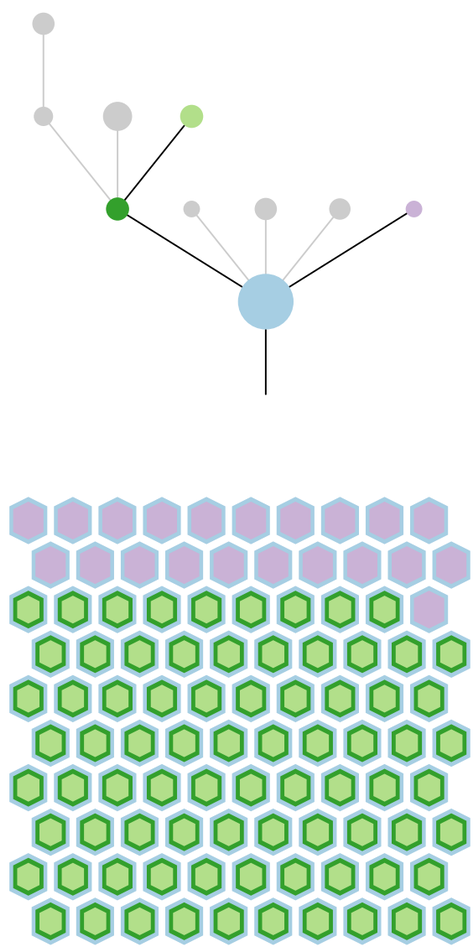
R1



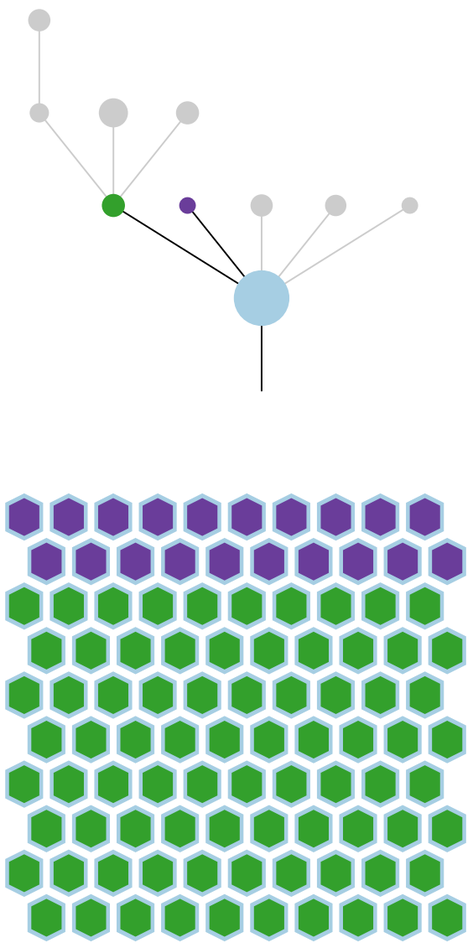
R2



R3



R4



R5

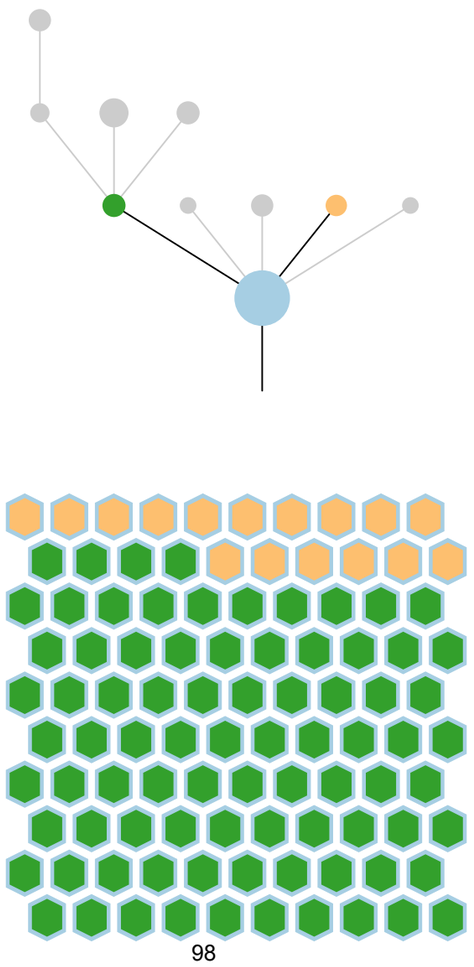
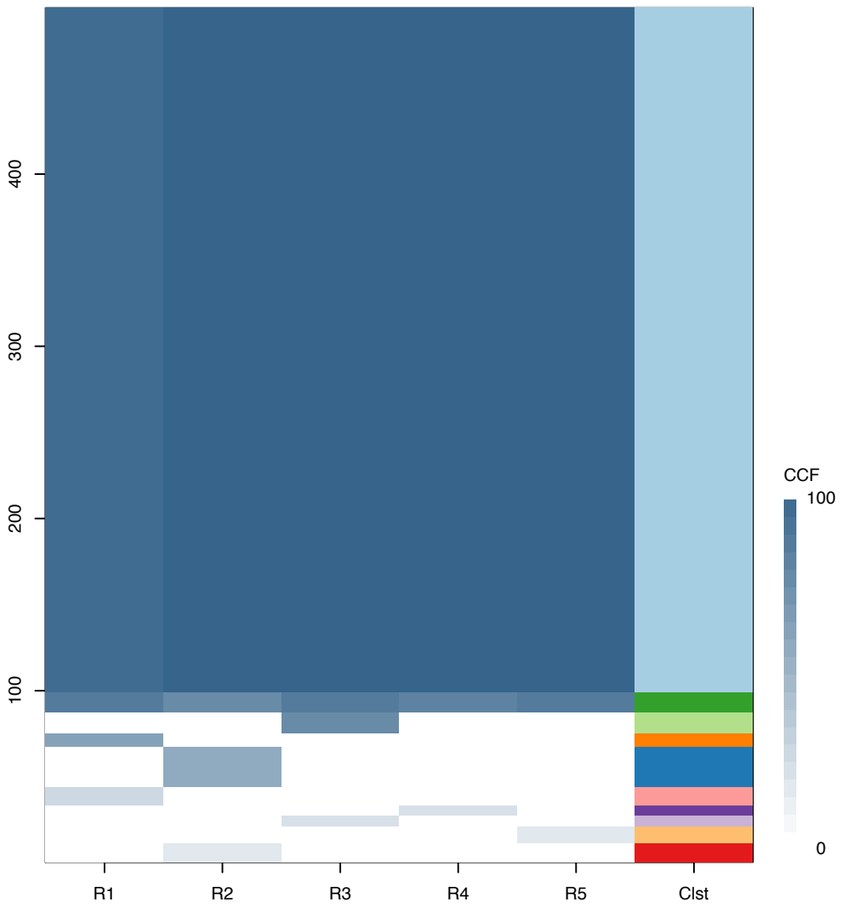
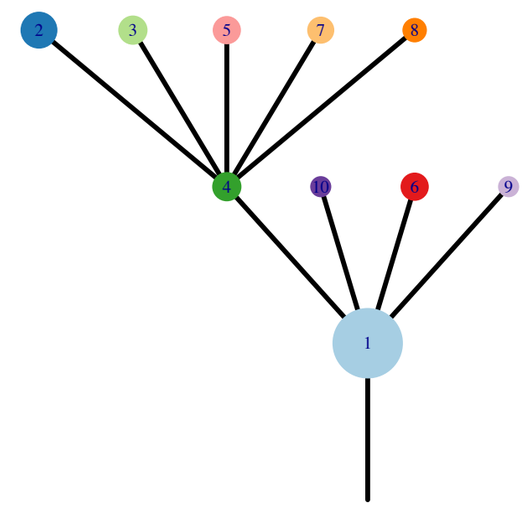


Fig.S12AK



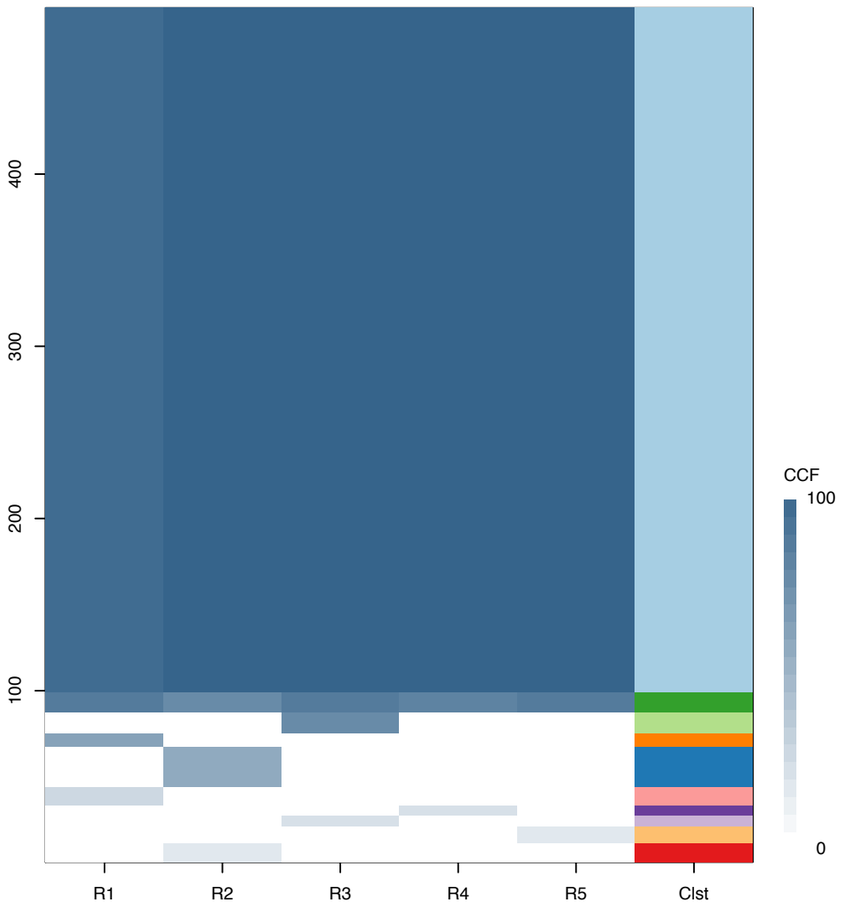
CRUK0037_O



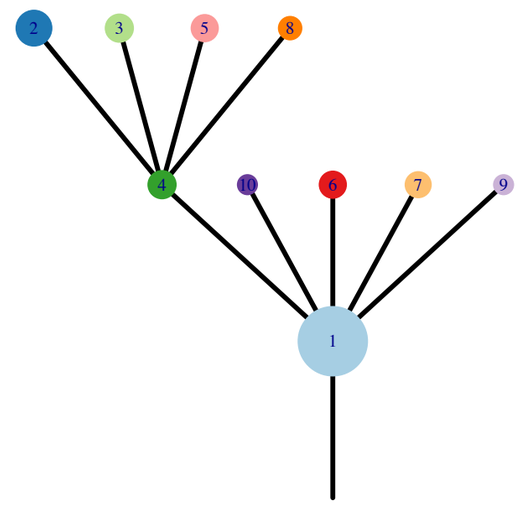
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

Fig.S12AK



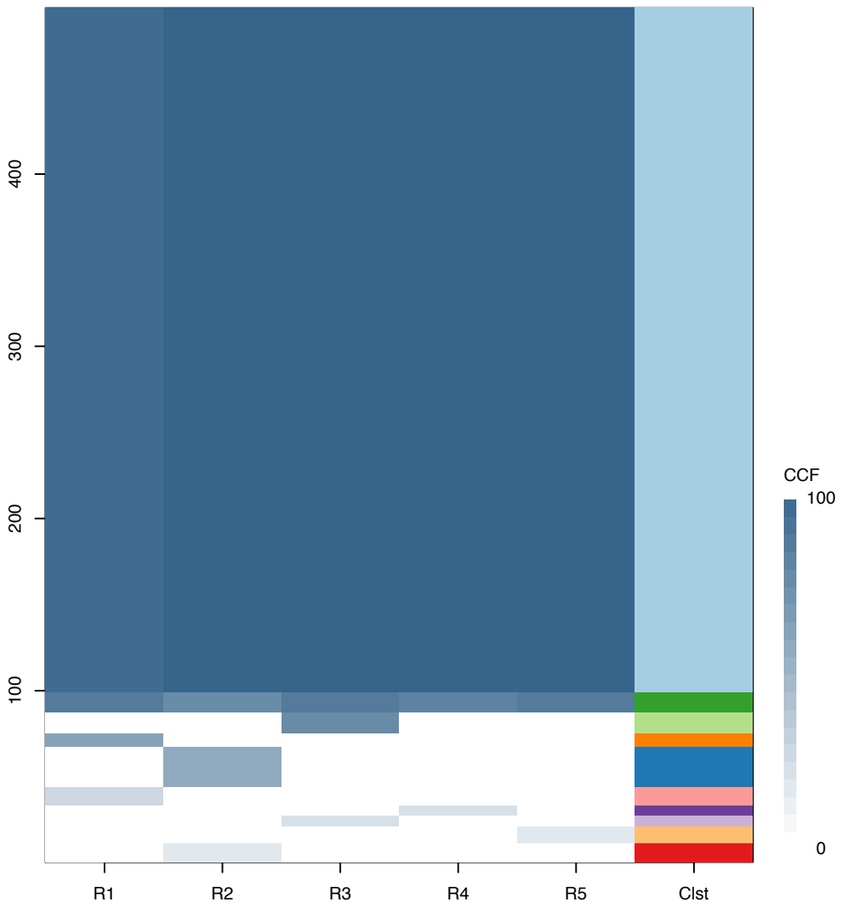
CRUK0037_P



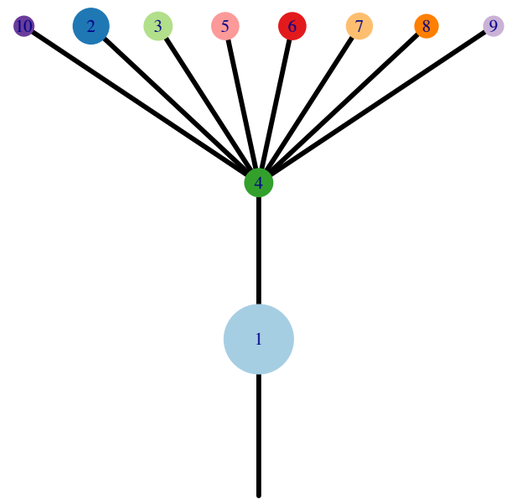
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

Fig.S12AK



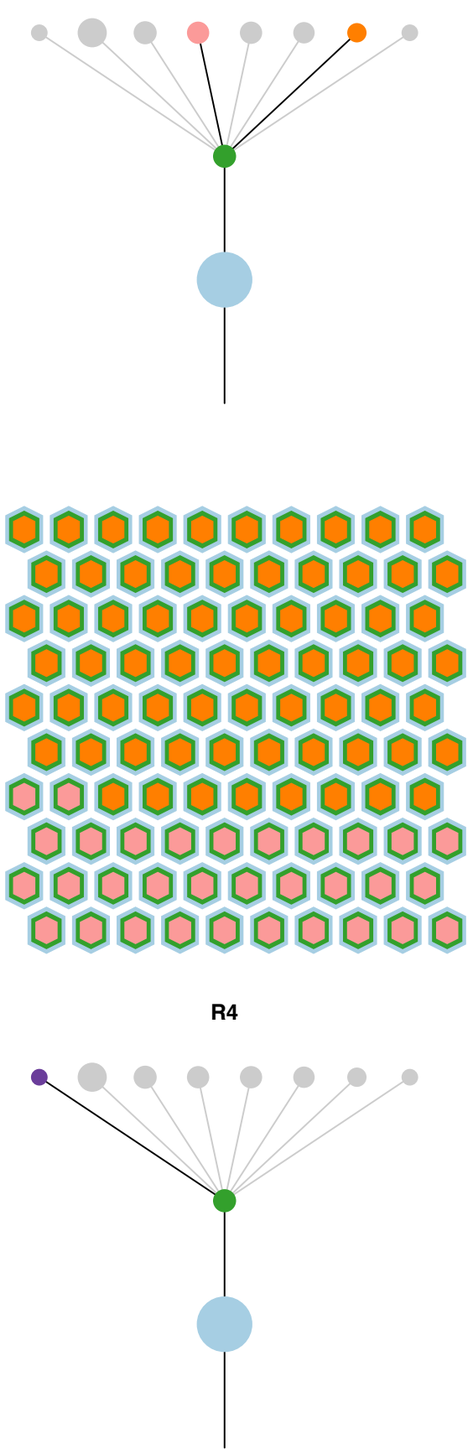
CRUK0037_Q



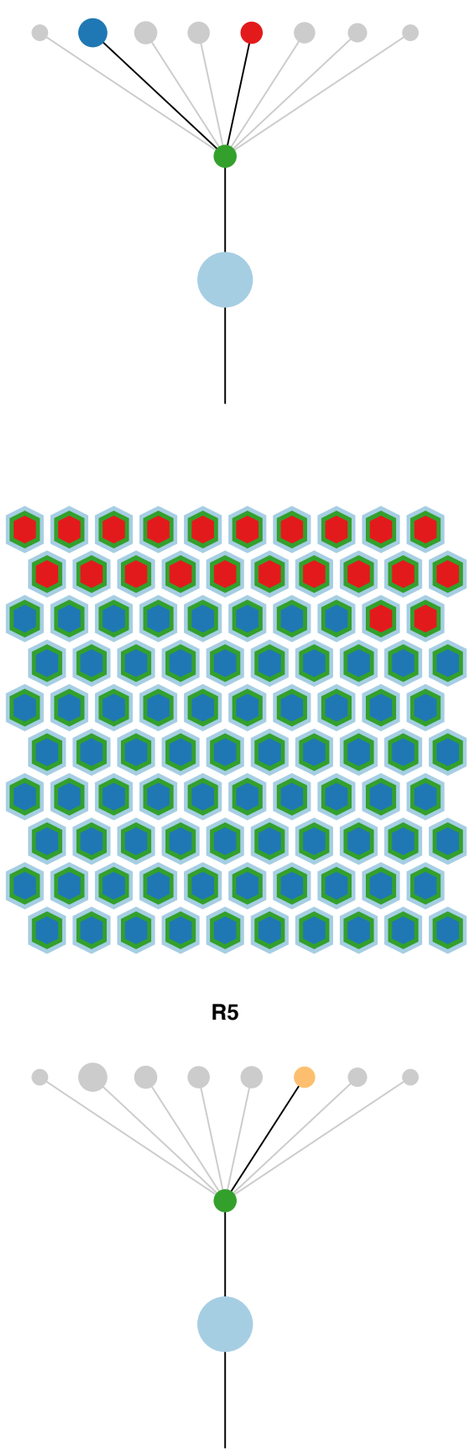
Histology:Adeno, Age:84, PackYears:9.75, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
CREBBP	1	16p13.3	SNV
NCOA6	1	20q11.22	SNV
GNAS	1	20q13.32	Amp
HOOK3	10	8p11.21	Amp
WHSC1L1	?	8p11.23	Amp
FGFR1	?	8p11.23	Amp
KAT6A	?	8p11.21	Amp
IKBKB	?	8p11.21	Amp
RBM10	?		SNV

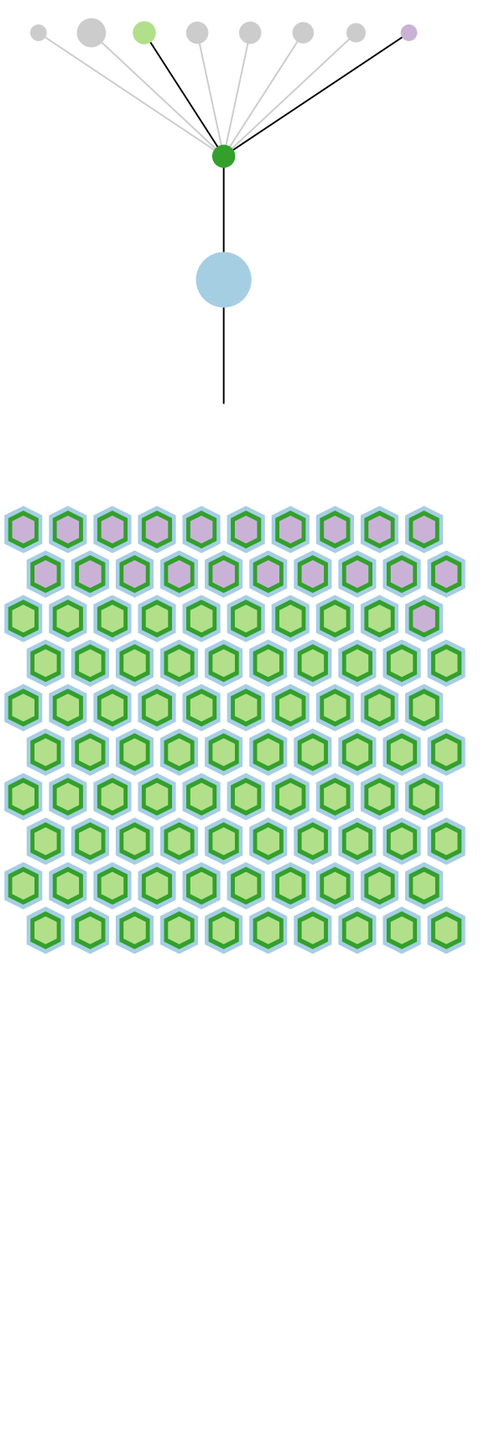
R1



R2



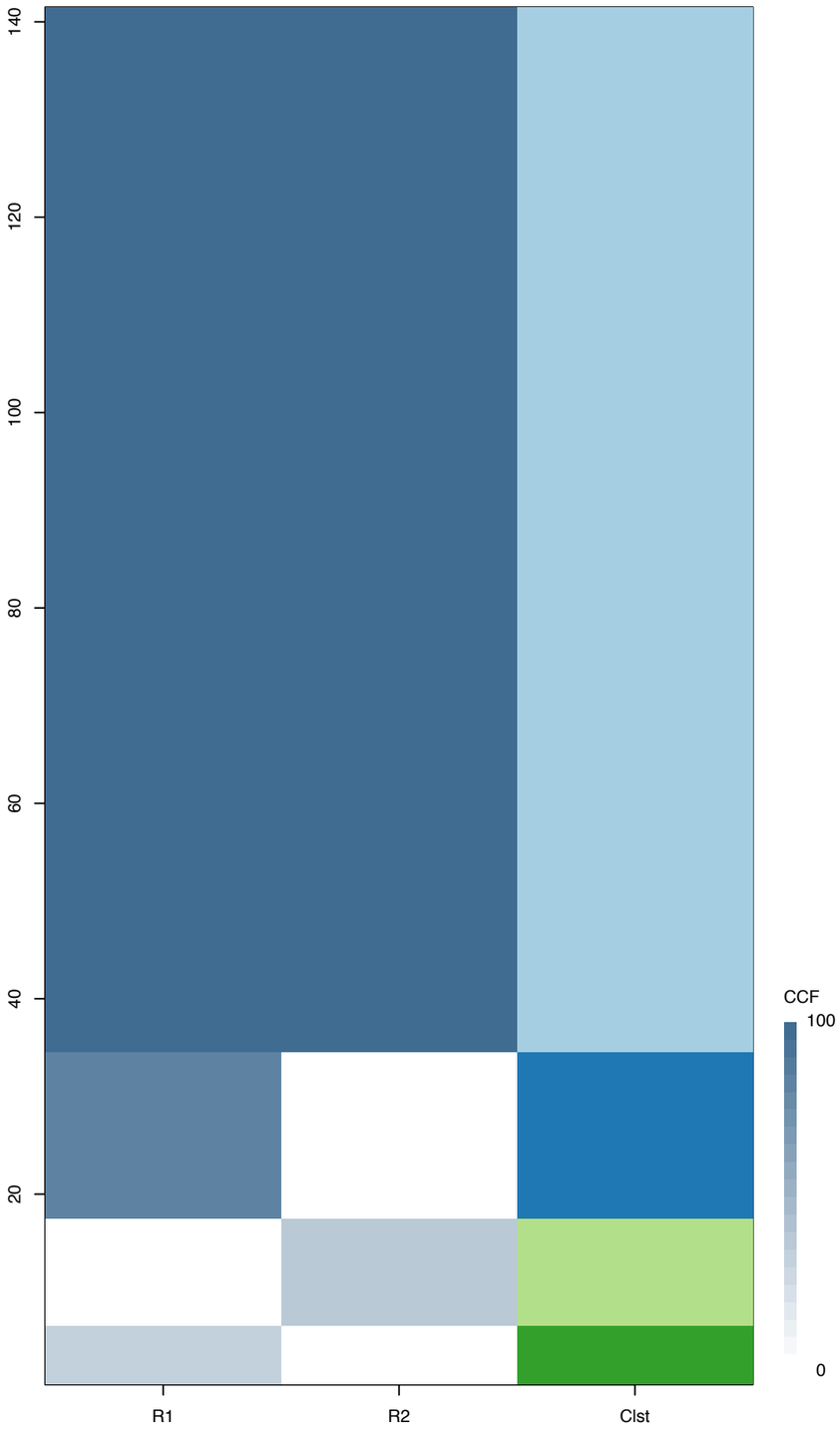
R3



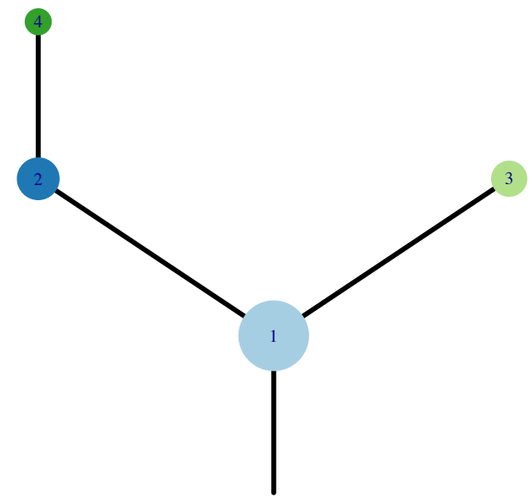
R4

R5

Fig.S12AL



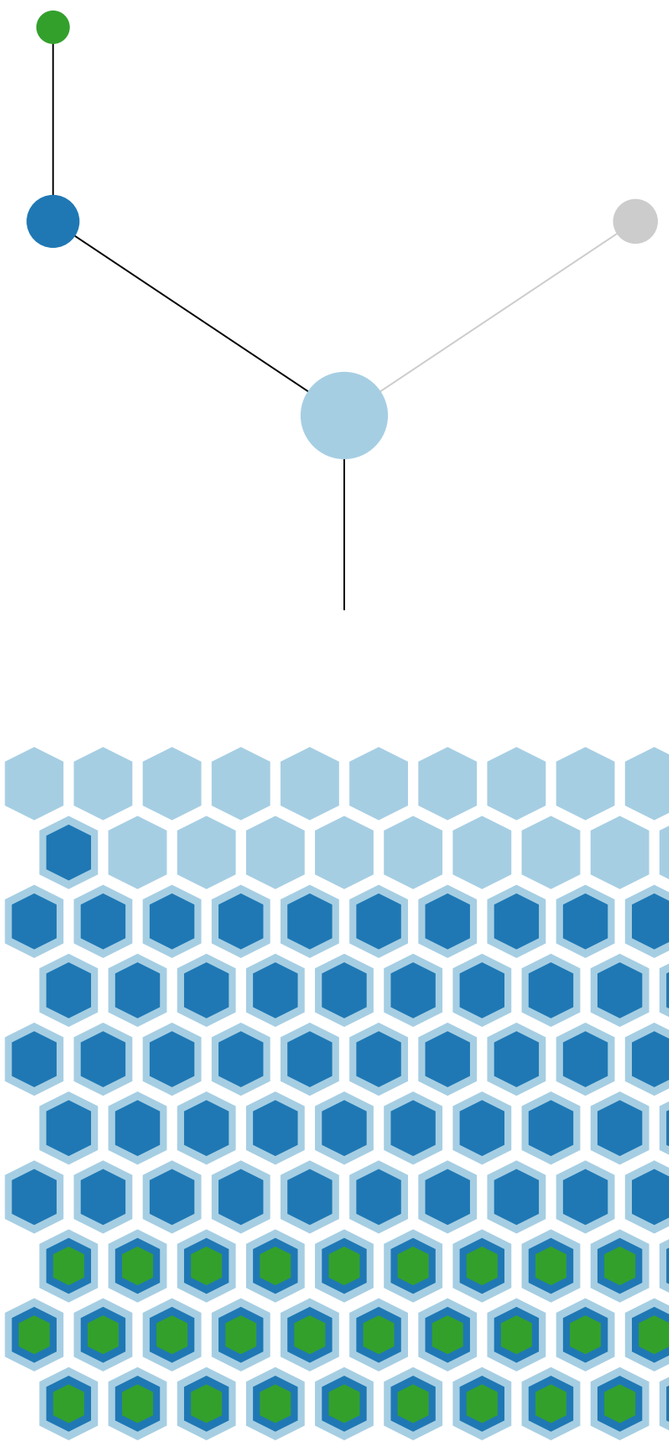
CRUK0038_A



Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
KMT2D	3	12q13.12	SNV

Histology:Adeno, Age:82, PackYears:32, Size:25
Stage:1b, Gender:Female, GD:Not GD, Recur:no

R1



R2

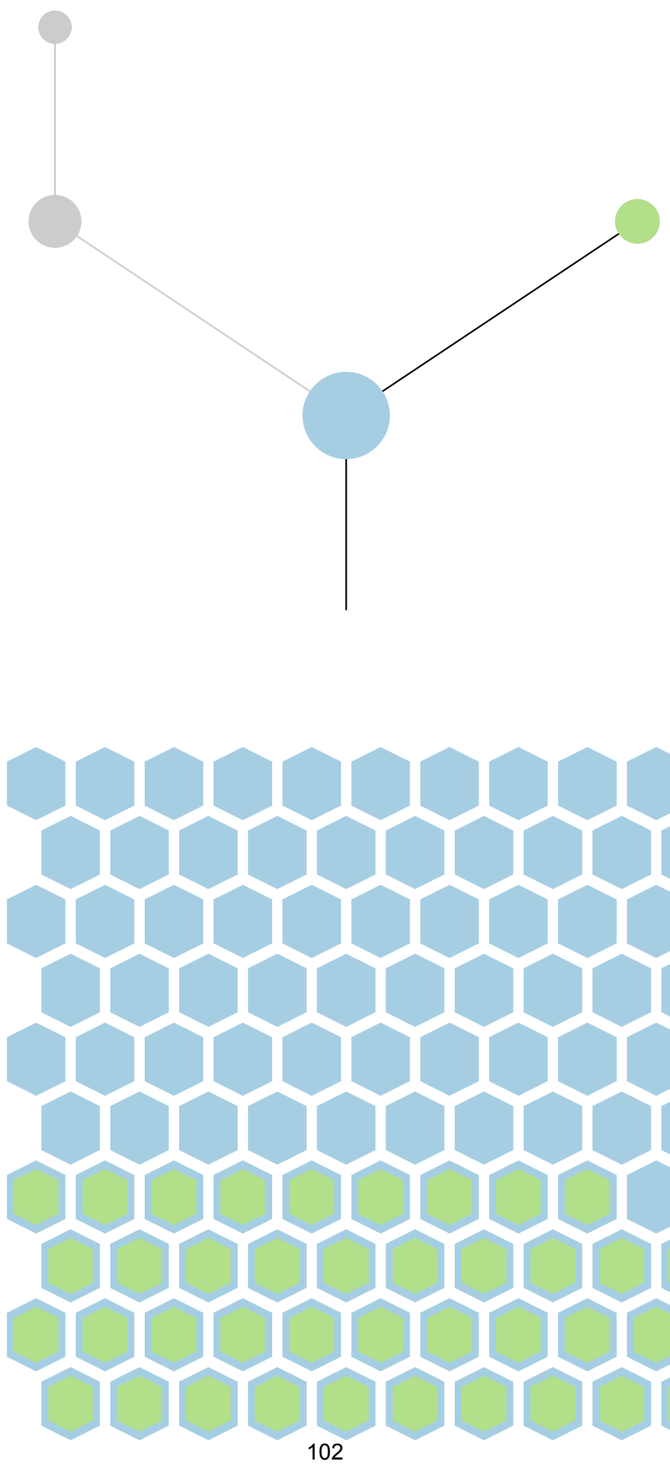
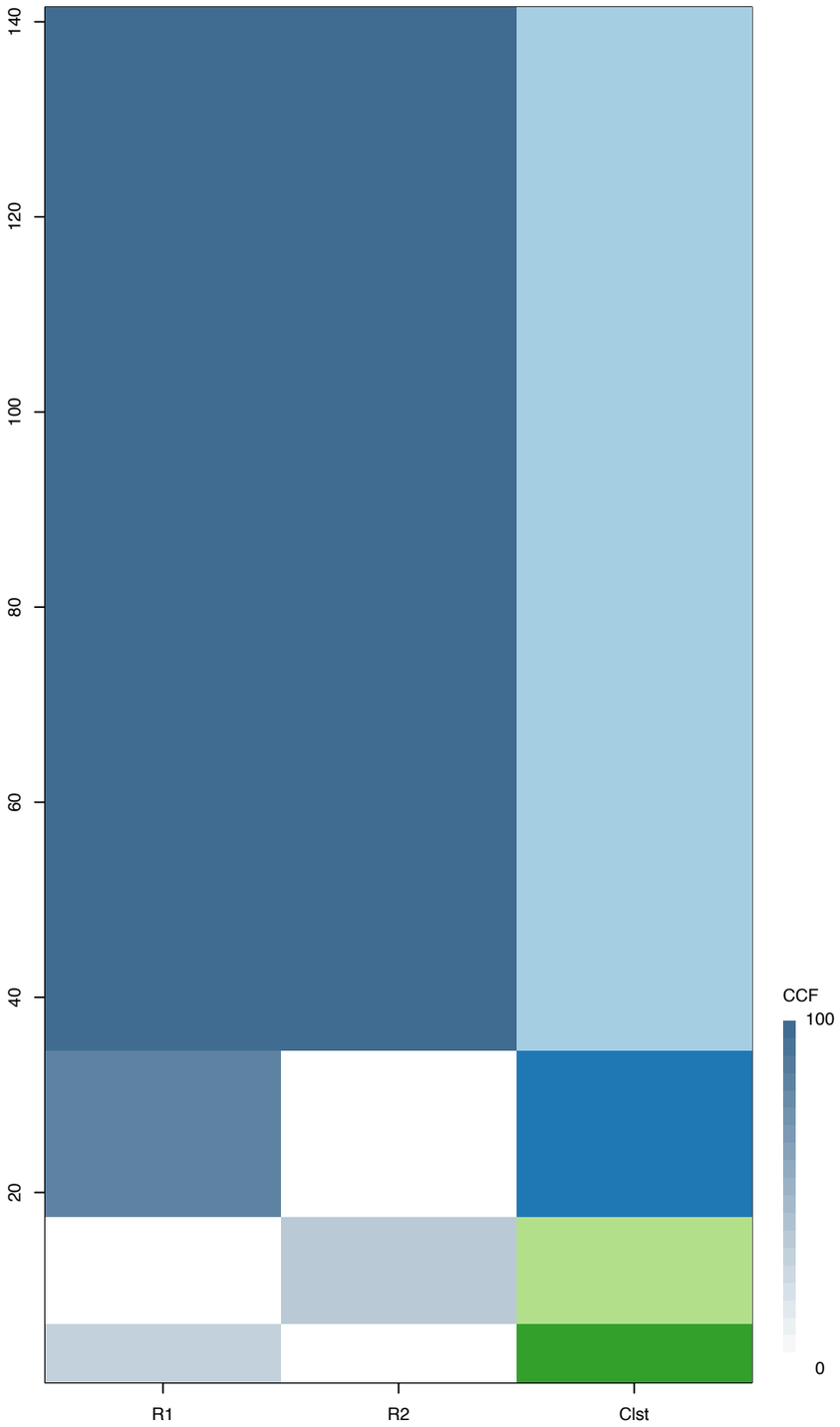
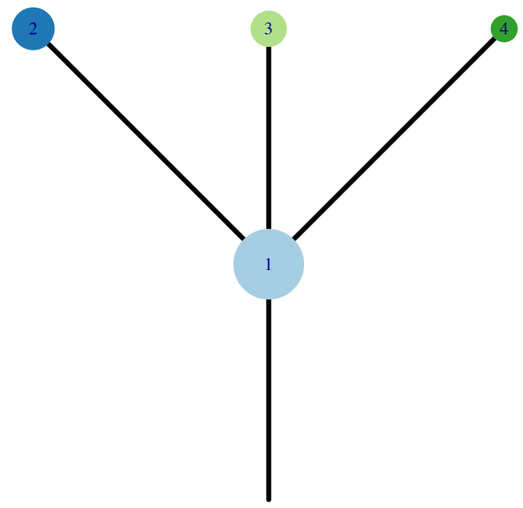


Fig.S12AL



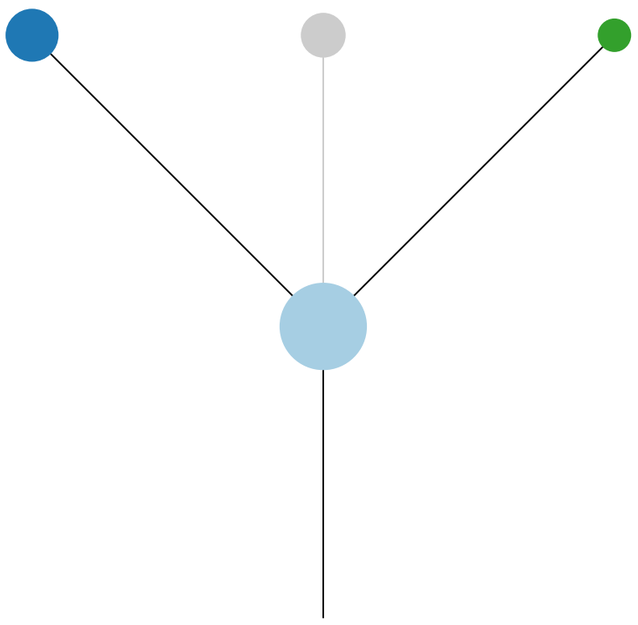
CRUK0038_B



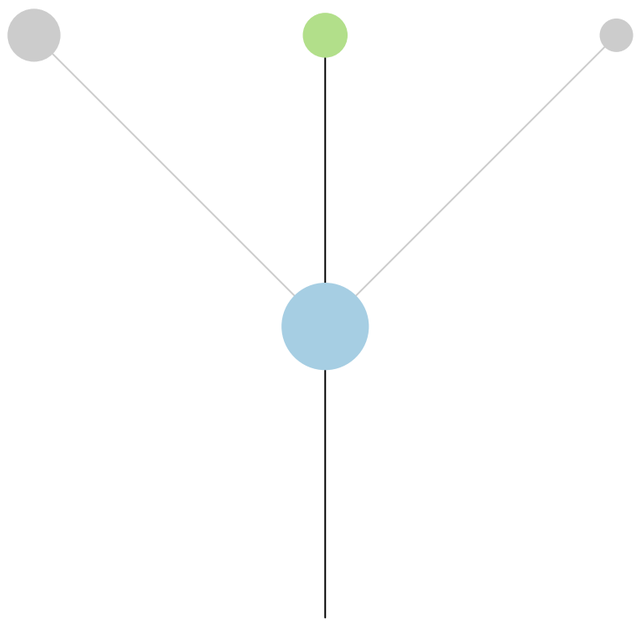
Gene	Cluster	Cytoband	Type
KRAS	1	12p12.1	SNV
KMT2D	3	12q13.12	SNV

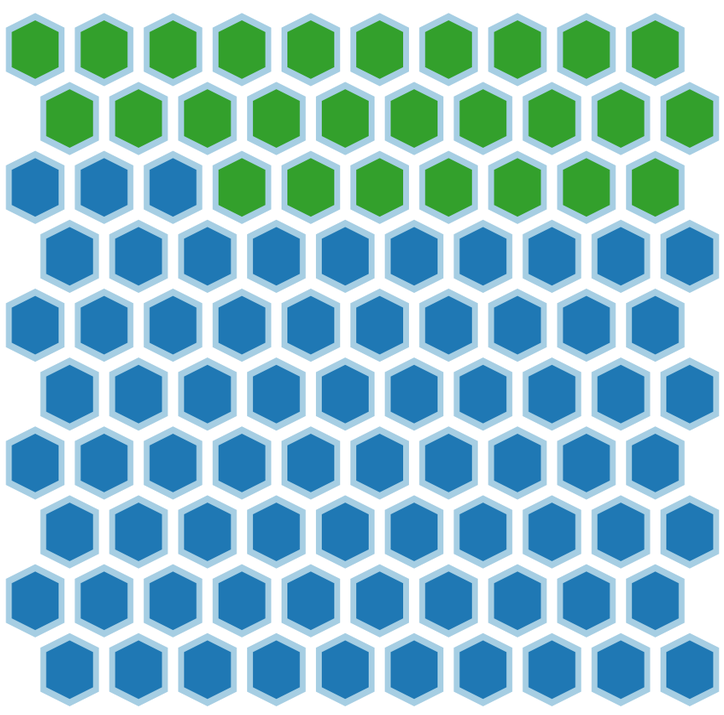
Histology:Adeno, Age:82, PackYears:32, Size:25
Stage:1b, Gender:Female, GD:Not GD, Recur:no

R1



R2





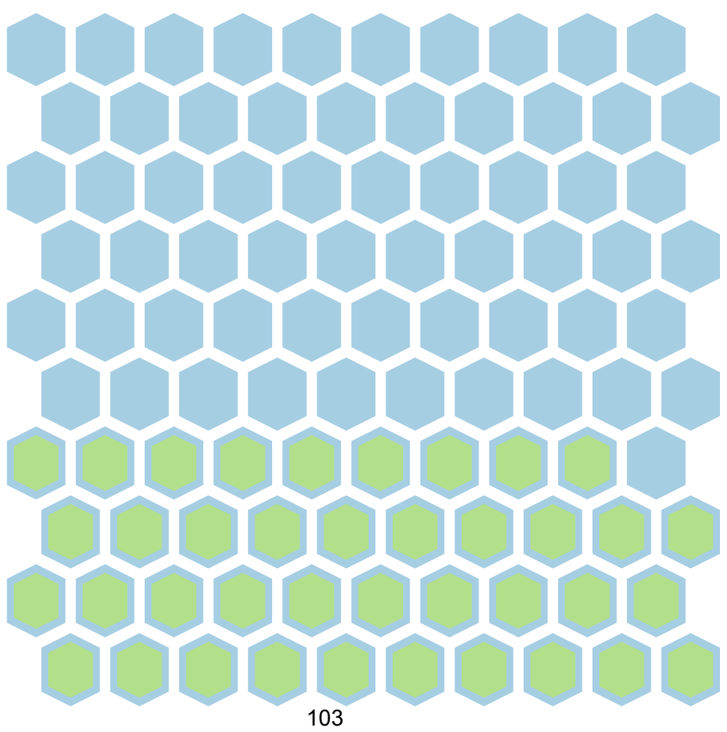
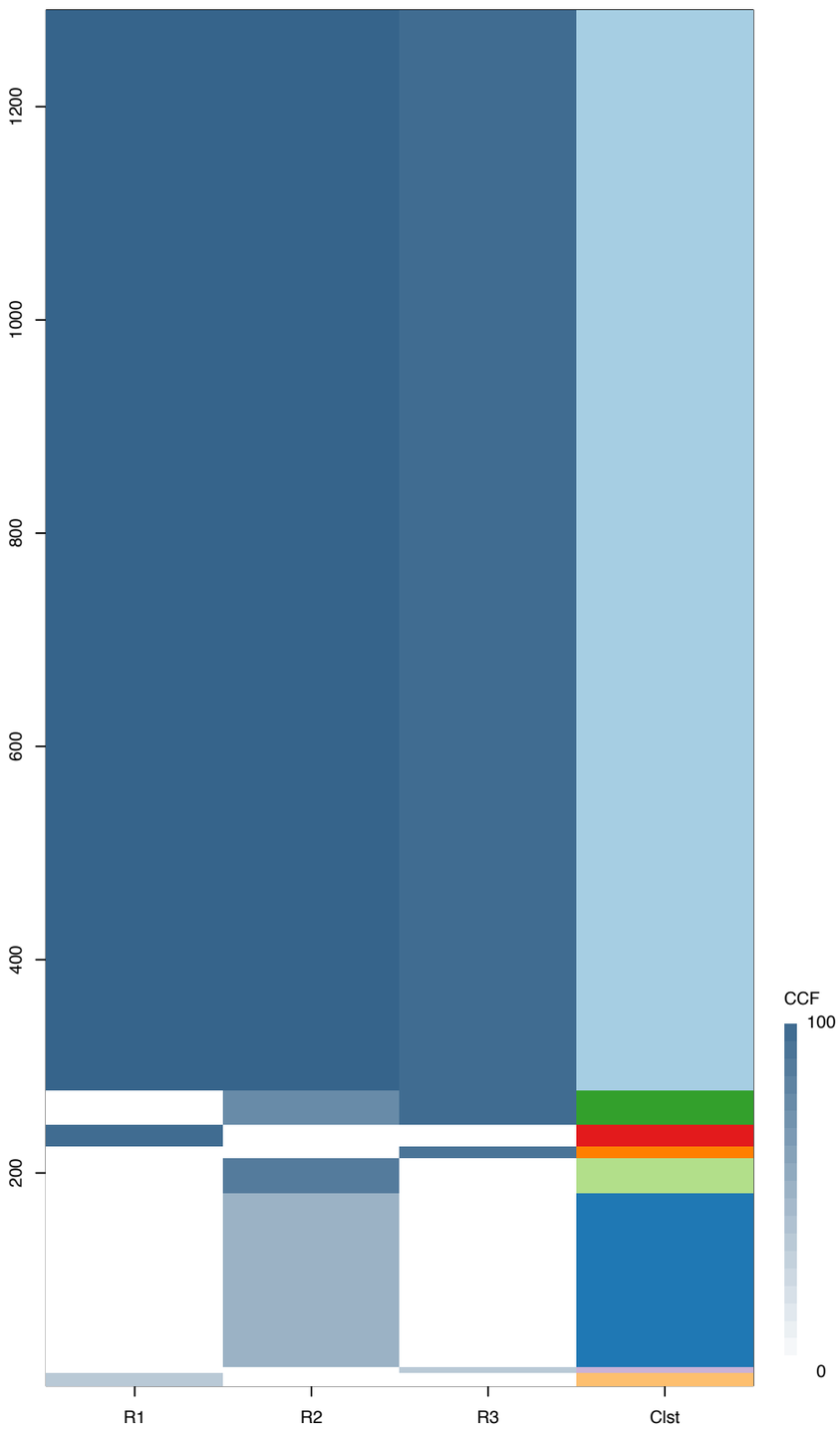
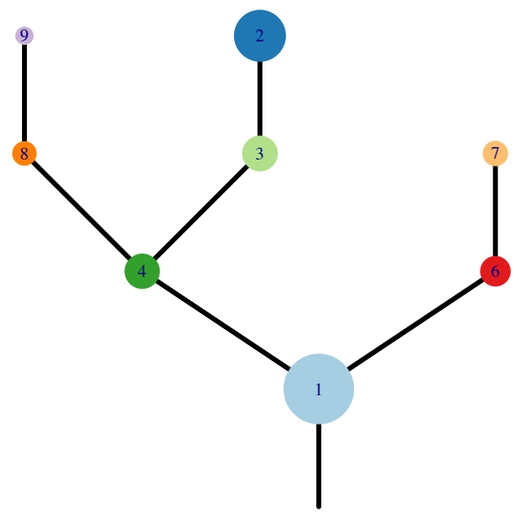


Fig.S12AM

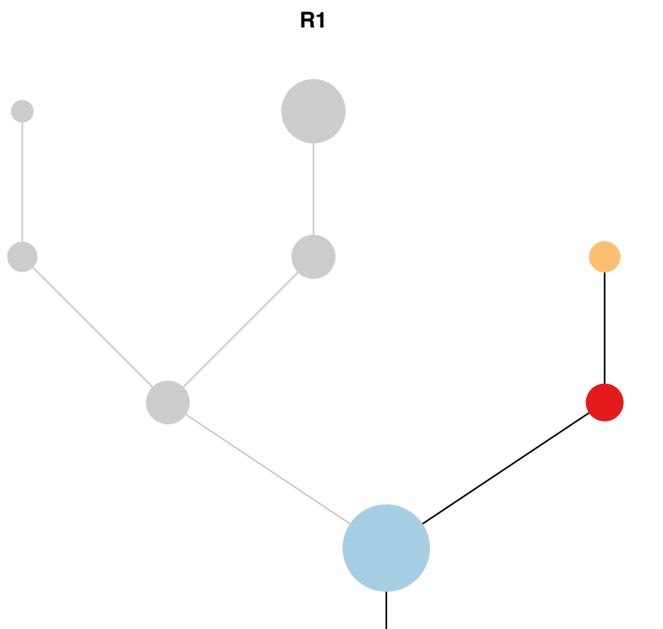


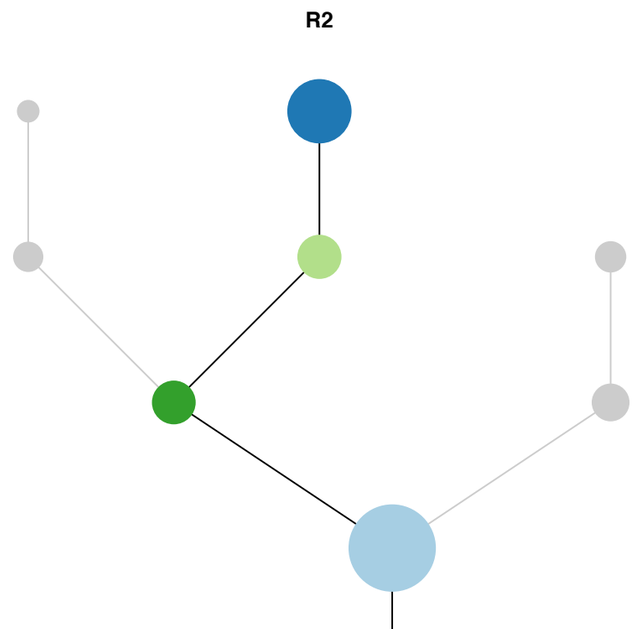
CRUK0039

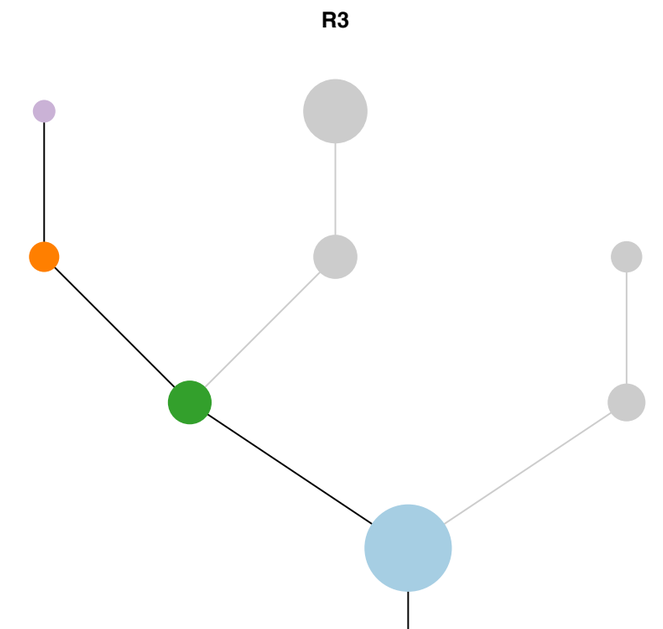


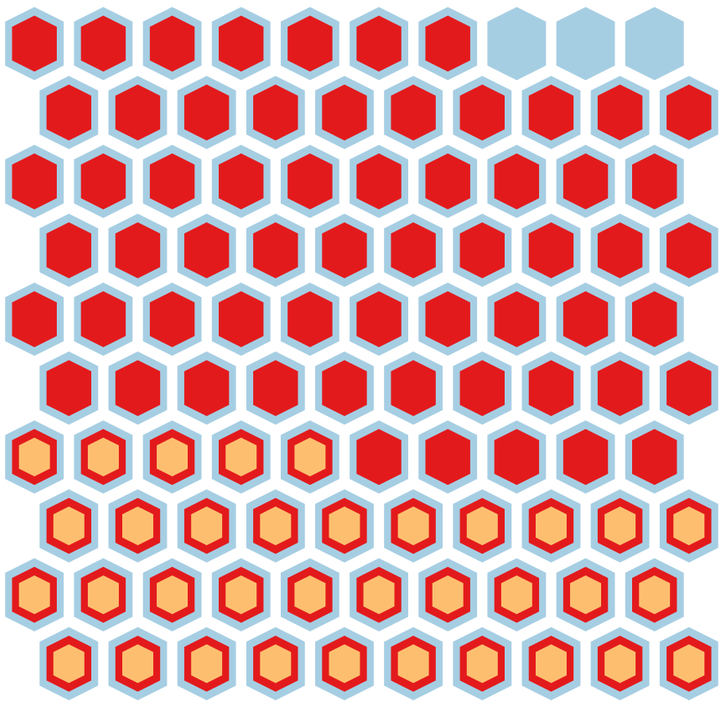
Histology:Adeno, Age:72, PackYears:52, Size:48
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

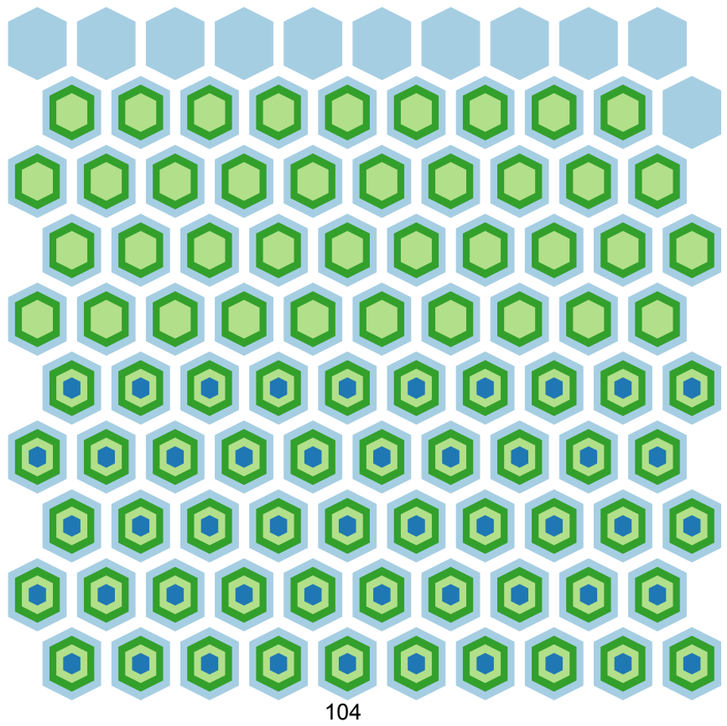
Gene	Cluster	Cytoband	Type
PHOX2B	1	4p13	SNV
KDM5A	1	12p13.33	Amp
ERC1	1	12p13.33	Amp
CCND2	1	12p13.32	Amp
ZNF384	1	12p13.31	Amp
ETV6	1	12p13.2	Amp
ATF7IP	1	12p13.1	Amp
ETNK1	1	12p12.1	Amp
KRAS	1	12p12.1	Amp
KRAS	1	12p12.1	SNV
PPFIBP1	1	12p11.23	Amp
GRIN2A	1	16p13.2	SNV
CMTR2	1	16q22.2	SNV
NF1	1	17q11.2	SNV
KRAS	2	12p12.1	SNV
FUBP1	?	1p31.1	SNV
NOTCH2	?	1p12	SNV











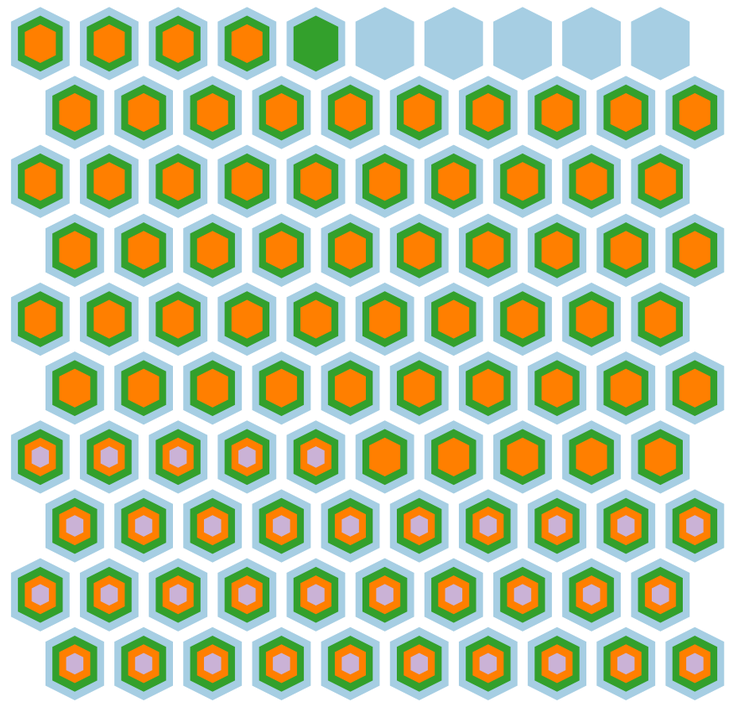
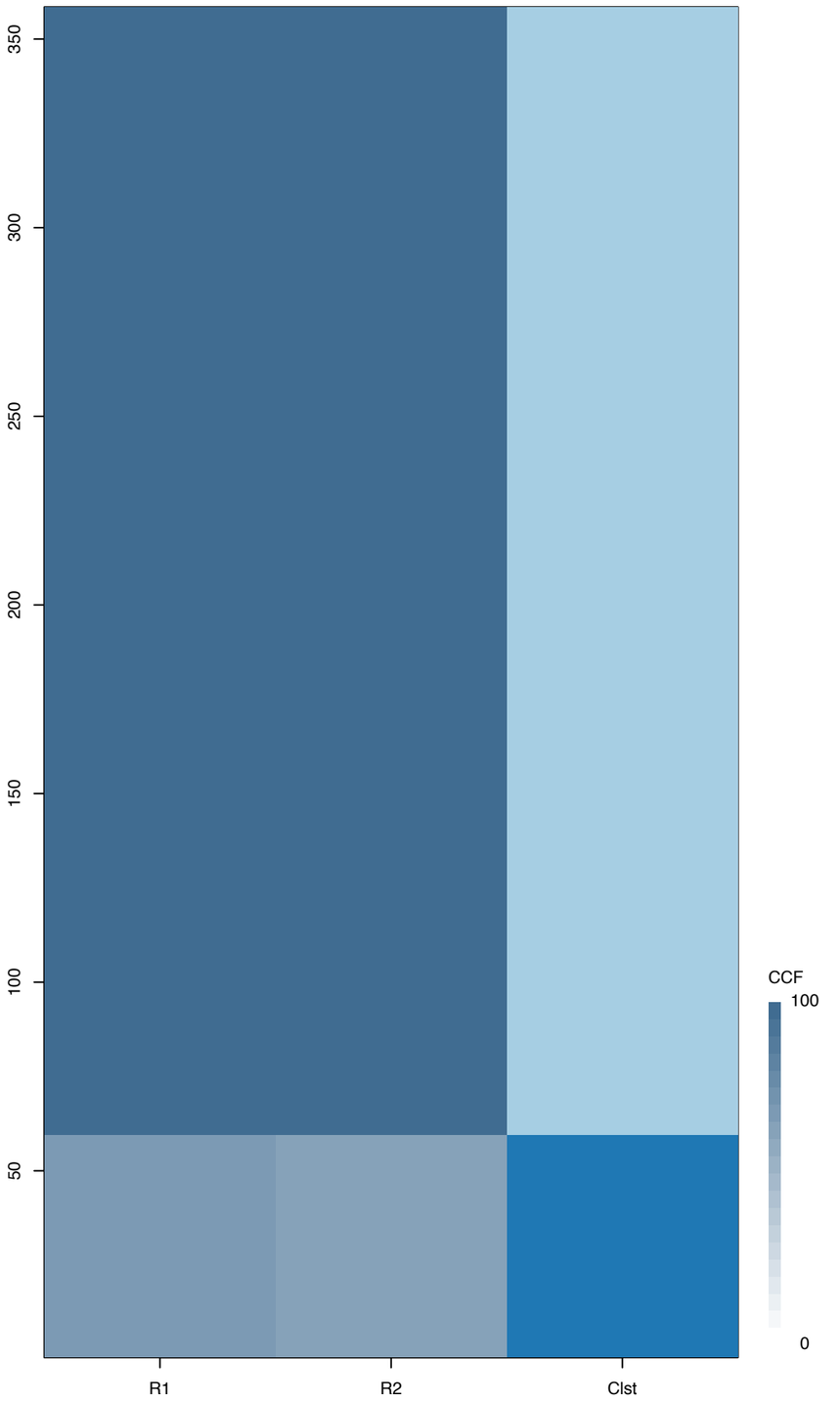
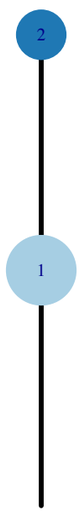


Fig.S12AN



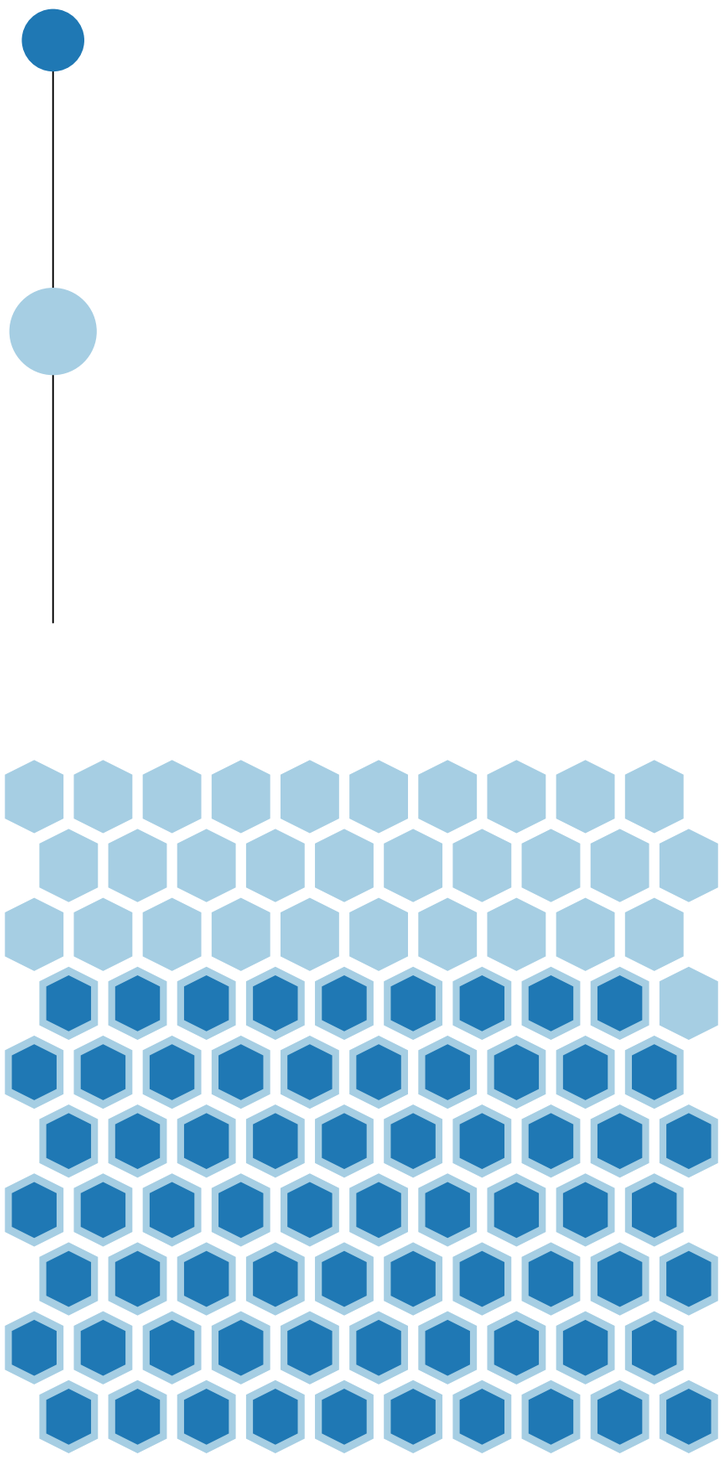
CRUK0040



Gene	Cluster	Cytoband	Type
RAD21	1	8q24.11	SNV
GATA3	1	10p14	SNV
KRAS	1	12p12.1	SNV
NCOR1	1	17p12	SNV

Histology:Adeno, Age:63, PackYears:40, Size:14
Stage:1a, Gender:Female, GD:Not GD, Recur:no

R1



R2

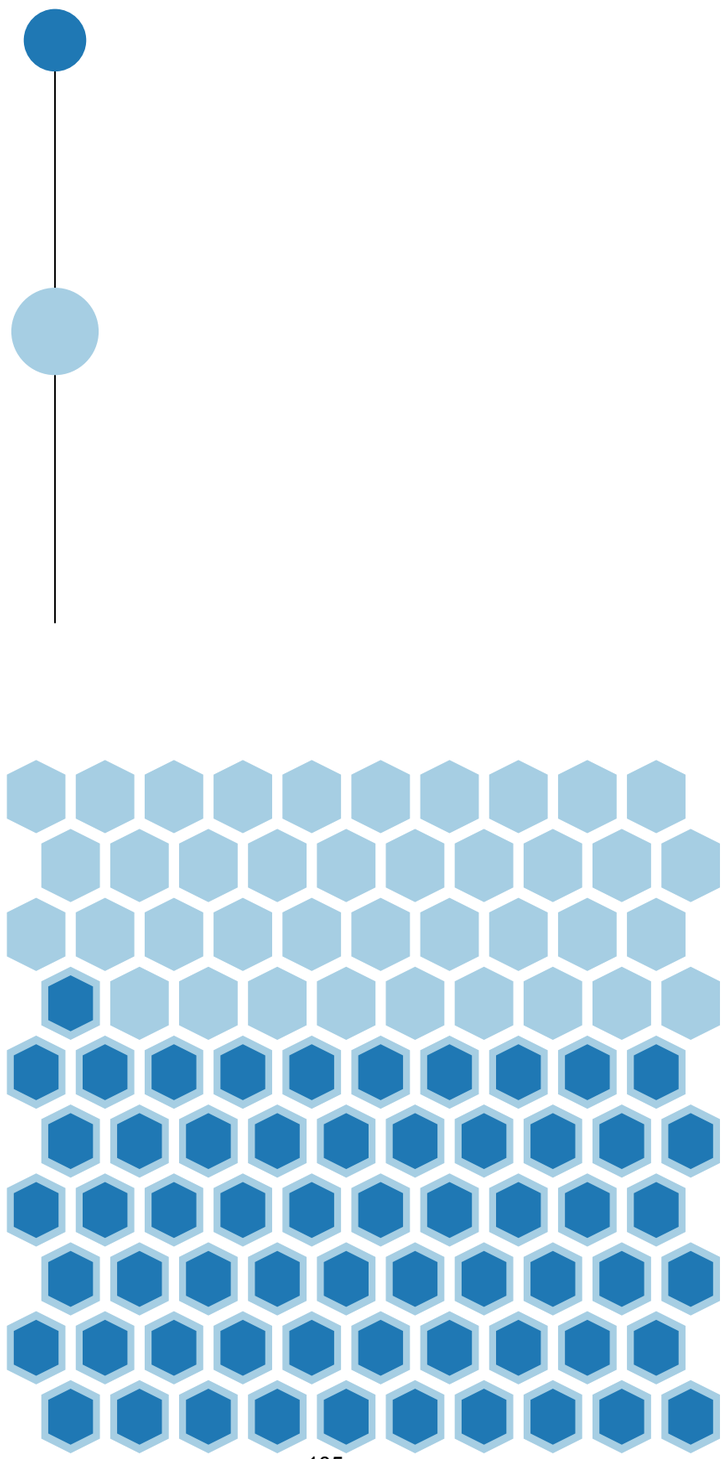
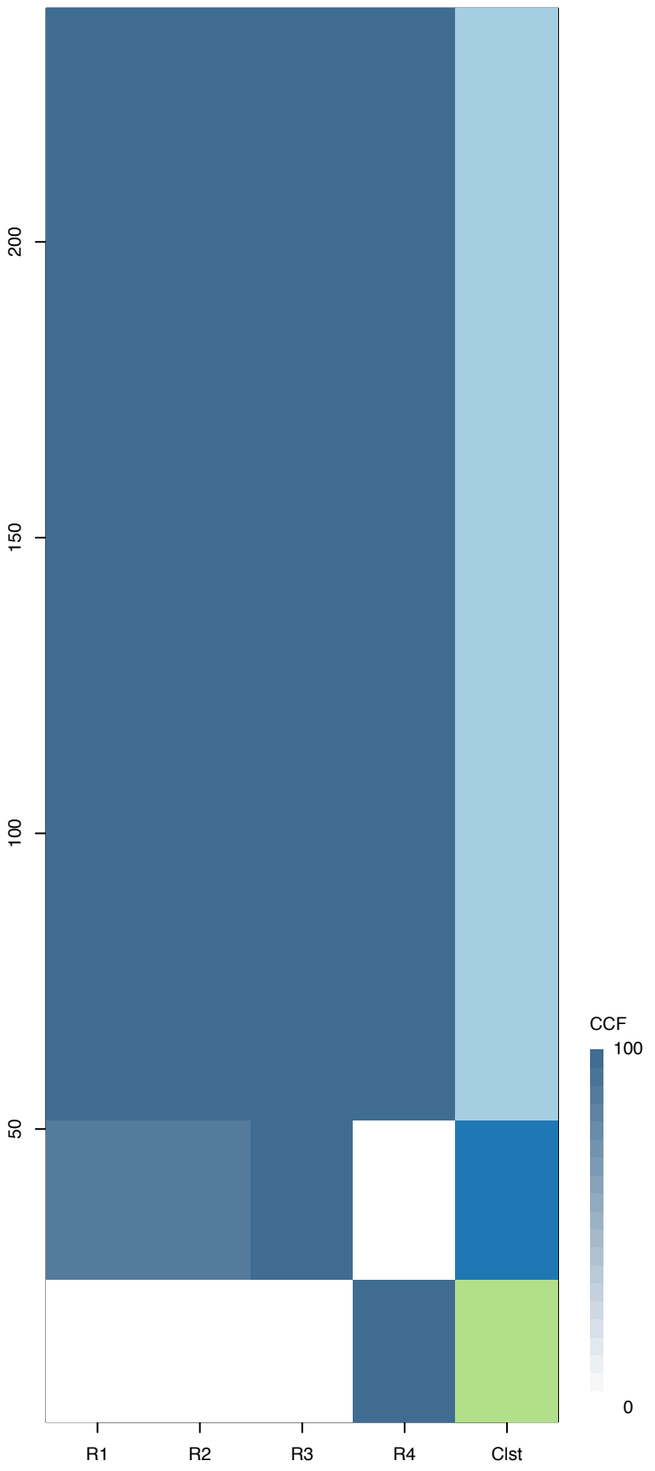
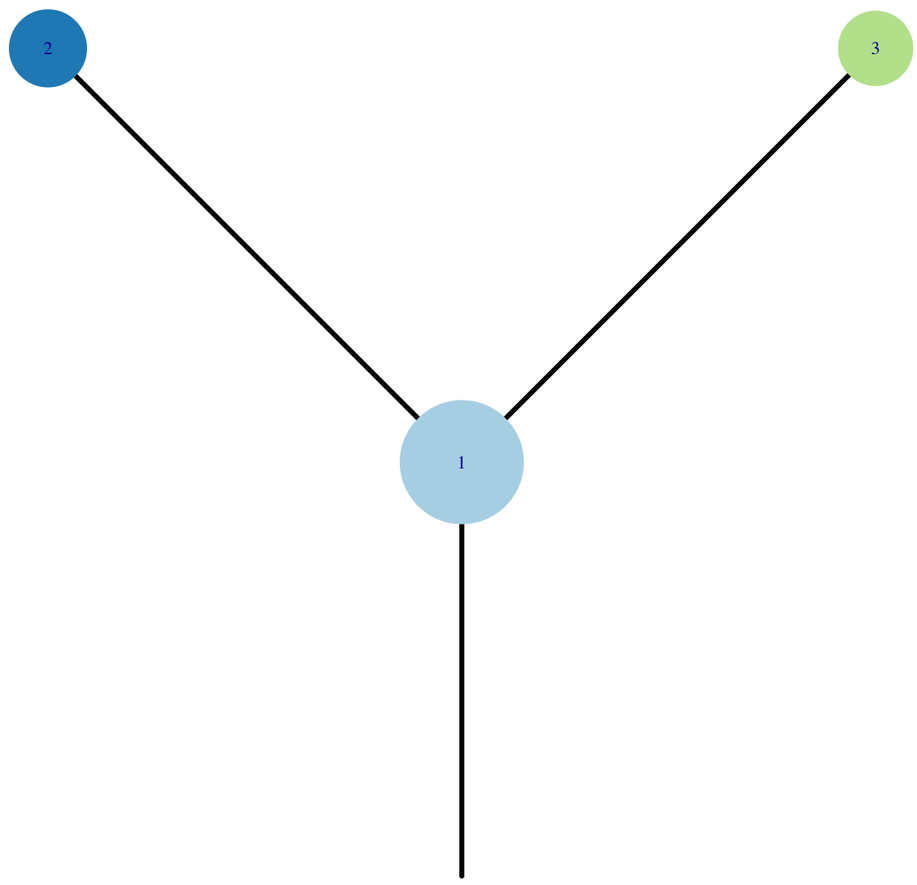


Fig.S12AO



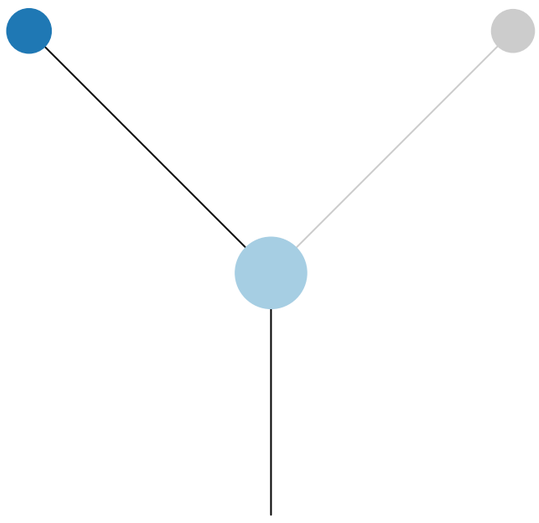
CRUK0041



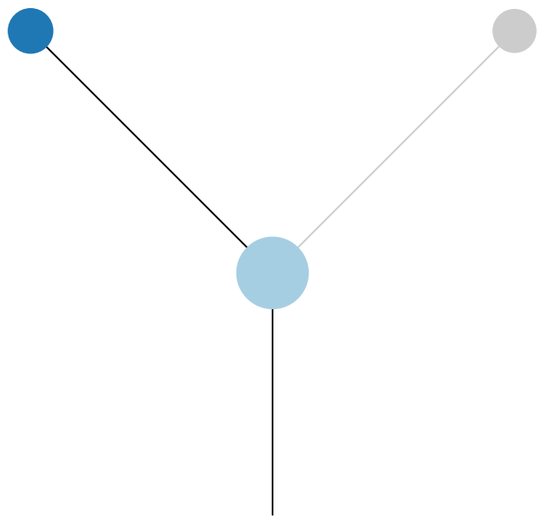
Histology:Adeno, Age:73, PackYears:62, Size:35
Stage:1b, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
TERT	1	5p15.33	Amp
EGFR	1	7p11.2	Amp
AKAP9	1	7q21.2	Amp
BRAF	1	7q34	SNV
HEY1	1	8q21.13	Amp
RUNX1T1	1	8q21.3	Amp
COX6C	1	8q22.2	Amp
RSPO2	1	8q23.1	Amp
KDM5A	1	12p13.33	Amp
CCNE1	1	19q12	Amp
EIF3E	?	8q23.1	Amp

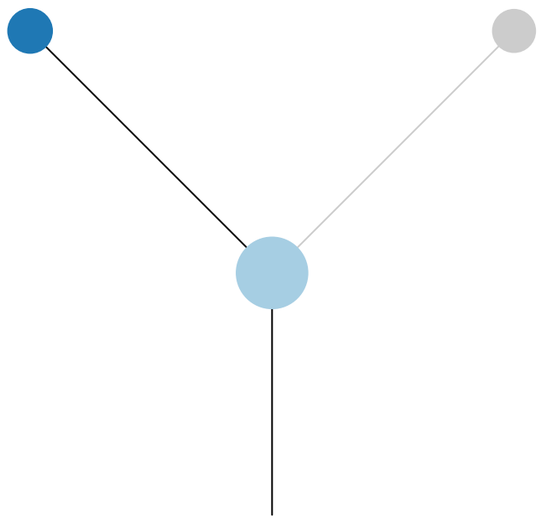
R1



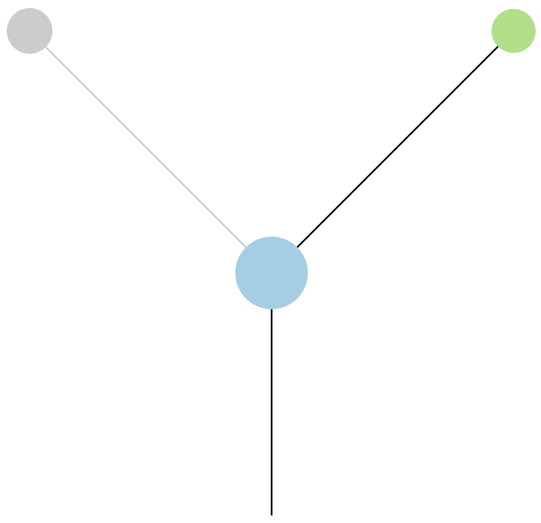
R2



R3



R4



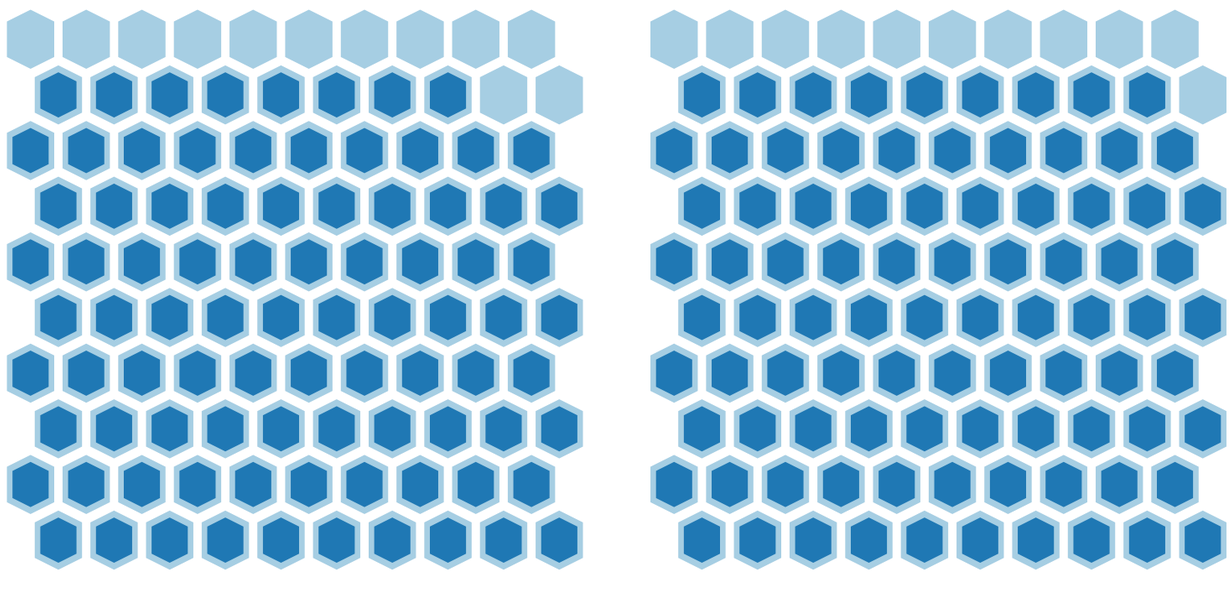
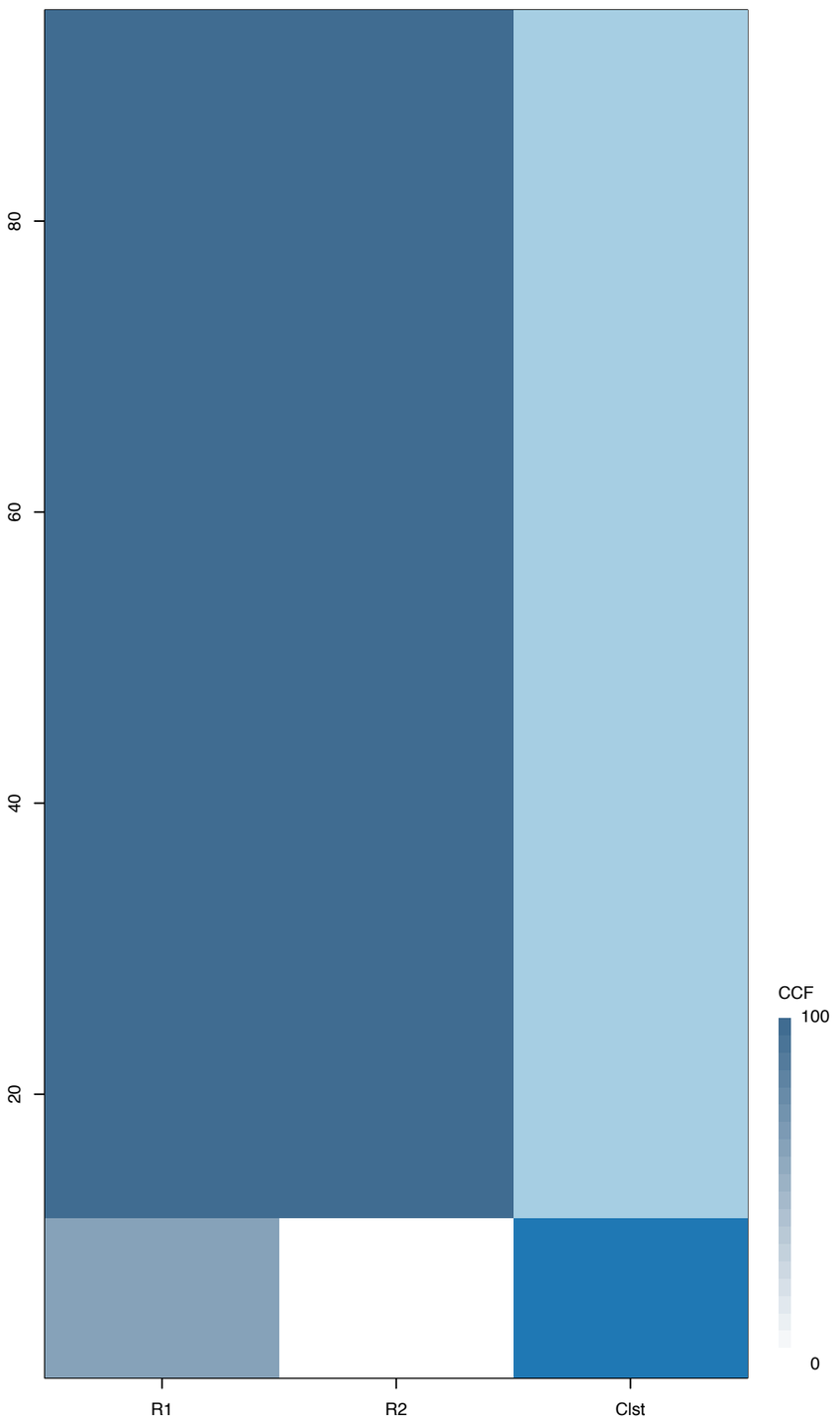
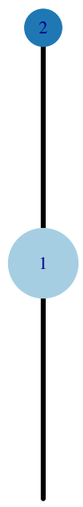


Fig.S12AP



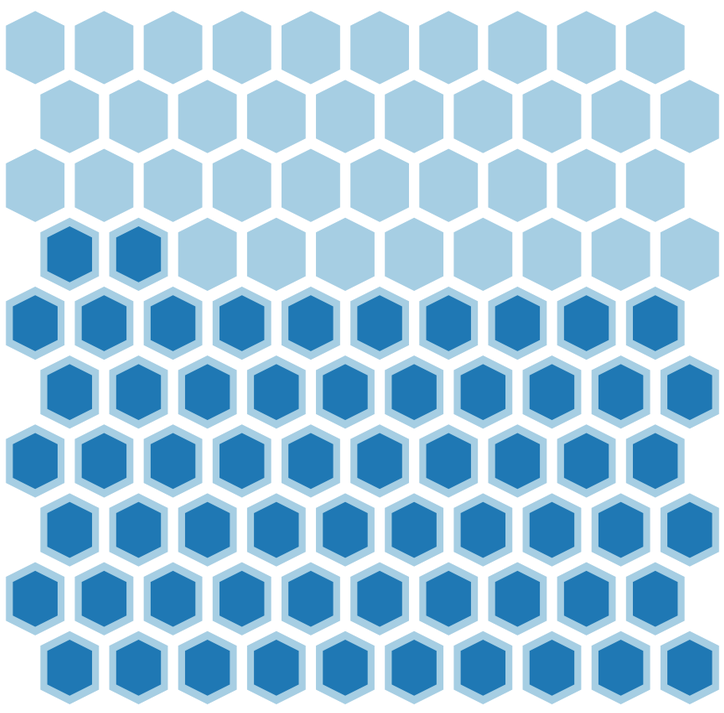
CRUK0043



Histology:Adeno, Age:70, PackYears:4, Size:20
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
MET	1	7q31.2	SNV
NKX2−1	1	14q13.3	Amp
FOXA1	1	14q21.1	Amp

R1



R2

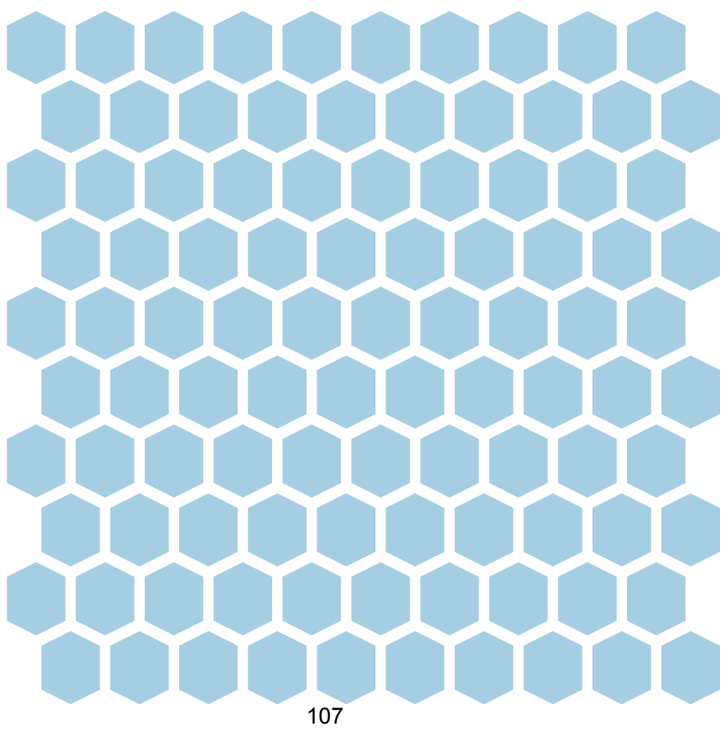
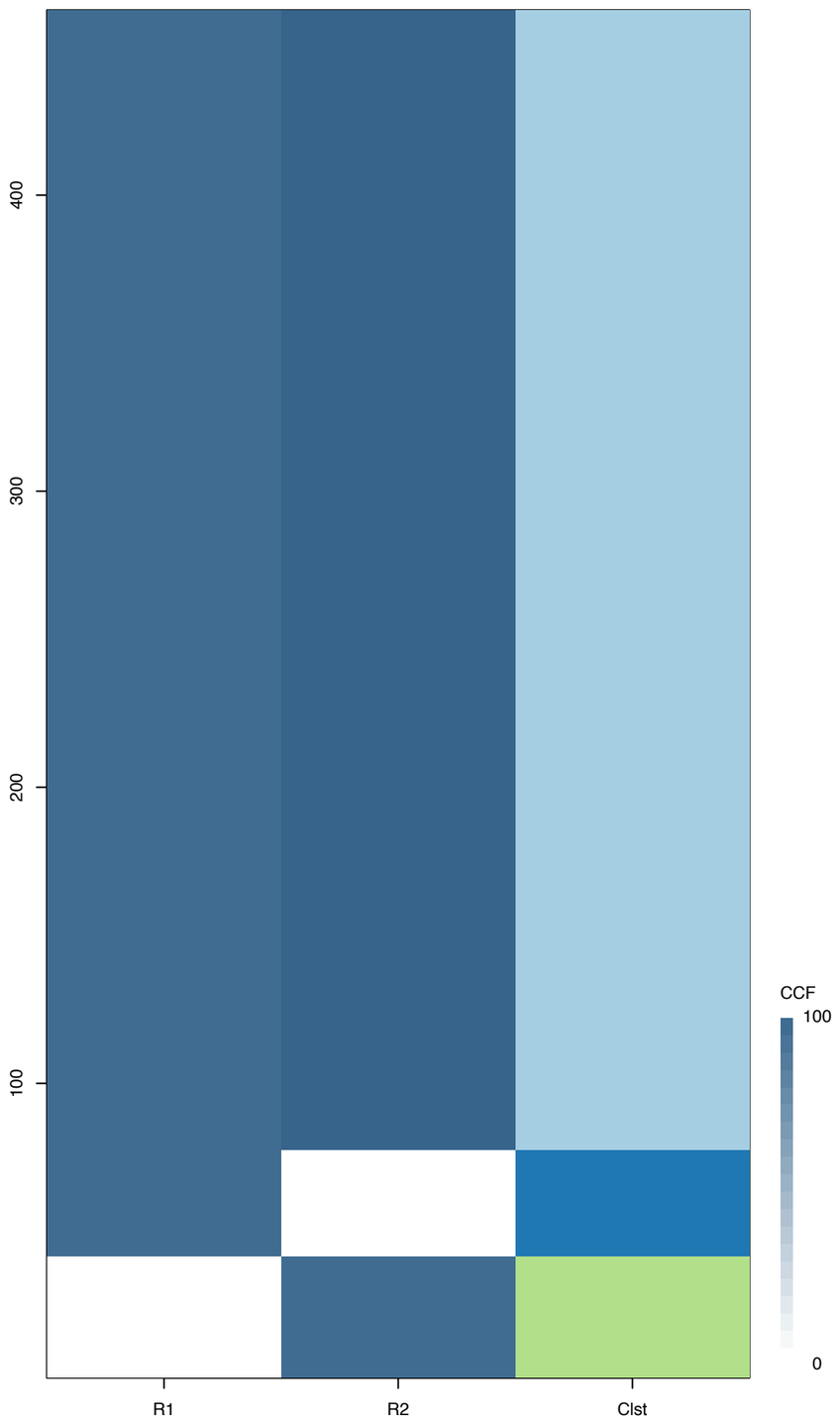
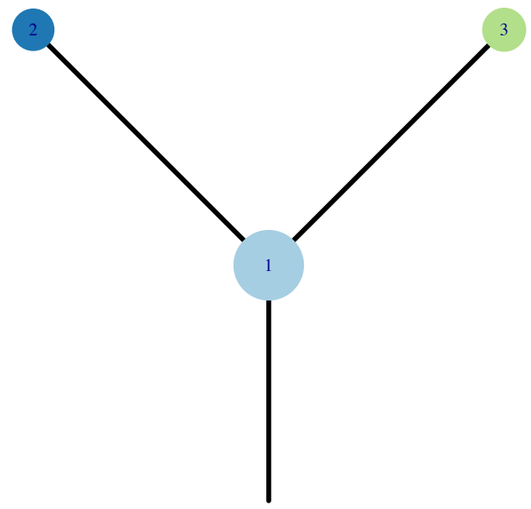
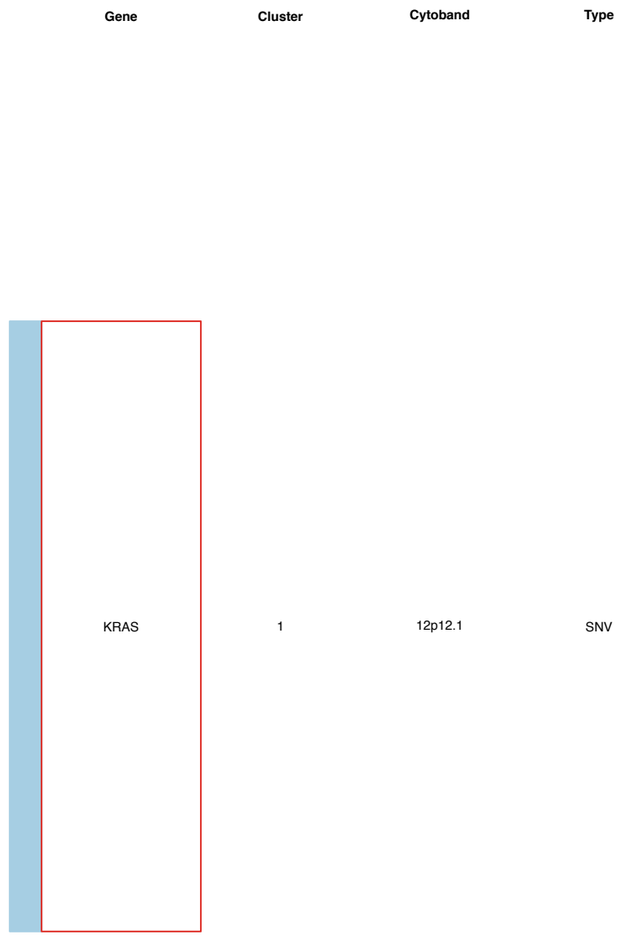


Fig.S12AQ



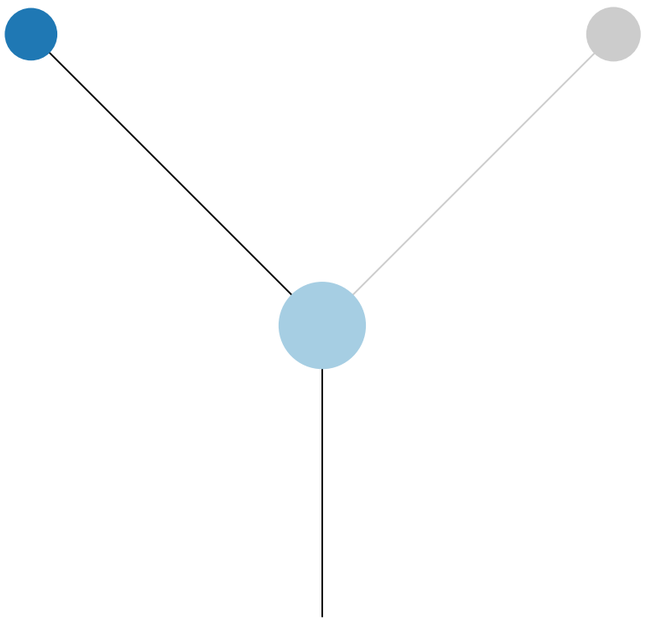
CRUK0044

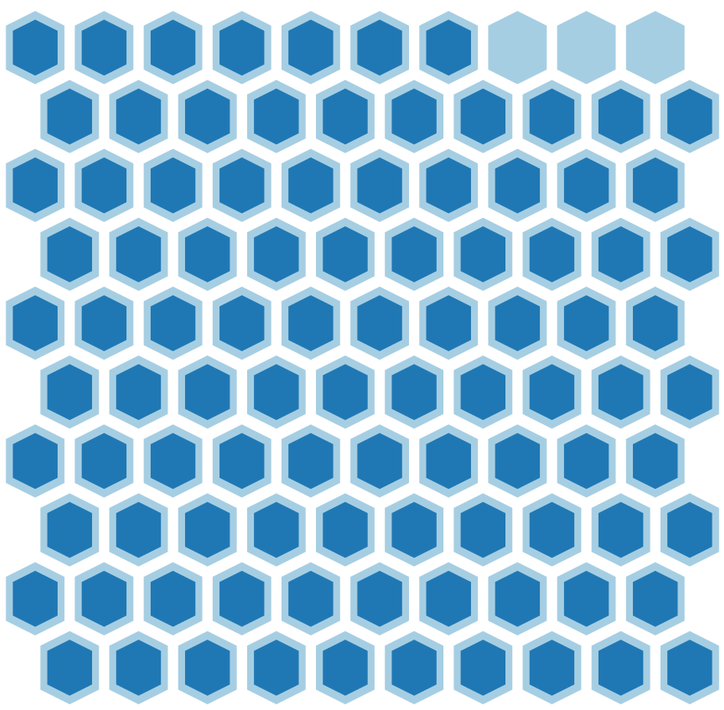




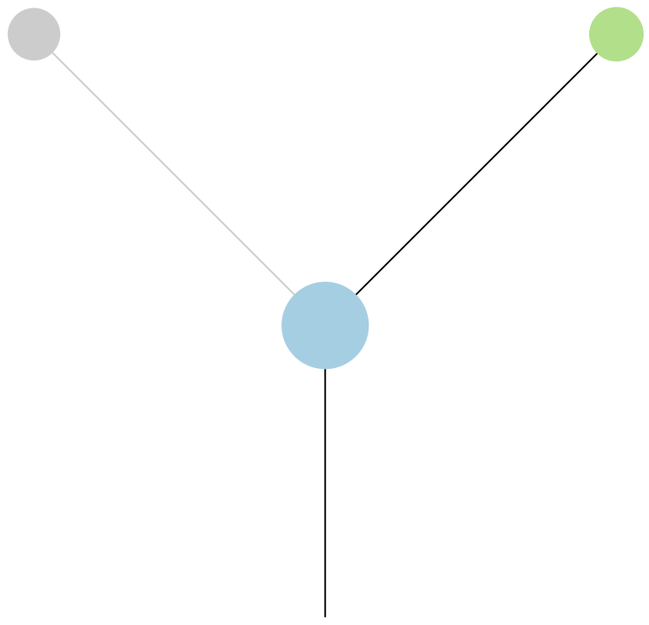
Histology:Adeno, Age:60, PackYears:33.75, Size:23
Stage:1a, Gender:Female, GD:Not GD, Recur:yes

R1





R2



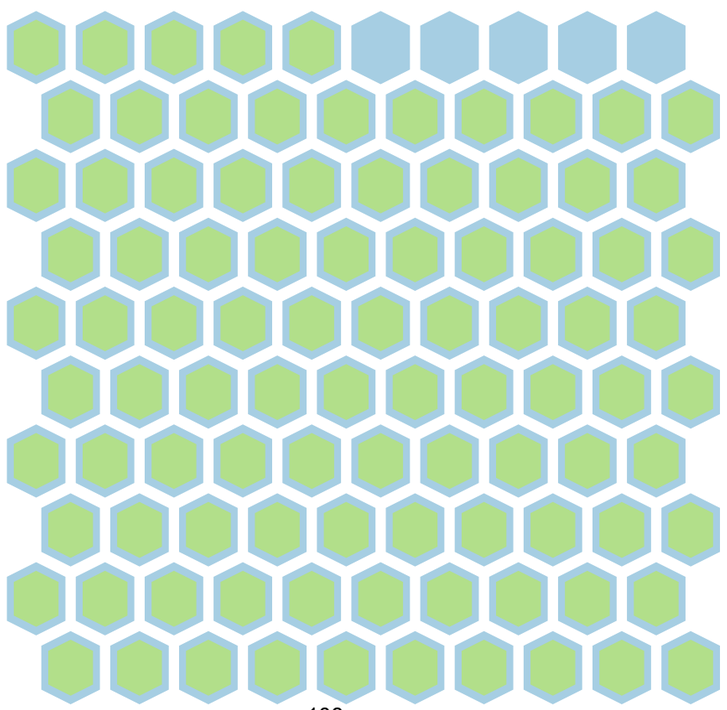
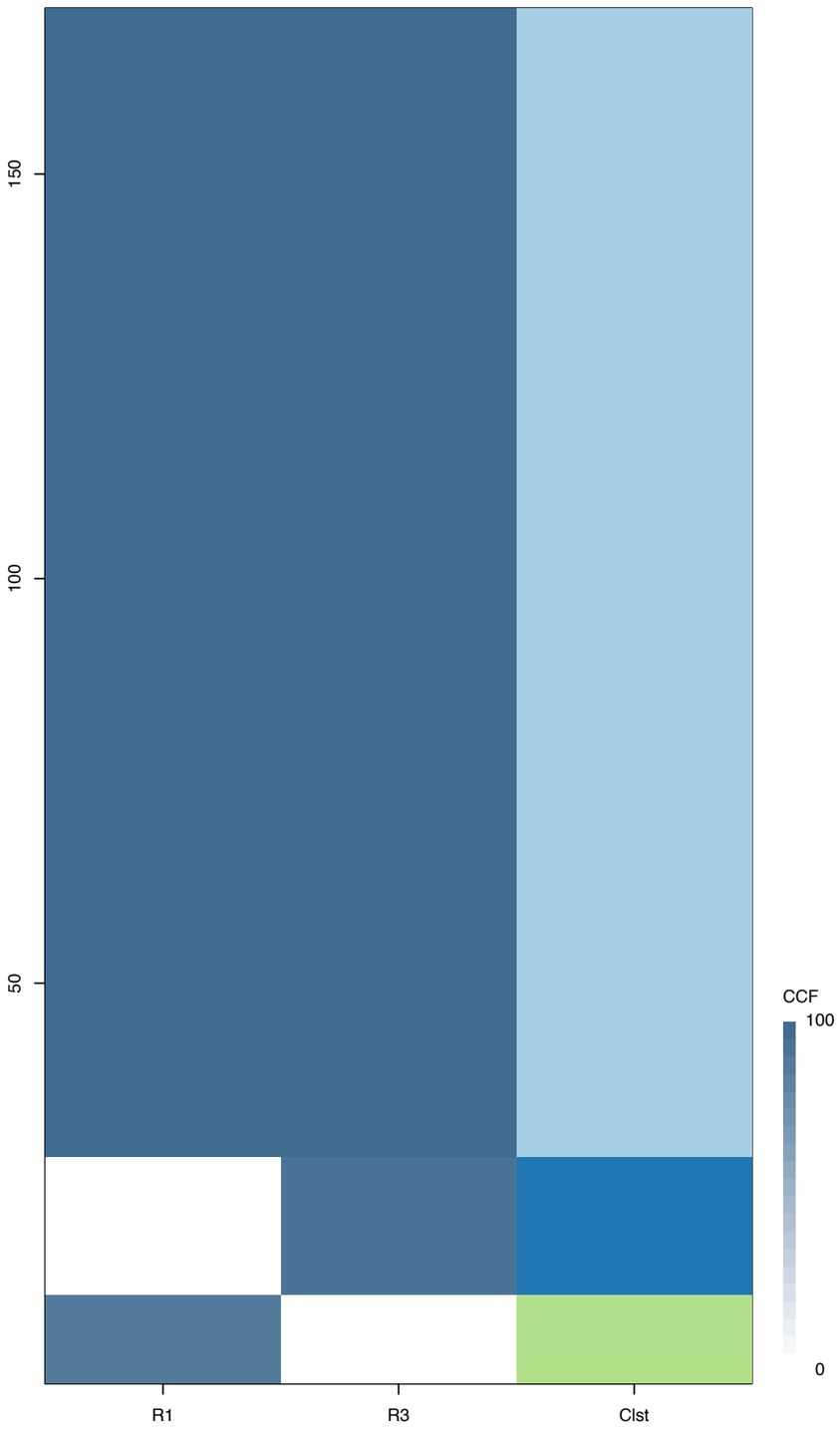
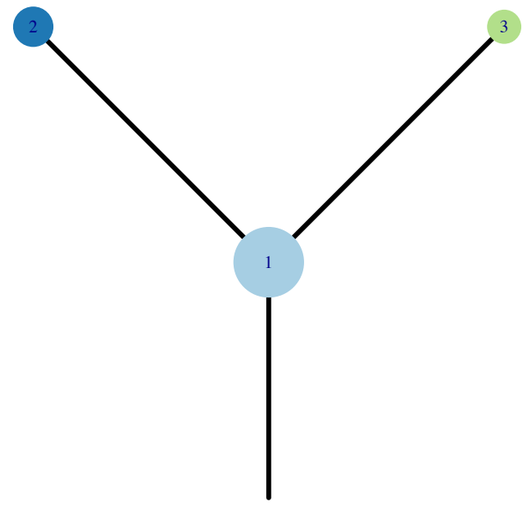


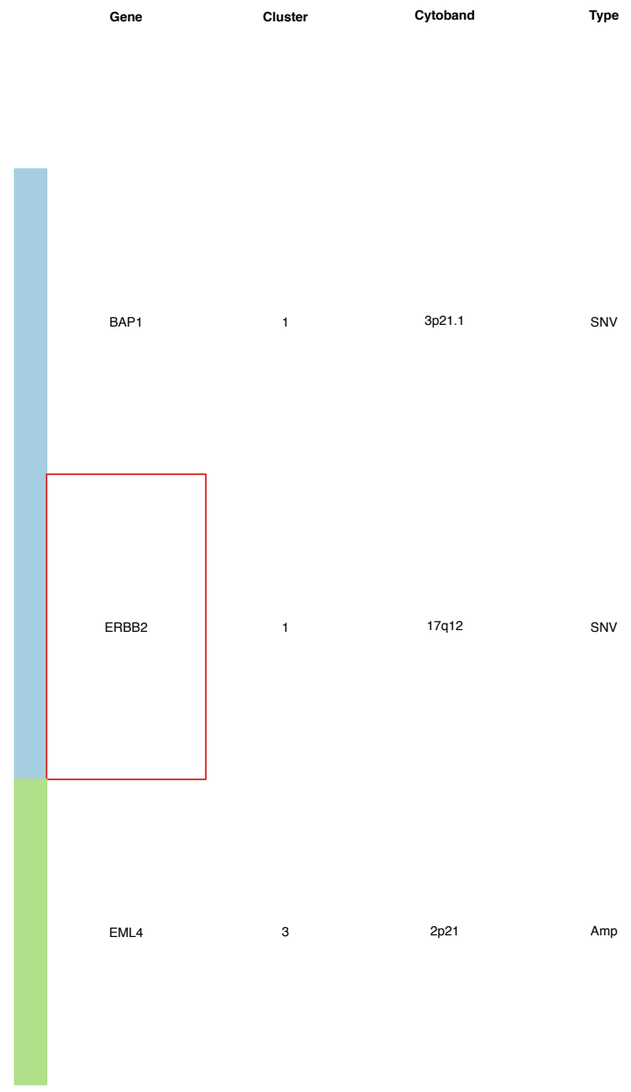
Fig.S12AR



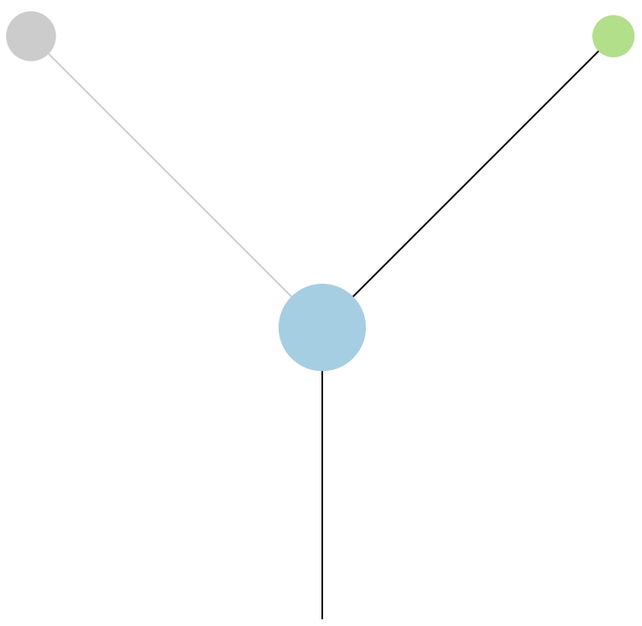
CRUK0045



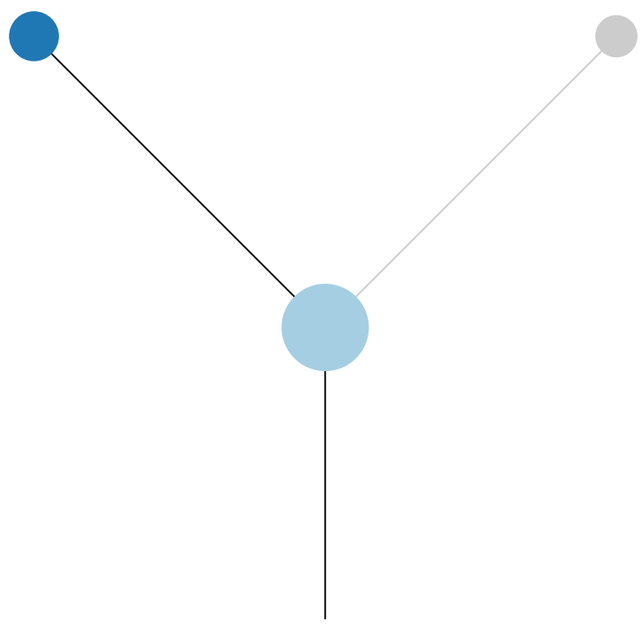
Histology:Adeno, Age:87, PackYears:11, Size:48
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

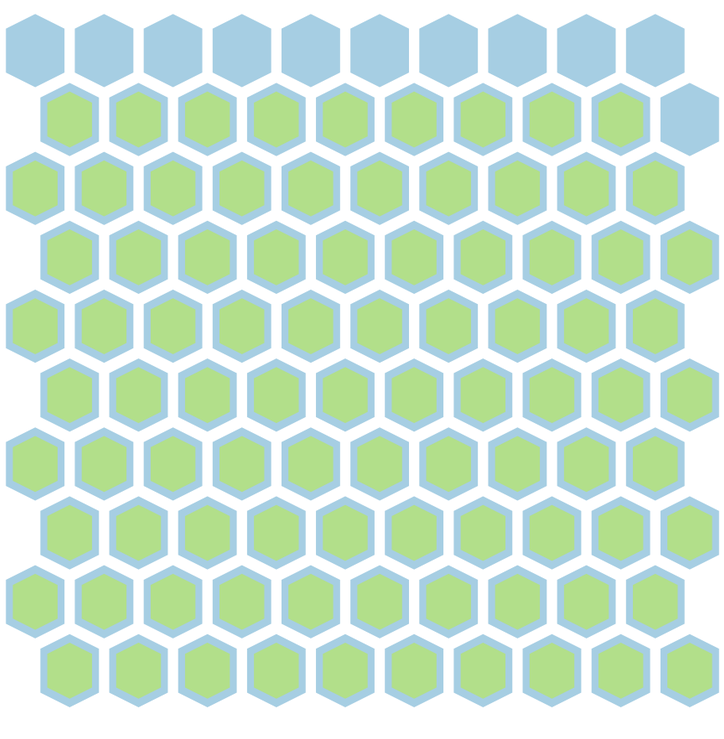


R1



R3





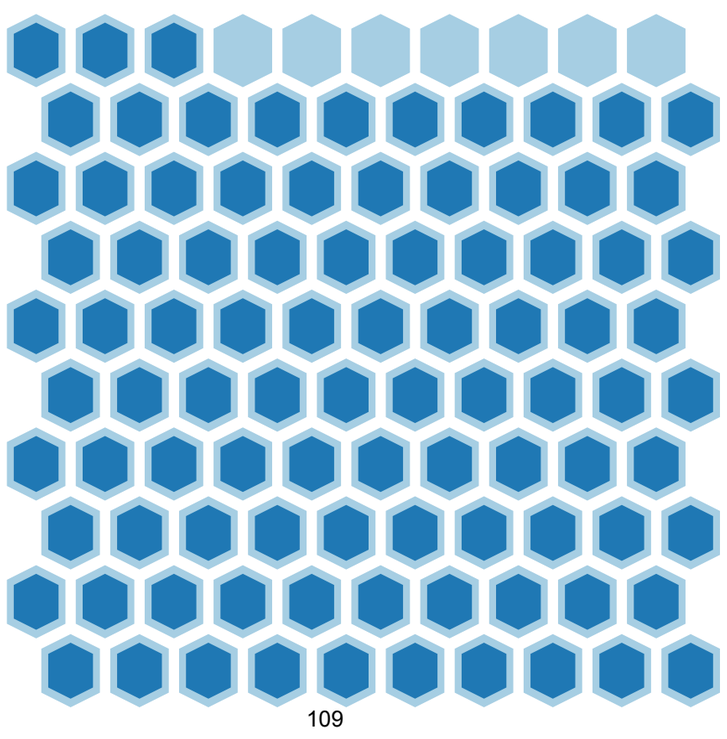
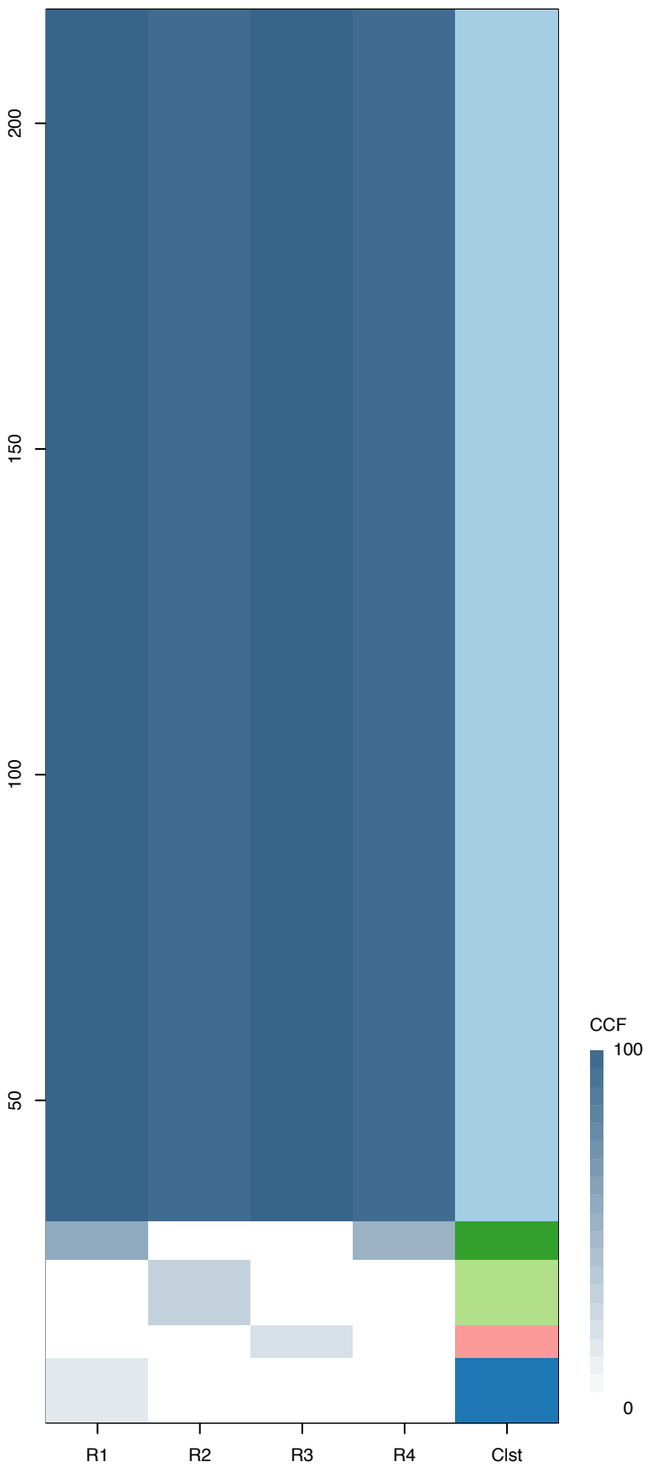
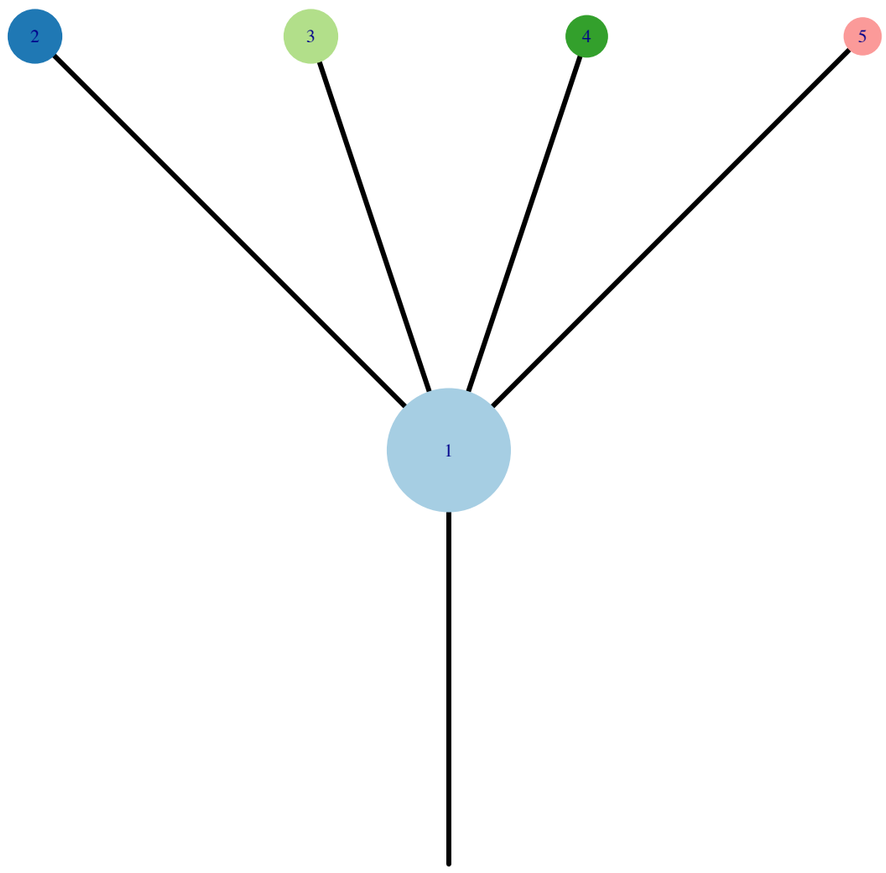


Fig.S12AS



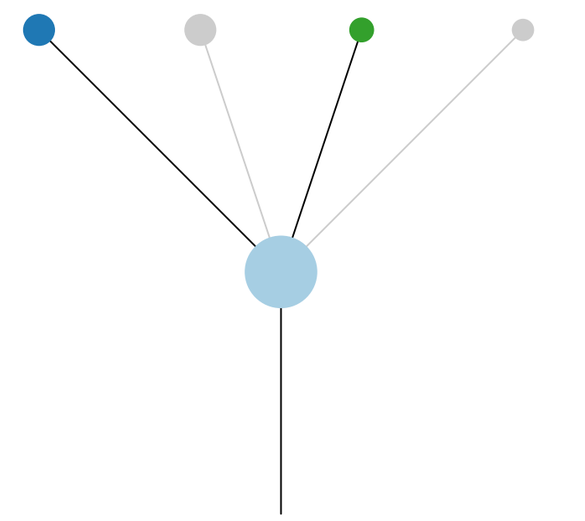
CRUK0046_A



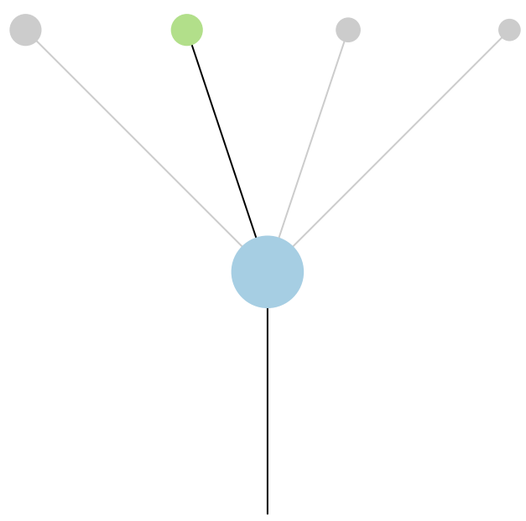
Histology:Adeno, Age:63, PackYears:15, Size:60
Stage:2a, Gender:Female, GD:Not GD, Recur:no

Gene	Cluster	Cytoband	Type
RHOH	1	4p14	Amp
APC	1	5q22.2	SNV
KEAP1	1	19p13.2	SNV
STK11	?	19p13.3	SNV

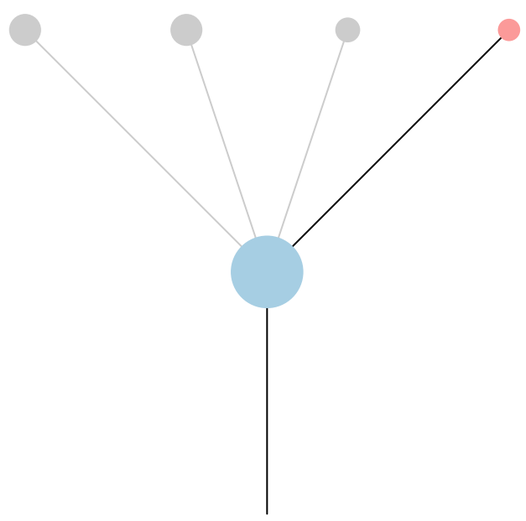
R1



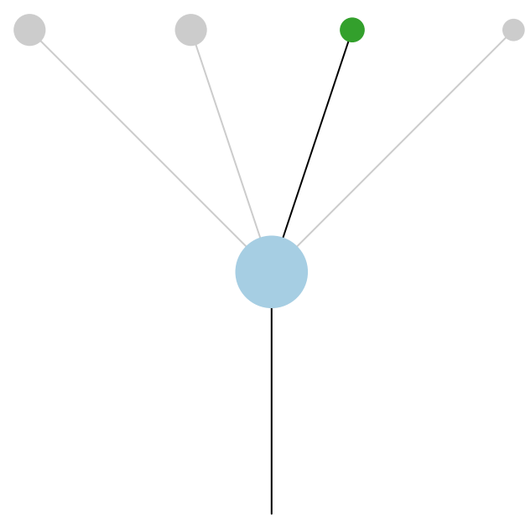
R2

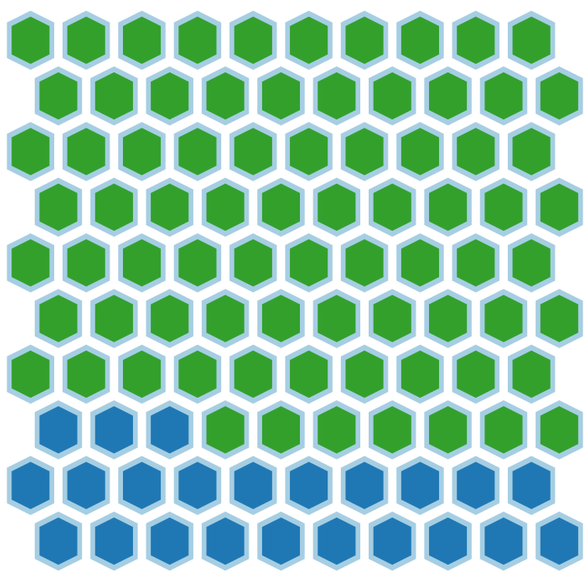


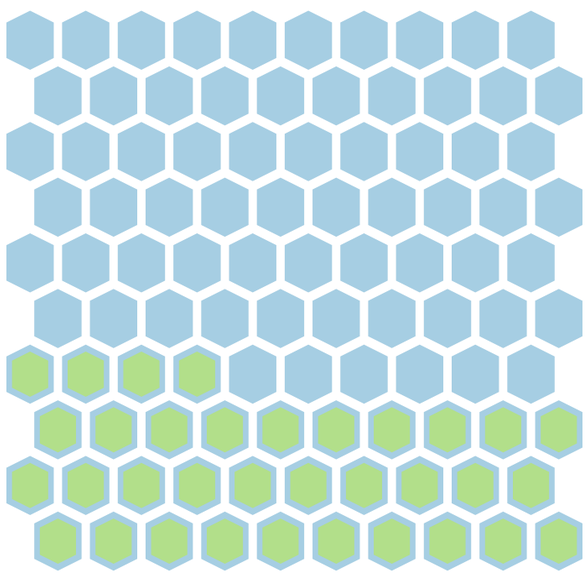
R3

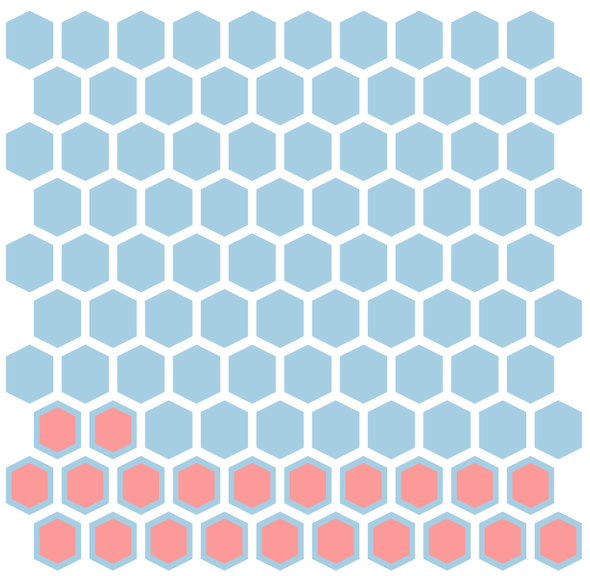


R4









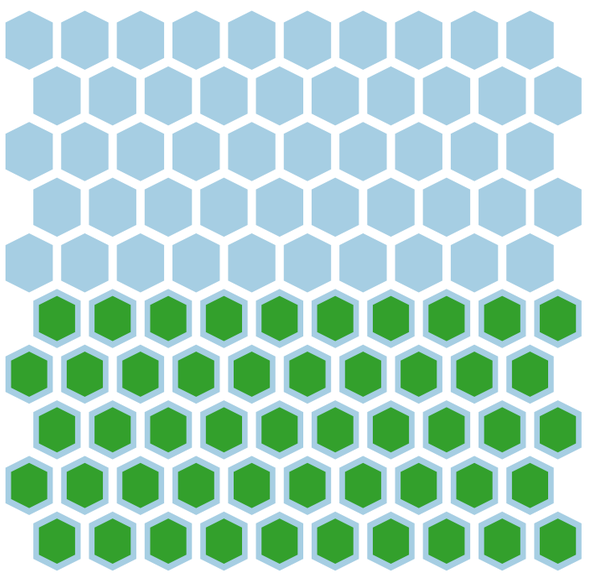
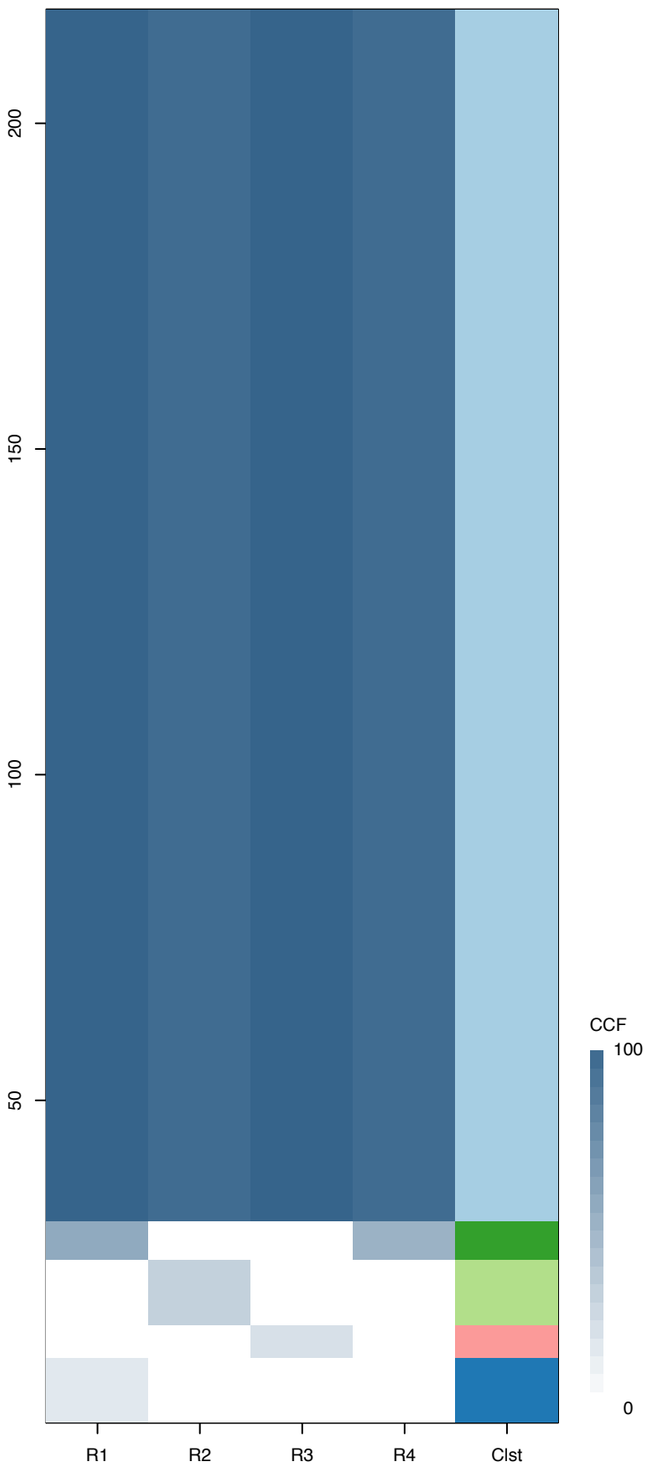
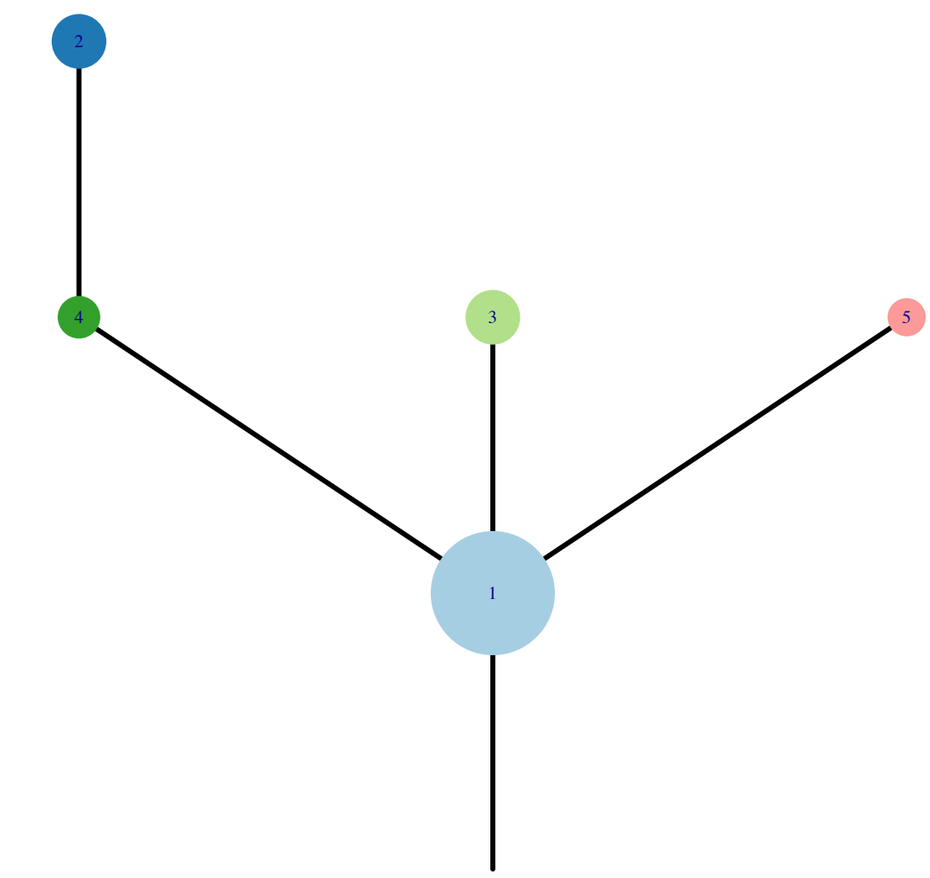


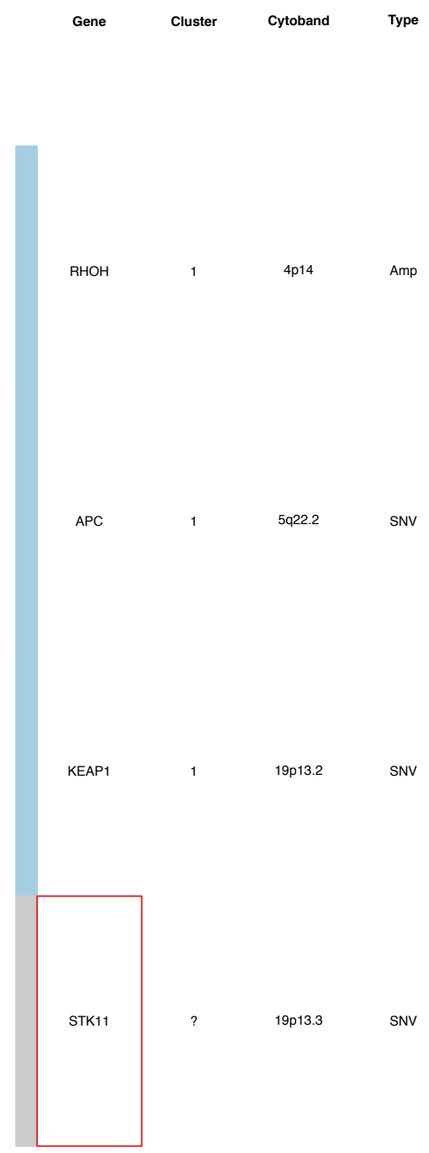
Fig.S12AS



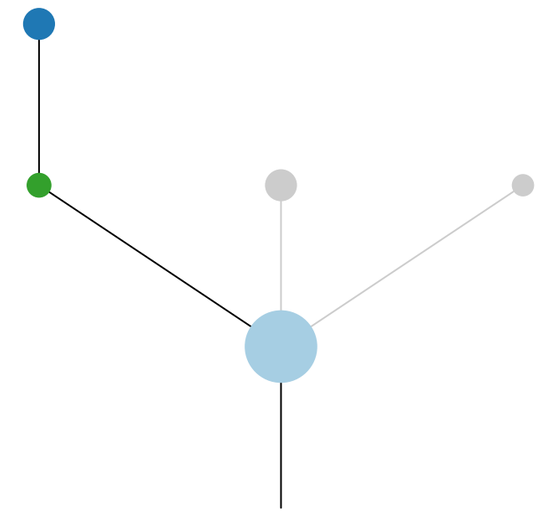
CRUK0046_B



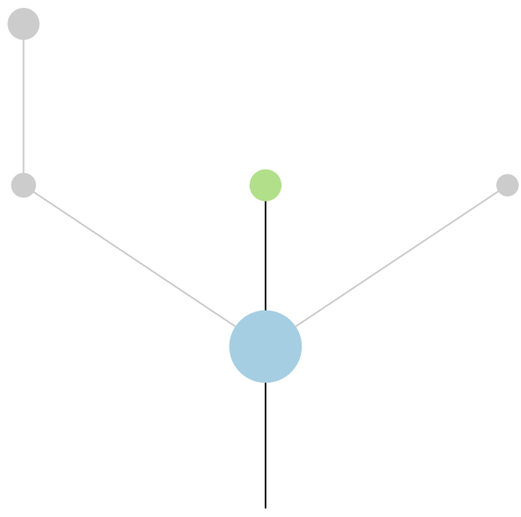
Histology:Adeno, Age:63, PackYears:15, Size:60
Stage:2a, Gender:Female, GD:Not GD, Recur:no



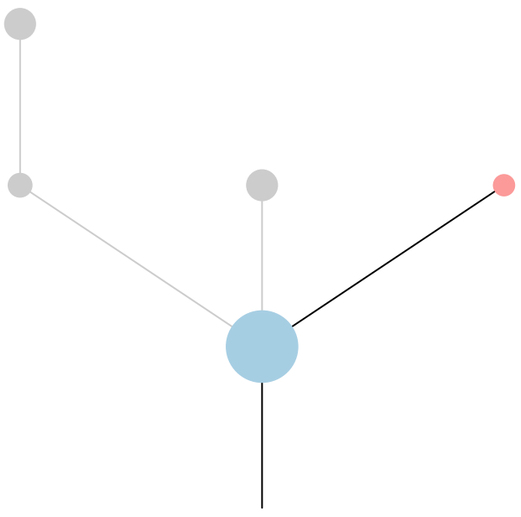
R1



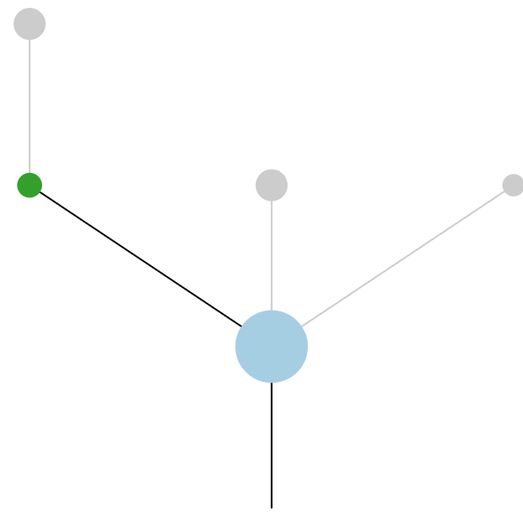
R2



R3



R4



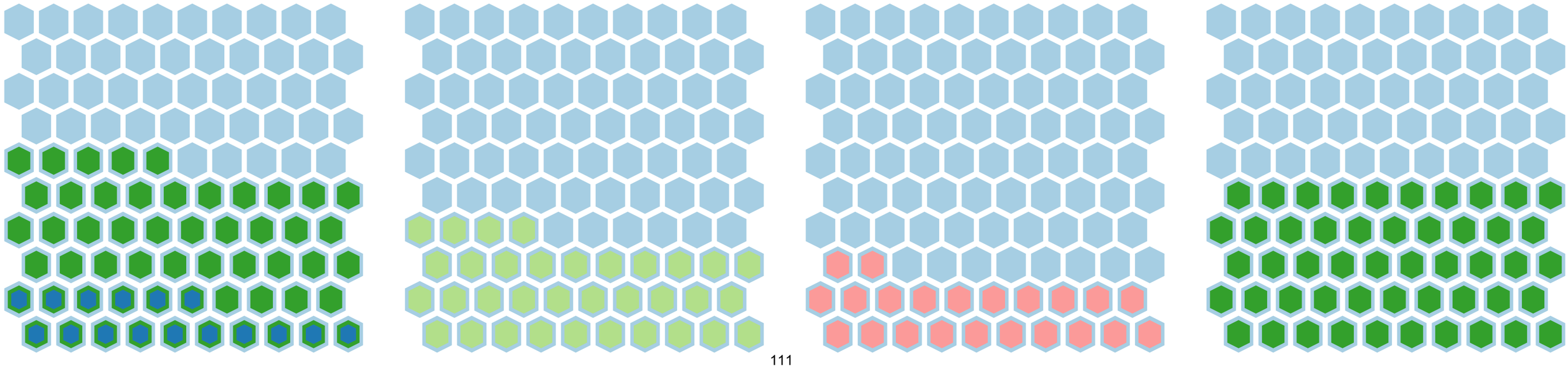
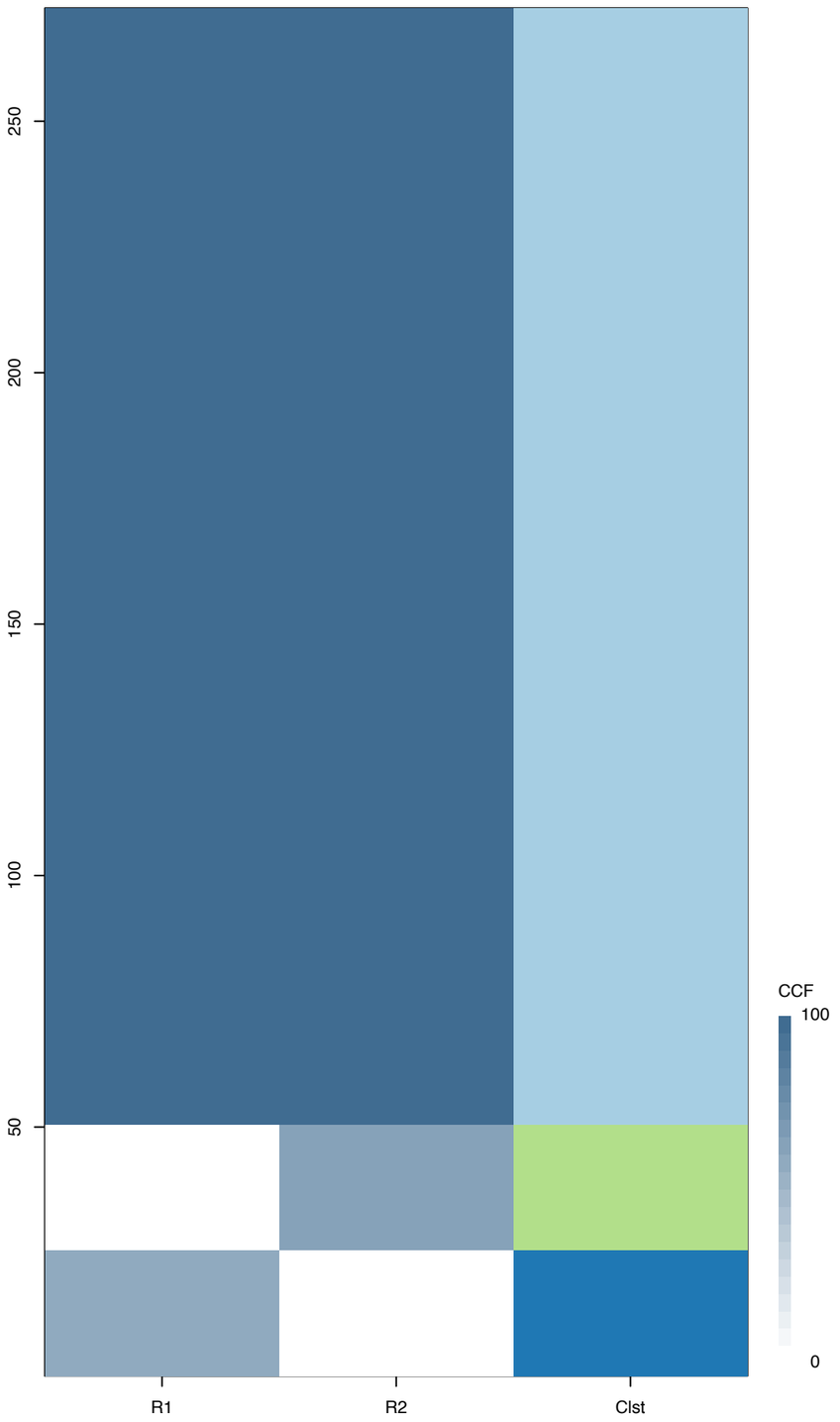
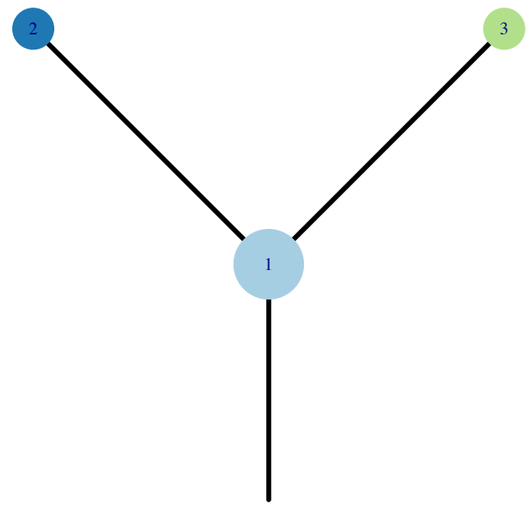


Fig.S12AT



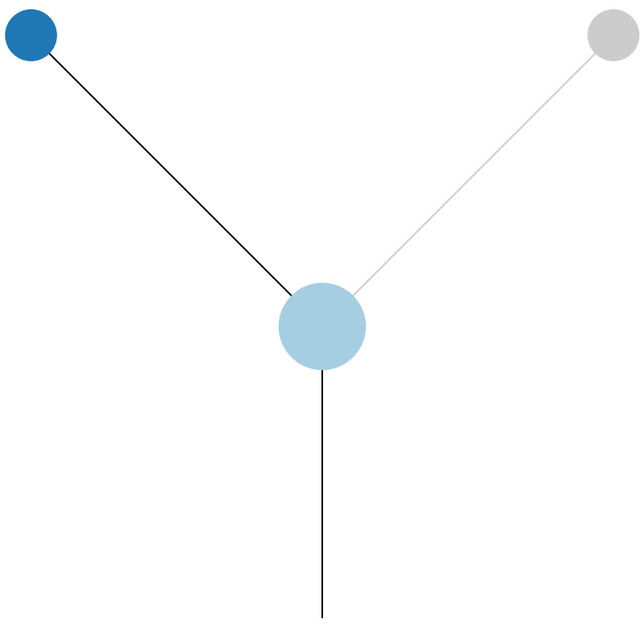
CRUK0047



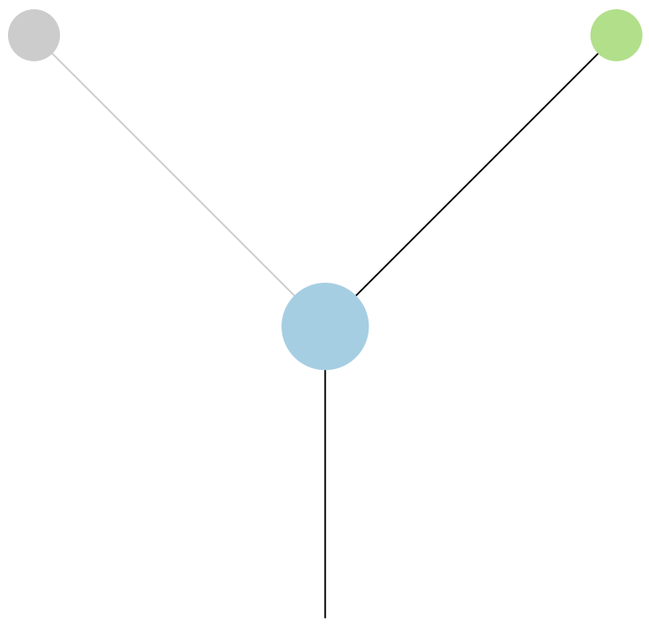
Histology:Adeno, Age:73, PackYears:33.75, Size:21
Stage:1a, Gender:Female, GD:Not GD, Recur:no

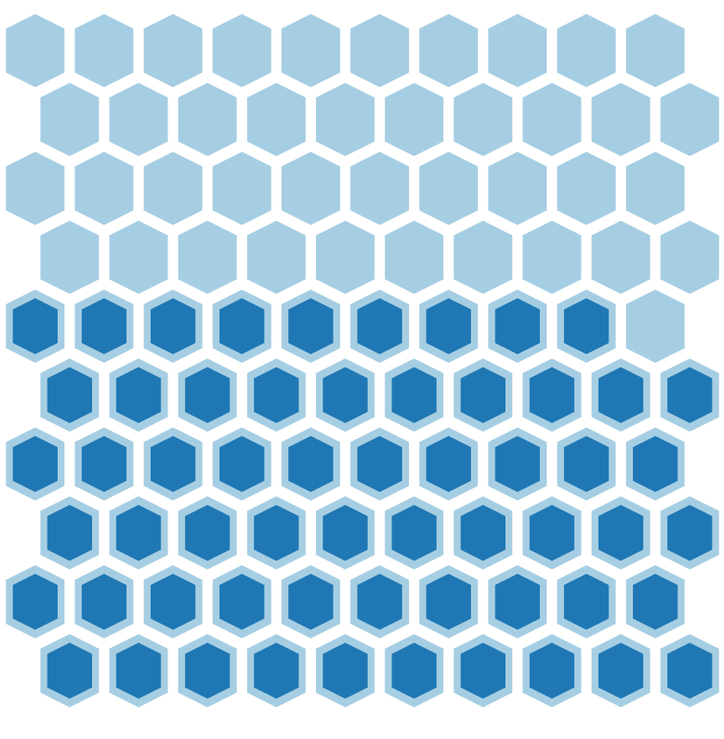
Gene	Cluster	Cytoband	Type
APC	1	5q22.2	SNV
TCEA1	1	8q11.23	Amp
PLAG1	1	8q12.1	Amp
CHCHD7	1	8q12.1	Amp
HEY1	1	8q21.13	Amp
COX6C	1	8q22.2	Amp
RSPO2	1	8q23.1	Amp
EIF3E	1	8q23.1	Amp
MYC	1	8q24.21	Amp
NDRG1	1	8q24.22	Amp
KRAS	1	12p12.1	SNV
STK11	1	19p13.3	SNV
AMER1	?		SNV

R1



R2





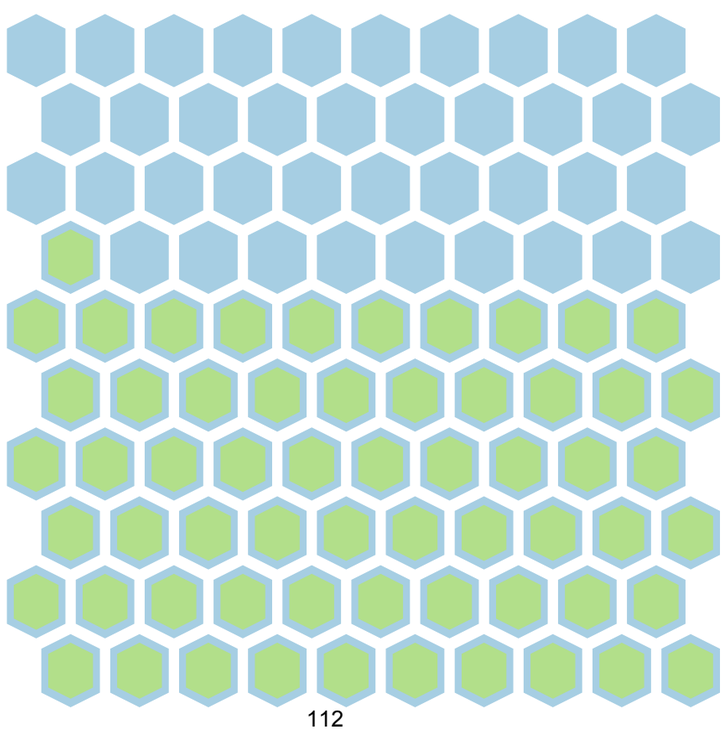
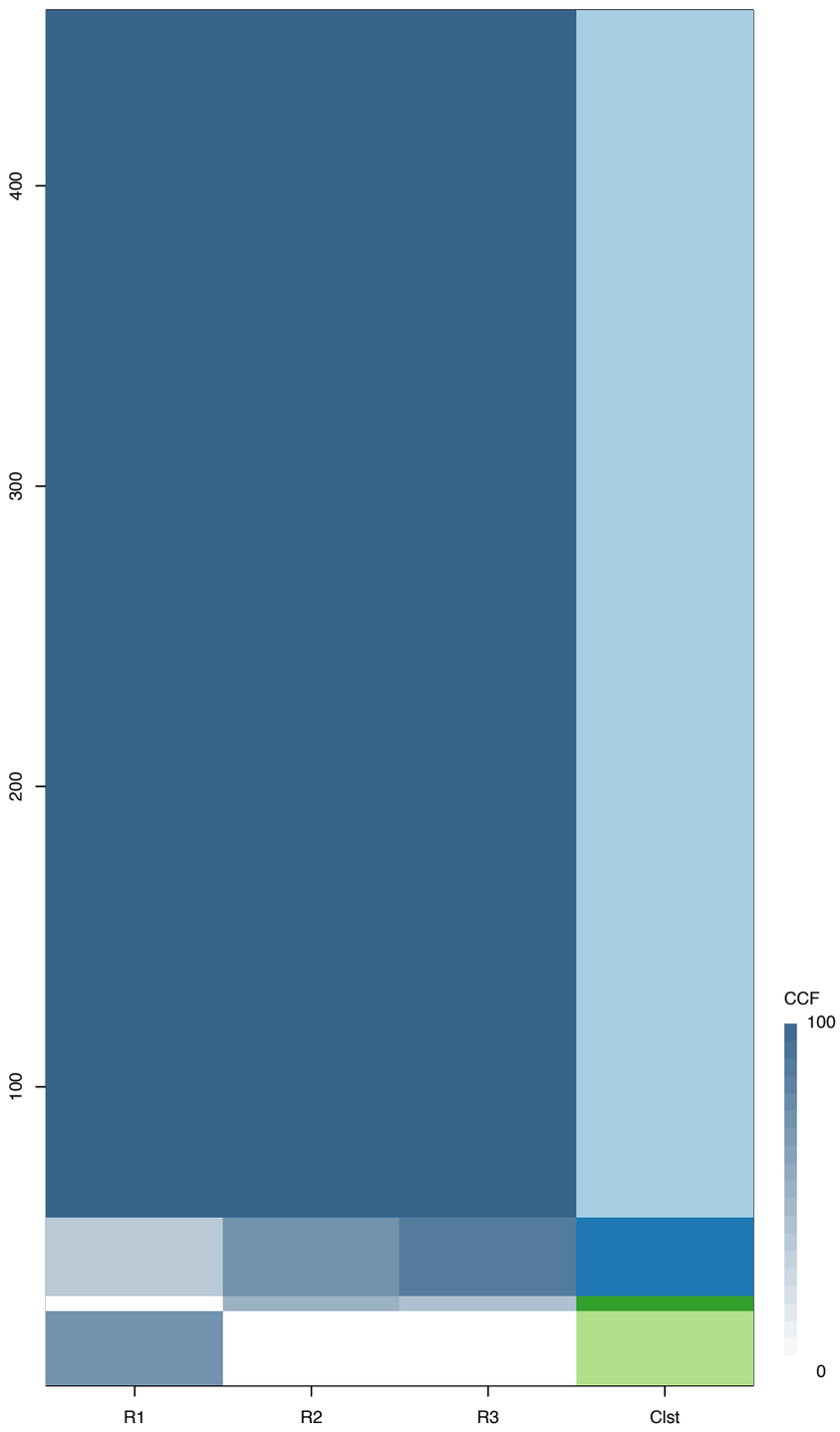
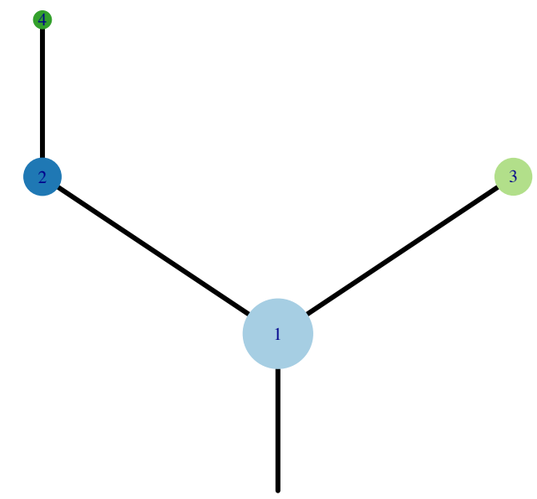


Fig.S12AU



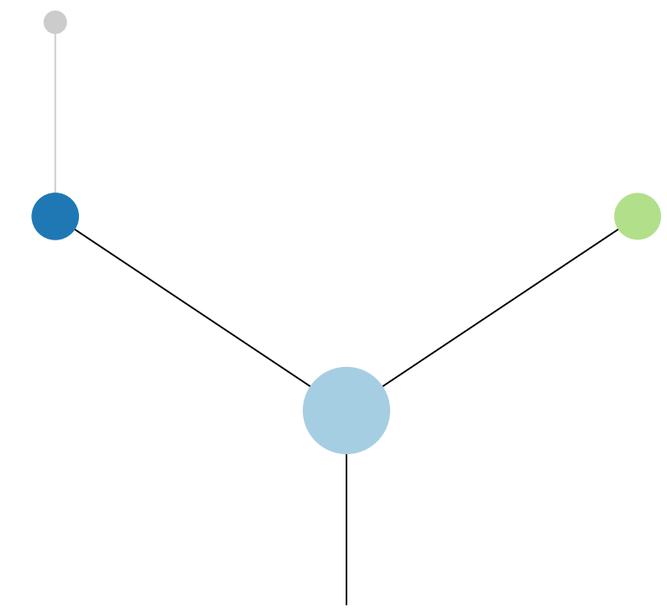
CRUK0048

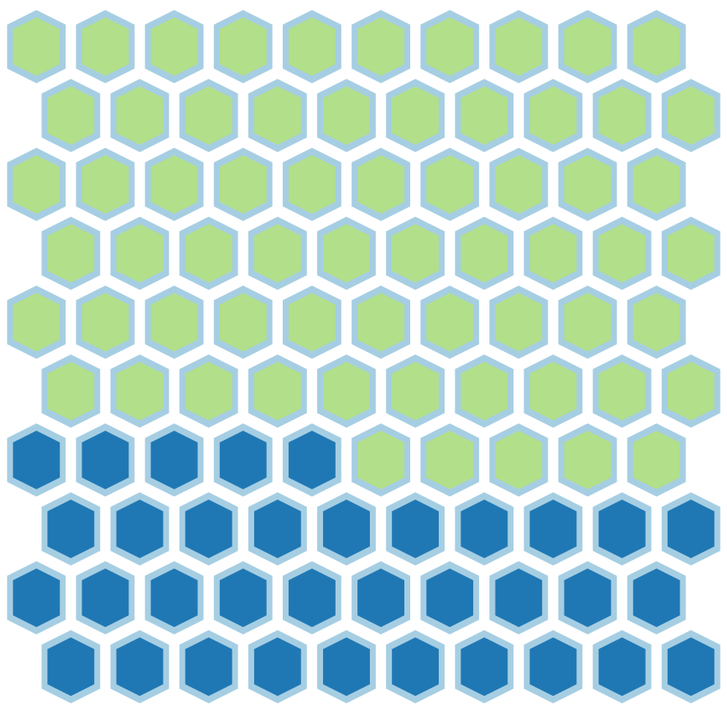


Histology:Adeno, Age:77, PackYears:49, Size:33
Stage:1b, Gender:Female, GD:Clonal GD, Recur:no

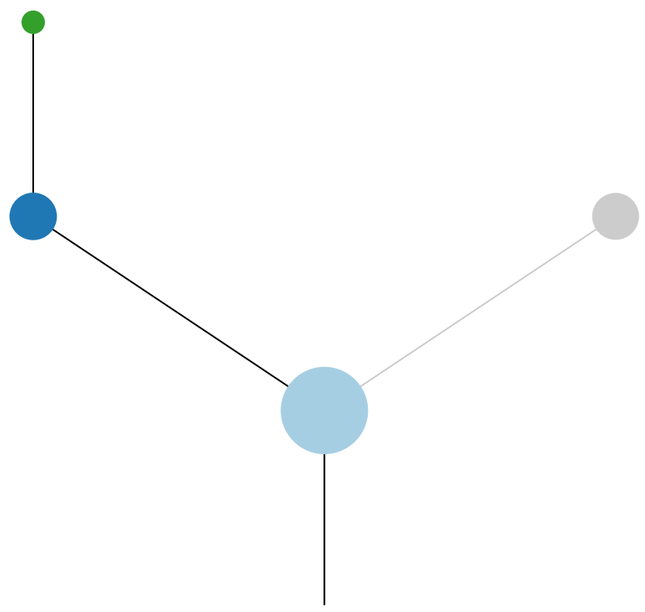
Gene	Cluster	Cytoband	Type
BCL9	1	1q21.2	Amp
TPM3	1	1q21.3	Amp
EML4	1	2p21	Amp
APC	1	5q22.2	SNV
PRDM1	1	6q21	SNV
EGFR	1	7p11.2	Amp
BRAF	1	7q34	SNV
MYC	1	8q24.21	Amp
NDRG1	1	8q24.22	Amp
KIF5B	1	10p11.22	Amp
FGFR2	1	10q26.13	Amp
TP53	1	17p13.1	SNV
ARHGAP35	1	19q13.32	SNV
KRAS	?	12p12.1	SNV

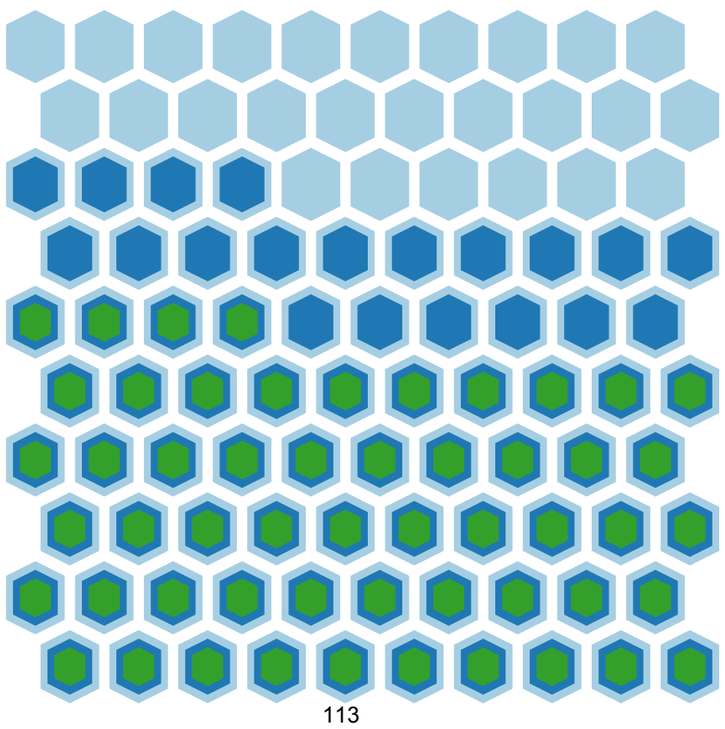
R1



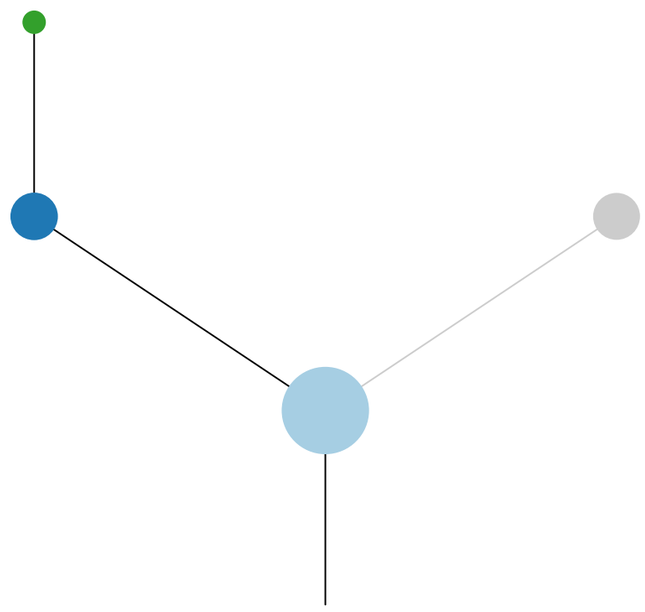


R2





R3



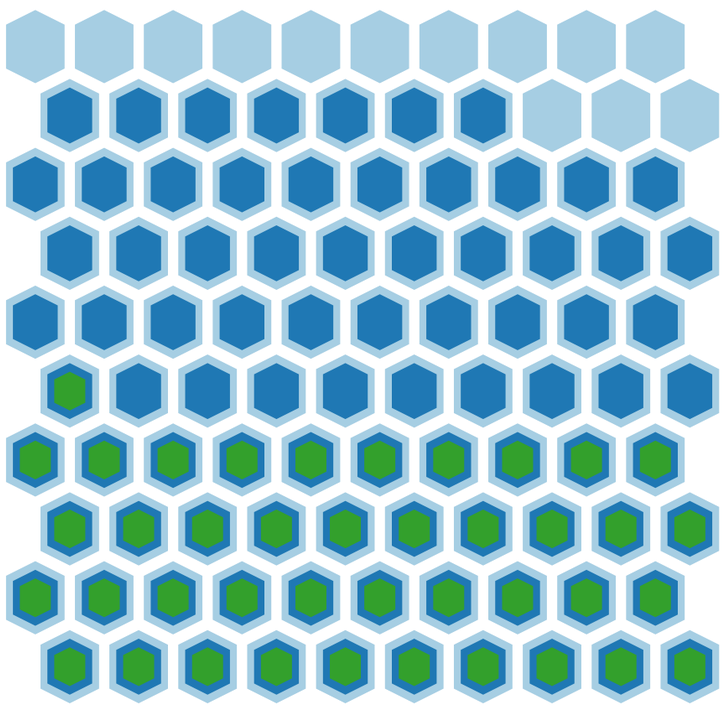
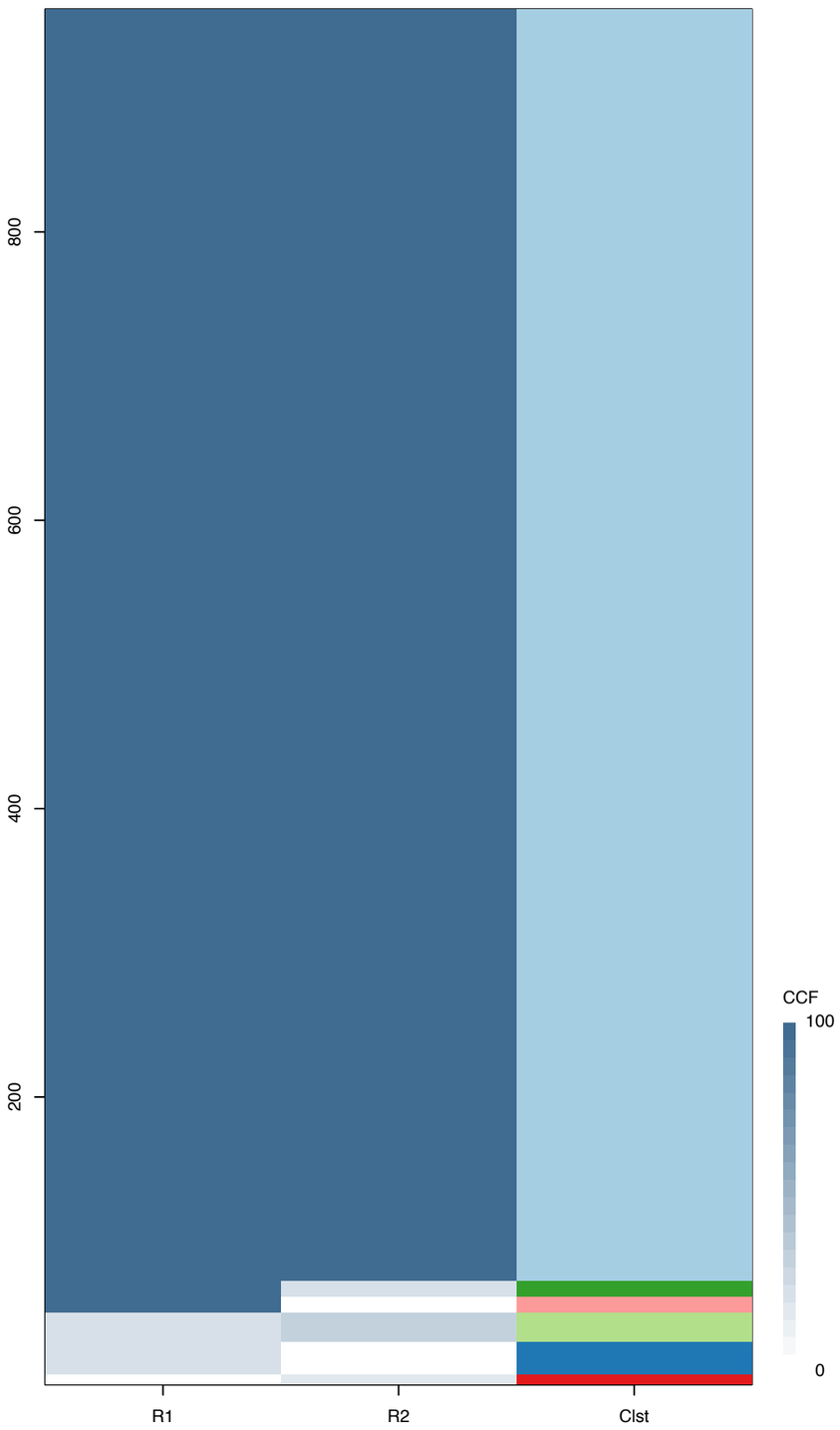
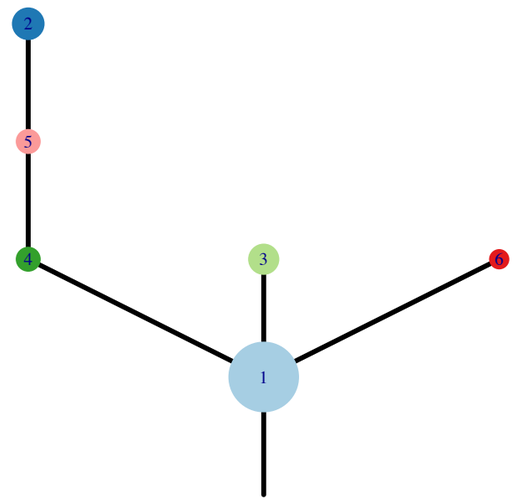


Fig.S12AV



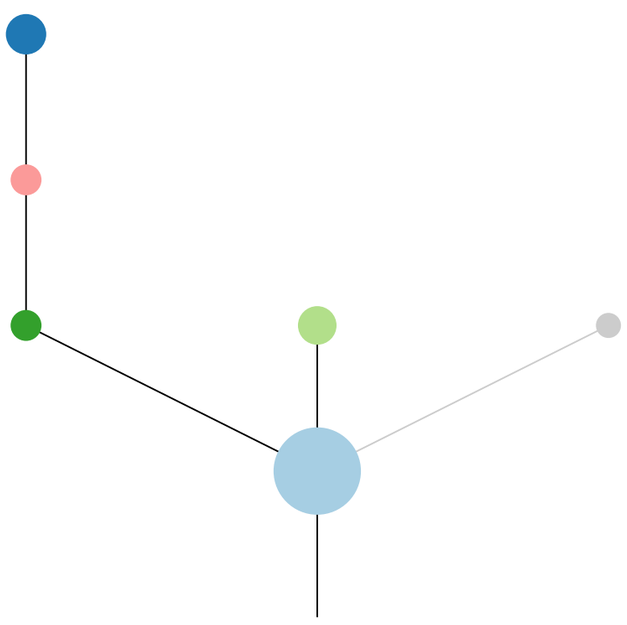
CRUK0049_A

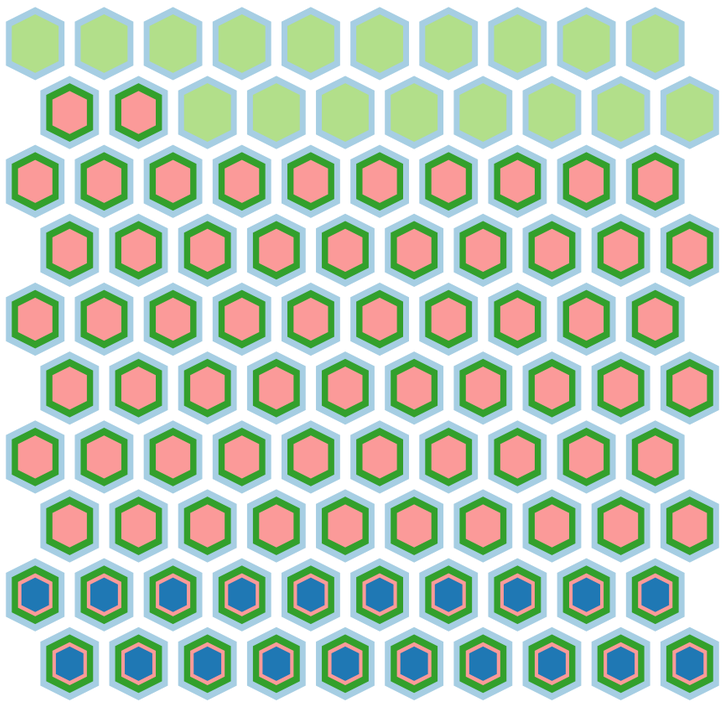


Histology:Adeno, Age:67, PackYears:48, Size:20
Stage:1a, Gender:Female, GD:Not GD, Recur:no

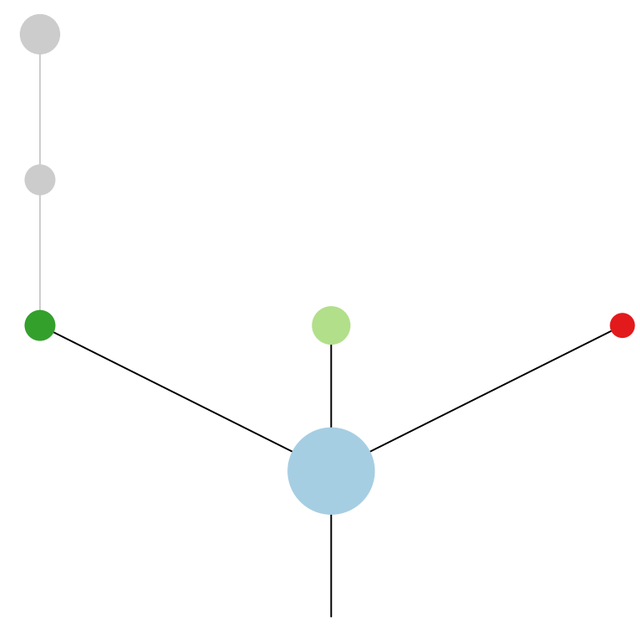
Gene	Cluster	Cytoband	Type
EGFR	1	7p11.2	Amp
ATF7IP	1	12p13.1	SNV
KRAS	1	12p12.1	Amp
PPFIBP1	1	12p11.23	Amp
COL2A1	1	12q13.11	SNV
RB1	1	13q14.2	SNV
TP53	1	17p13.1	SNV
DOT1L	?	19p13.3	SNV
BCOR	?		SNV
RBM10	?		SNV

R1





R2



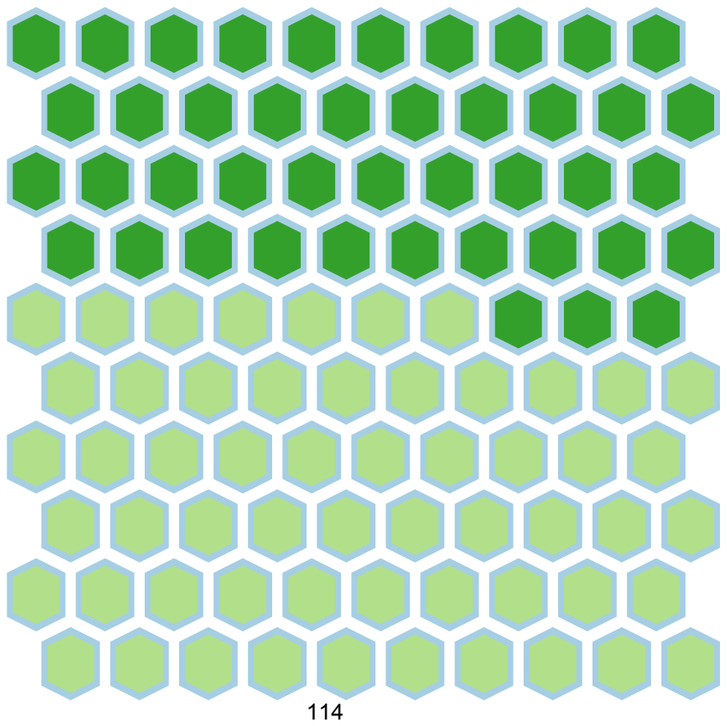
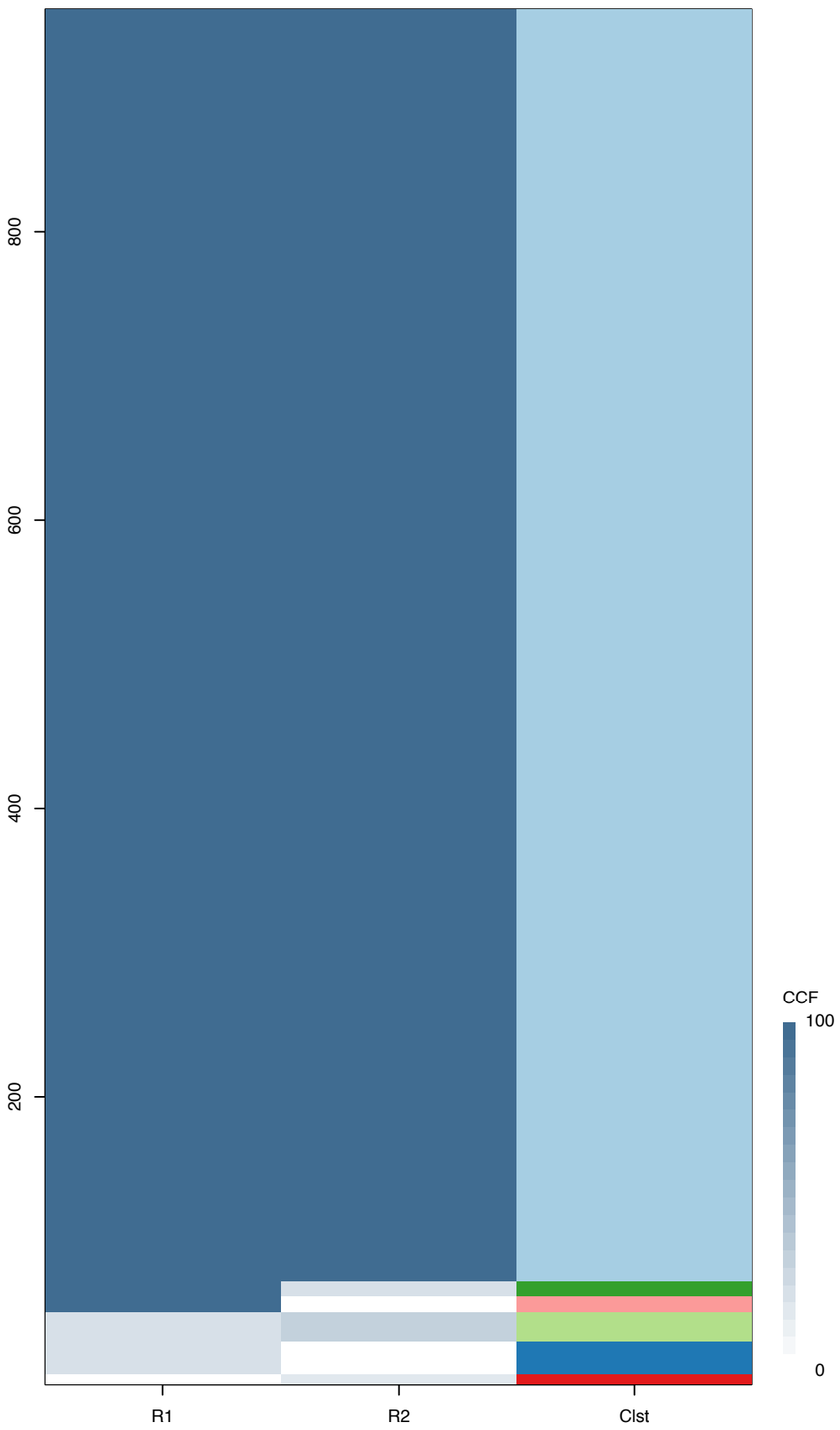
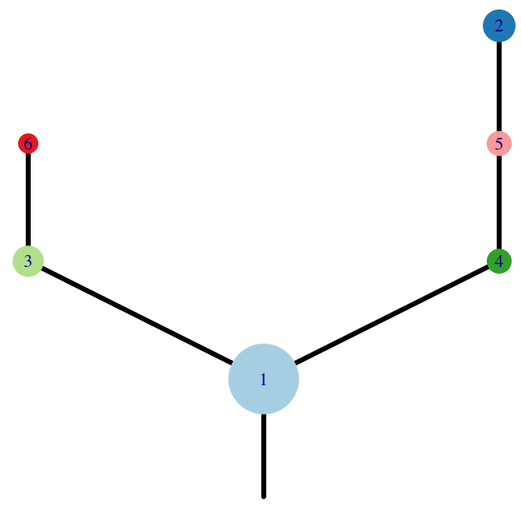


Fig.S12AV



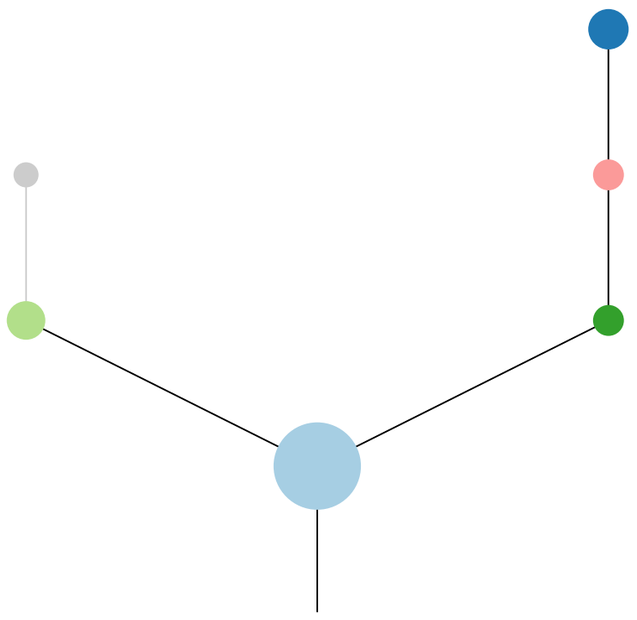
CRUK0049_B



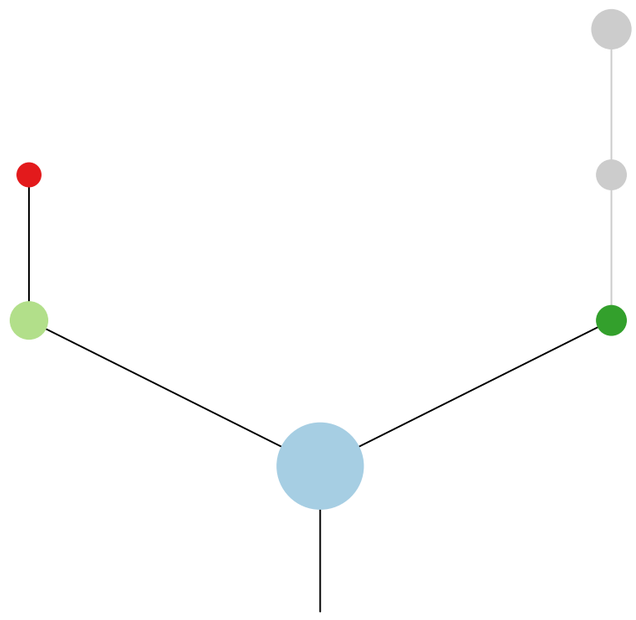
Histology:Adeno, Age:67, PackYears:48, Size:20
Stage:1a, Gender:Female, GD:Not GD, Recur:no

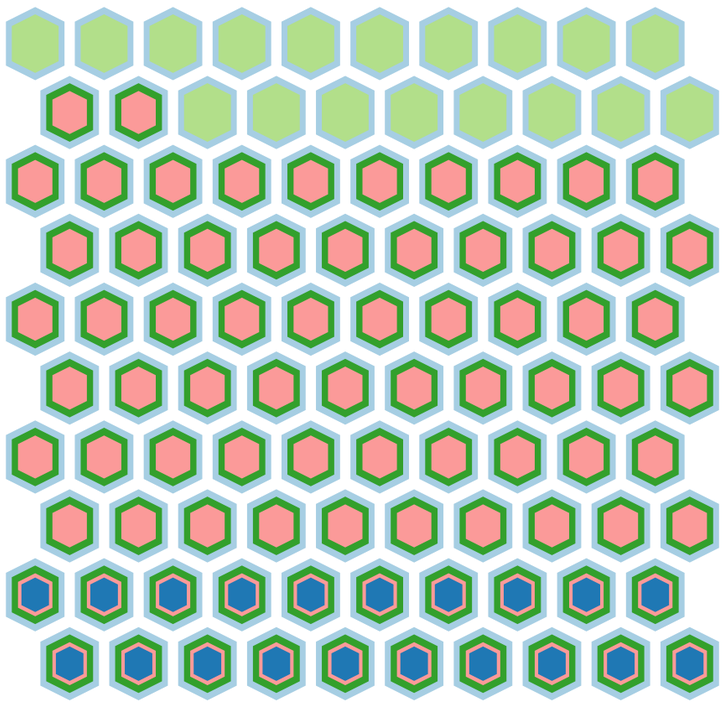
Gene	Cluster	Cytoband	Type
EGFR	1	7p11.2	Amp
ATF7IP	1	12p13.1	SNV
KRAS	1	12p12.1	Amp
PPFIBP1	1	12p11.23	Amp
COL2A1	1	12q13.11	SNV
RB1	1	13q14.2	SNV
TP53	1	17p13.1	SNV
DOT1L	?	19p13.3	SNV
BCOR	?		SNV
RBM10	?		SNV

R1



R2





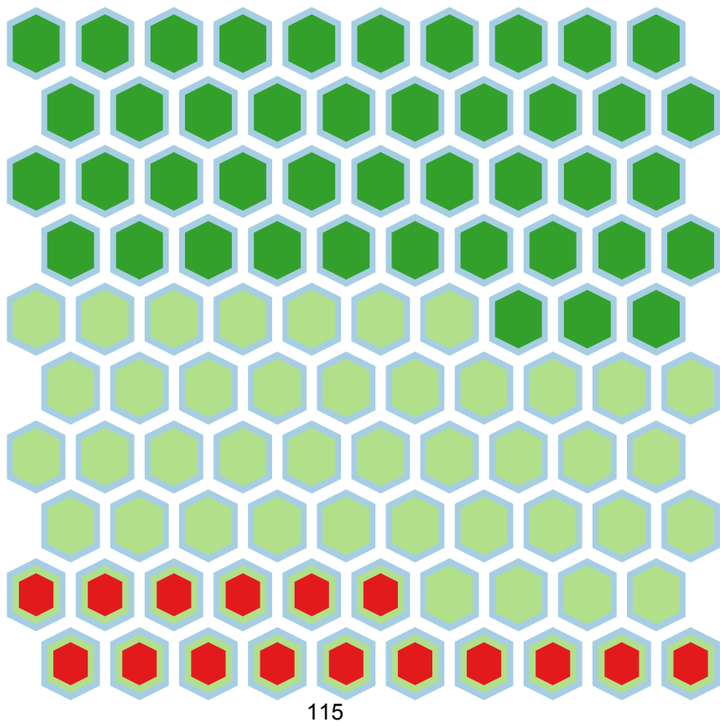
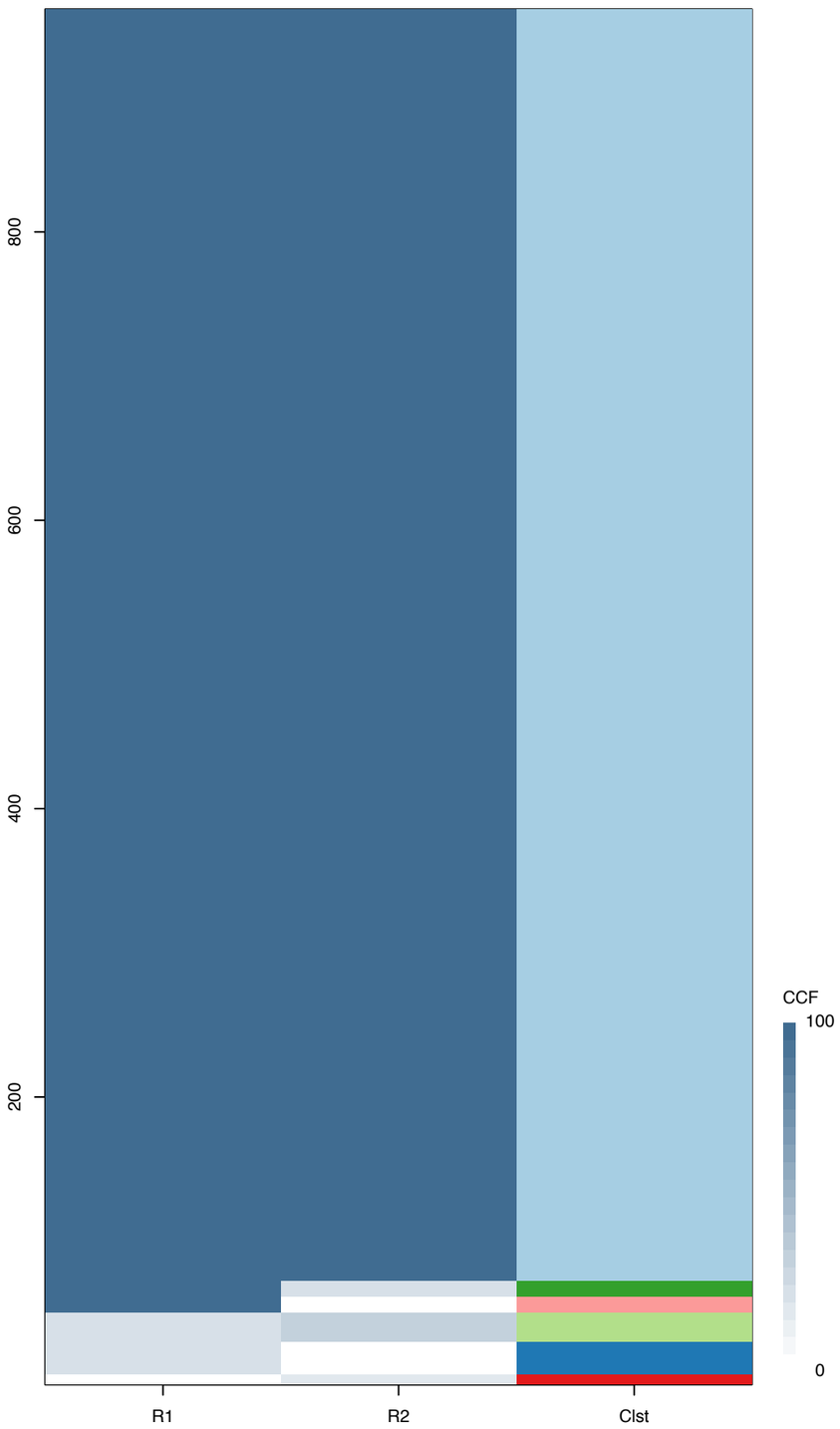
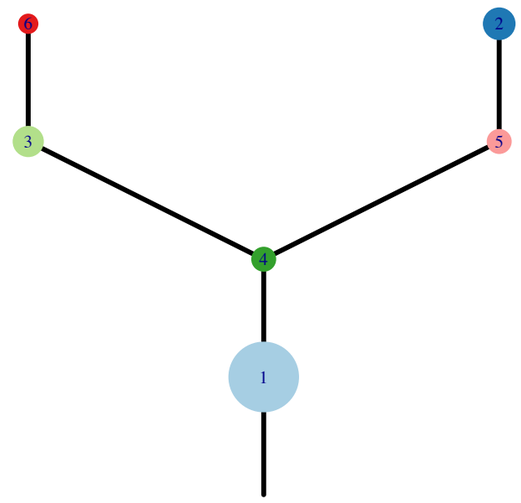


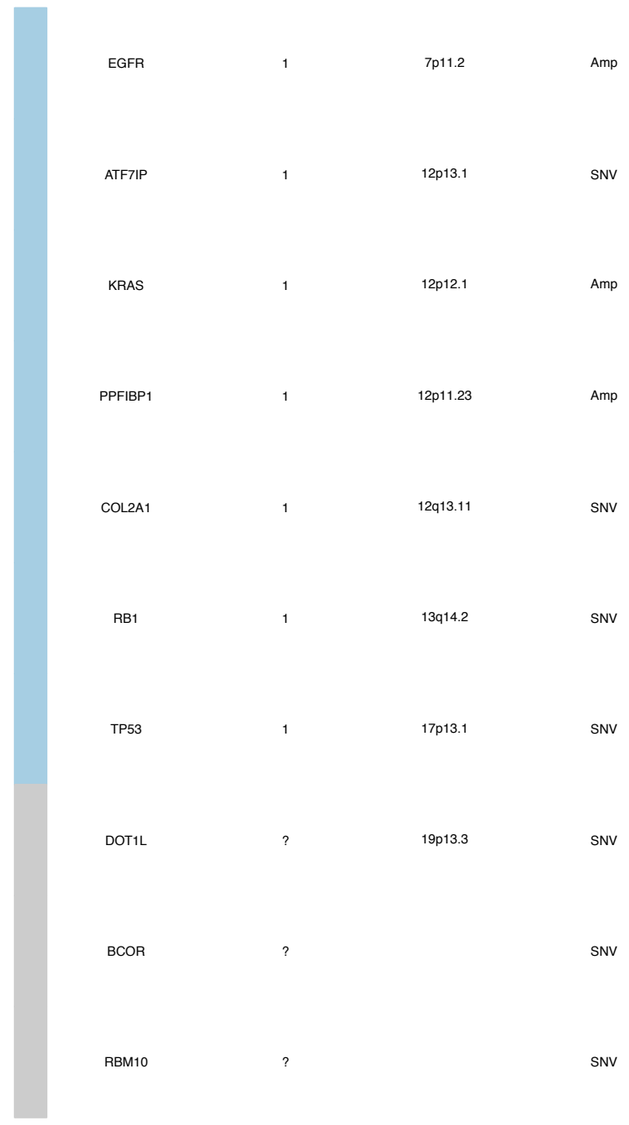
Fig.S12AV



CRUK0049_C

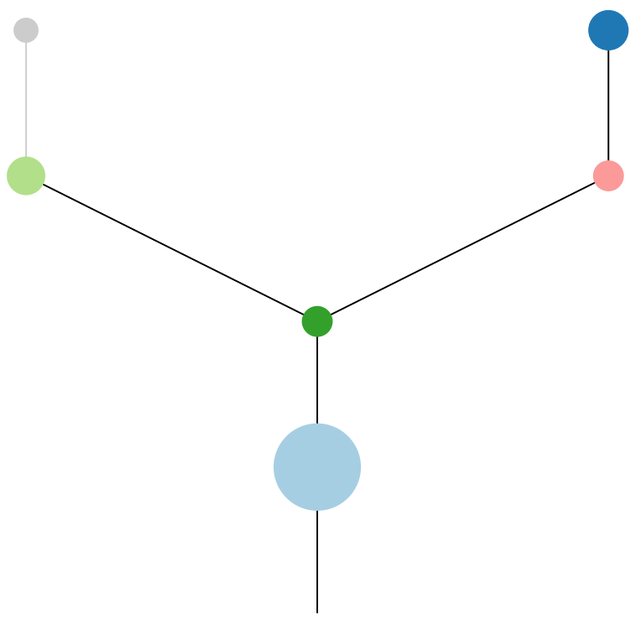


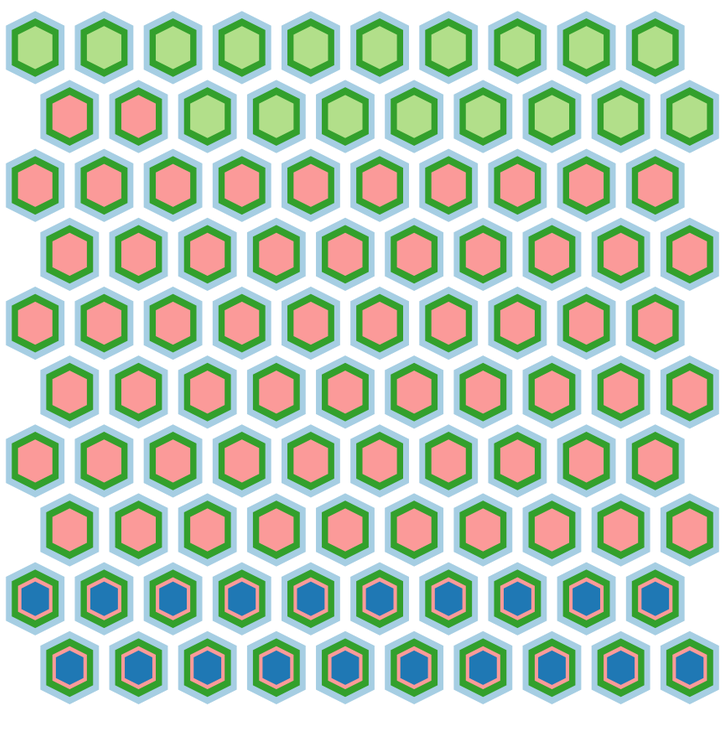
Histology:Adeno, Age:67, PackYears:48, Size:20
Stage:1a, Gender:Female, GD:Not GD, Recur:no



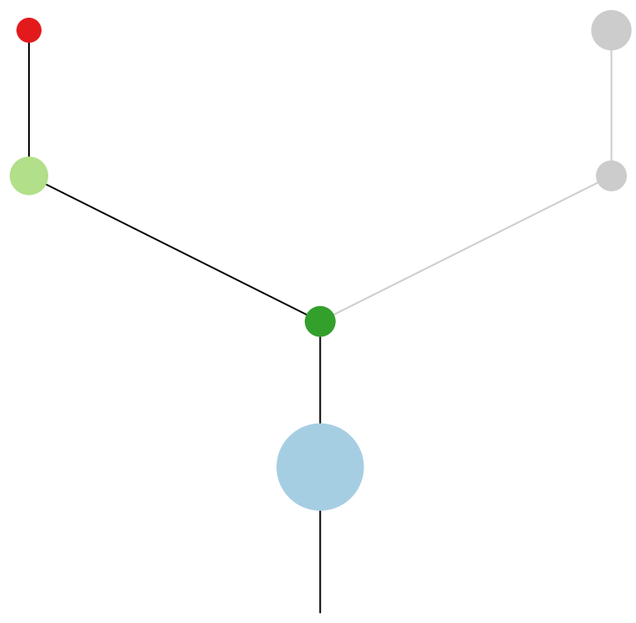
Gene	Cluster	Cytoband	Type
EGFR	1	7p11.2	Amp
ATF7IP	1	12p13.1	SNV
KRAS	1	12p12.1	Amp
PPFIBP1	1	12p11.23	Amp
COL2A1	1	12q13.11	SNV
RB1	1	13q14.2	SNV
TP53	1	17p13.1	SNV
DOT1L	?	19p13.3	SNV
BCOR	?		SNV
RBM10	?		SNV

R1





R2



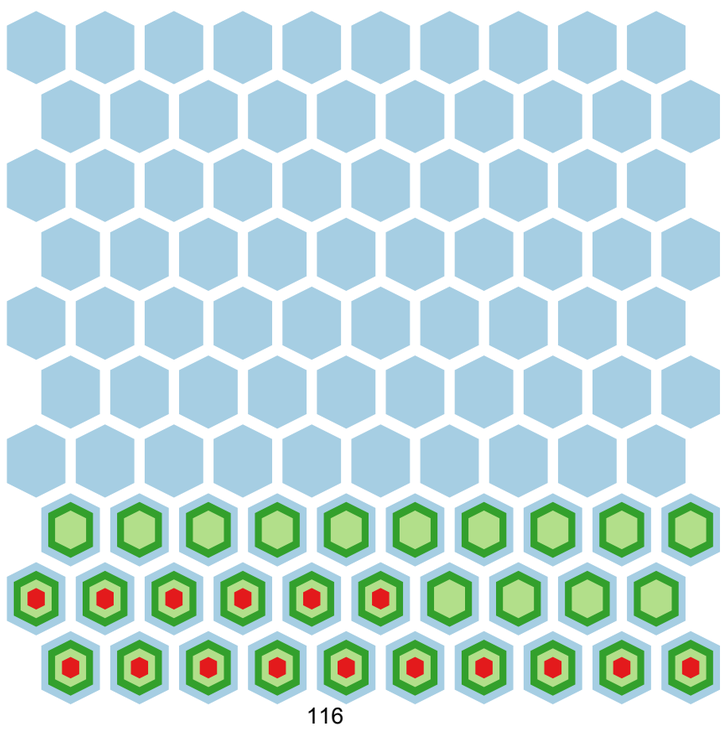
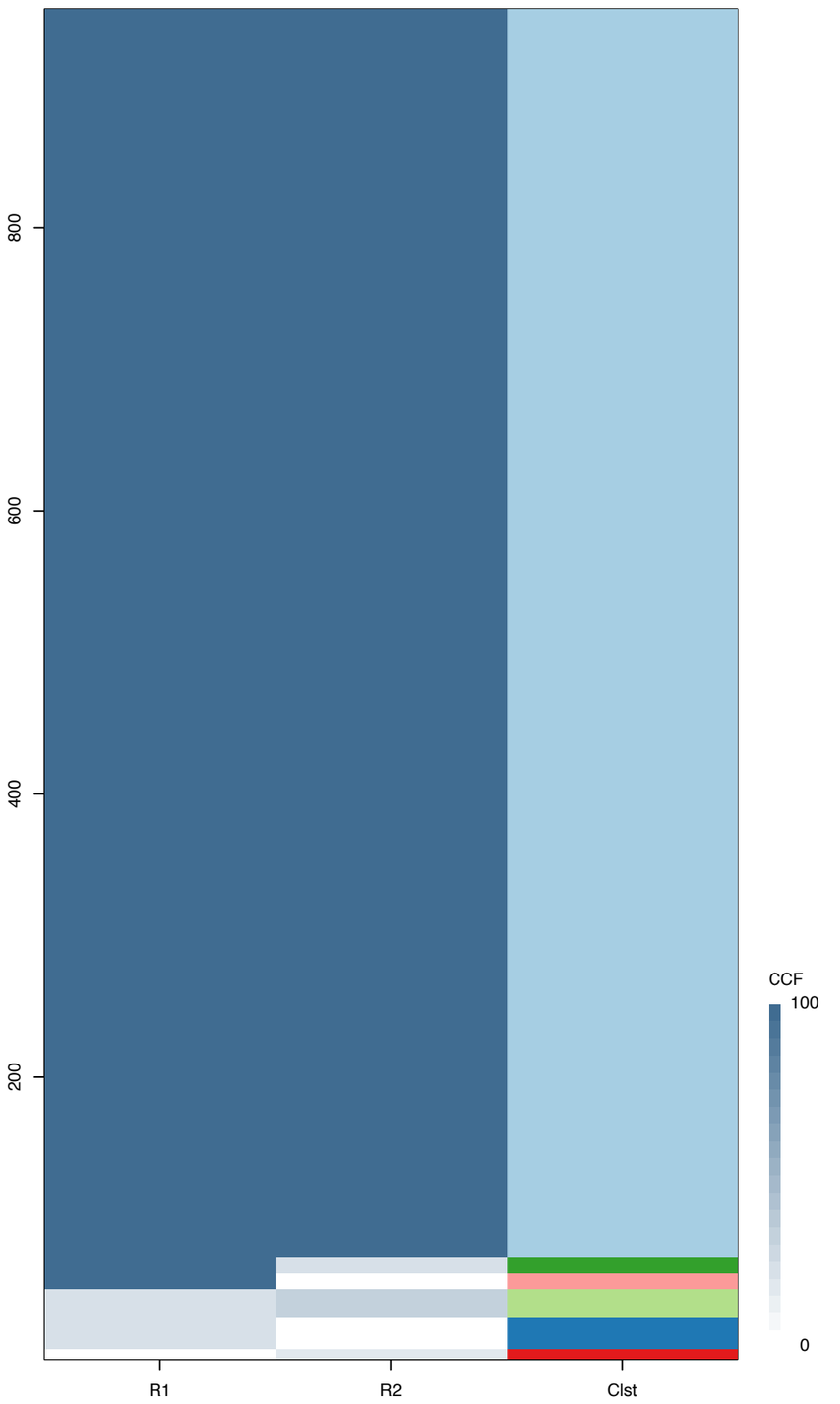
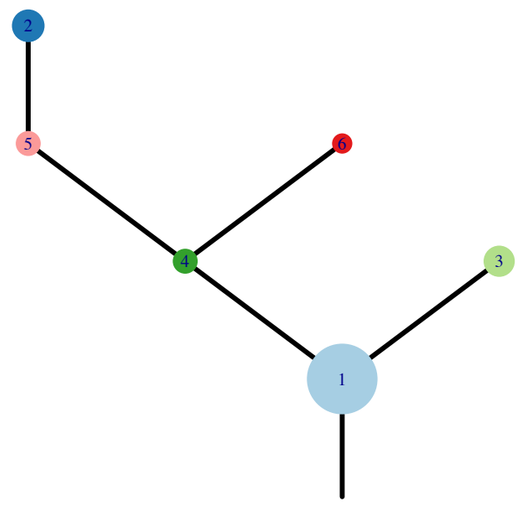


Fig.S12AV



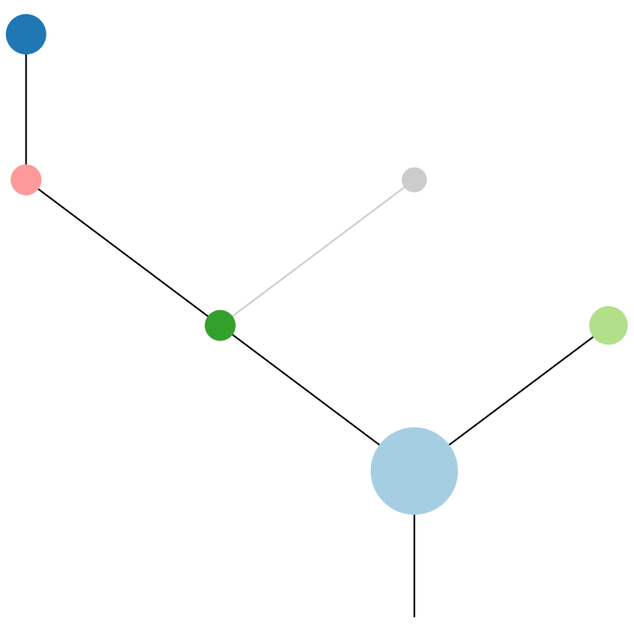
CRUK0049_D

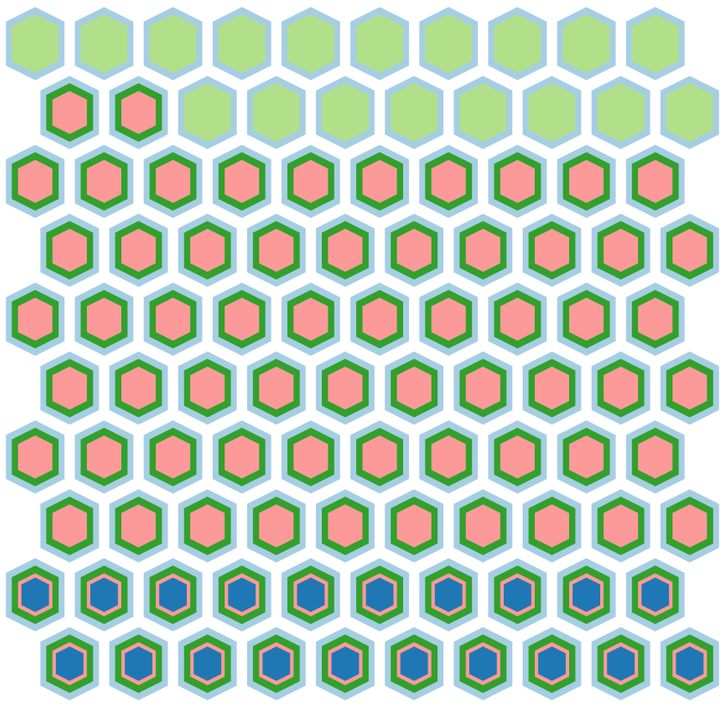


Histology:Adeno, Age:67, PackYears:48, Size:20
Stage:1a, Gender:Female, GD:Not GD, Recur:no

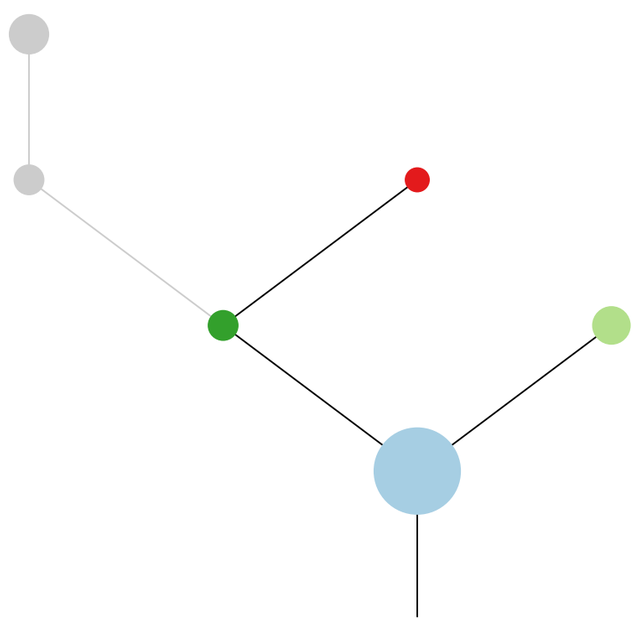
Gene	Cluster	Cytoband	Type
EGFR	1	7p11.2	Amp
ATF7IP	1	12p13.1	SNV
KRAS	1	12p12.1	Amp
PPFIBP1	1	12p11.23	Amp
COL2A1	1	12q13.11	SNV
RB1	1	13q14.2	SNV
TP53	1	17p13.1	SNV
DOT1L	?	19p13.3	SNV
BCOR	?		SNV
RBM10	?		SNV

R1





R2



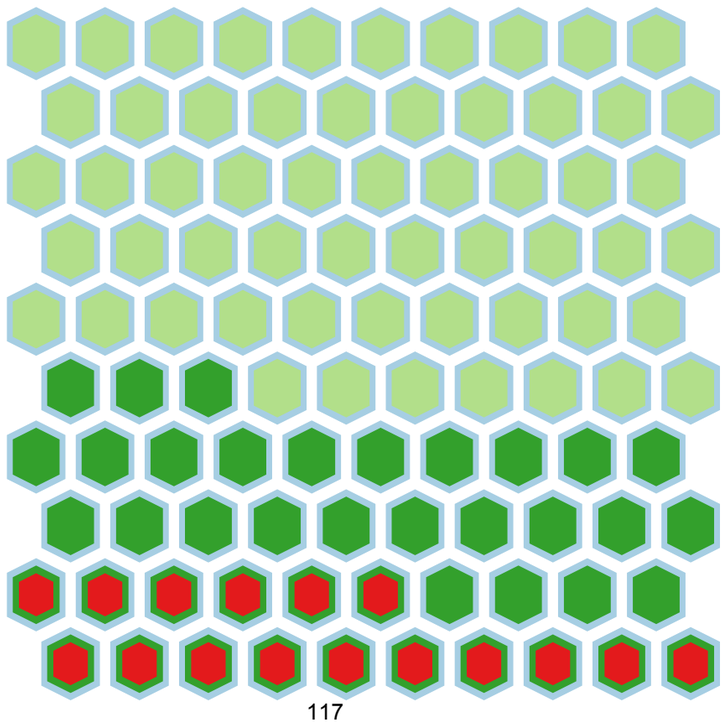
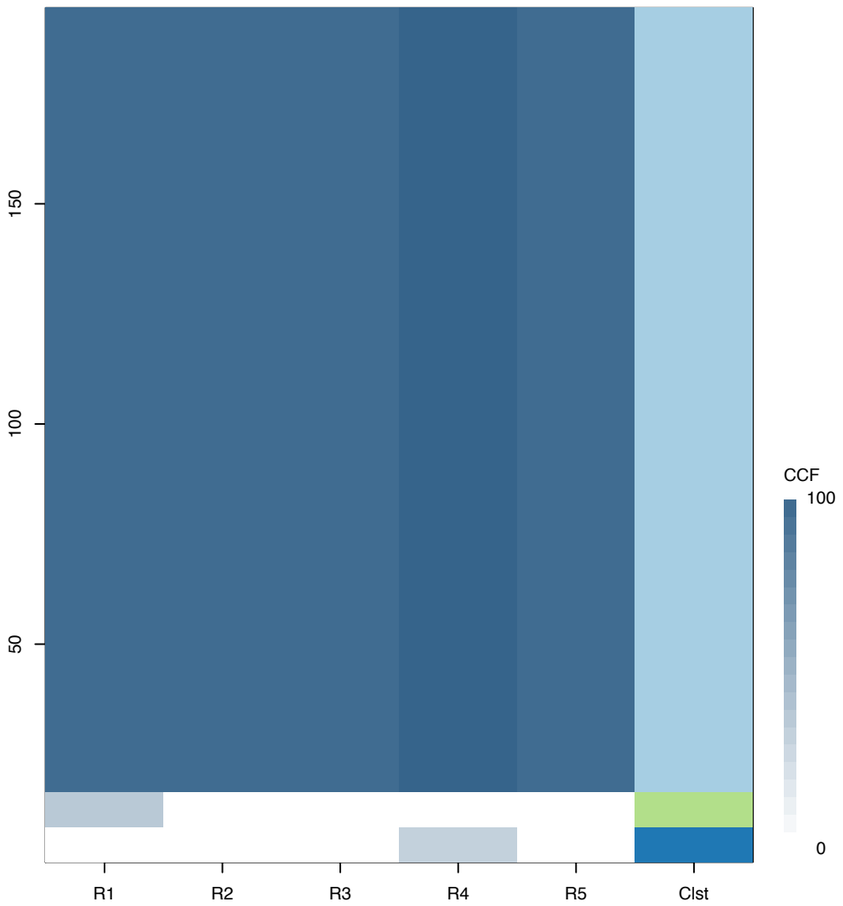
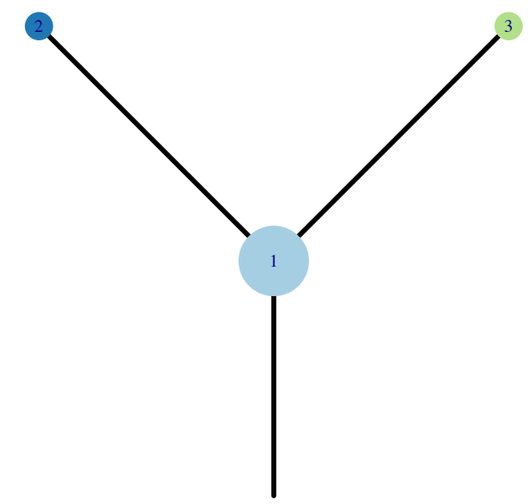


Fig.S12AW



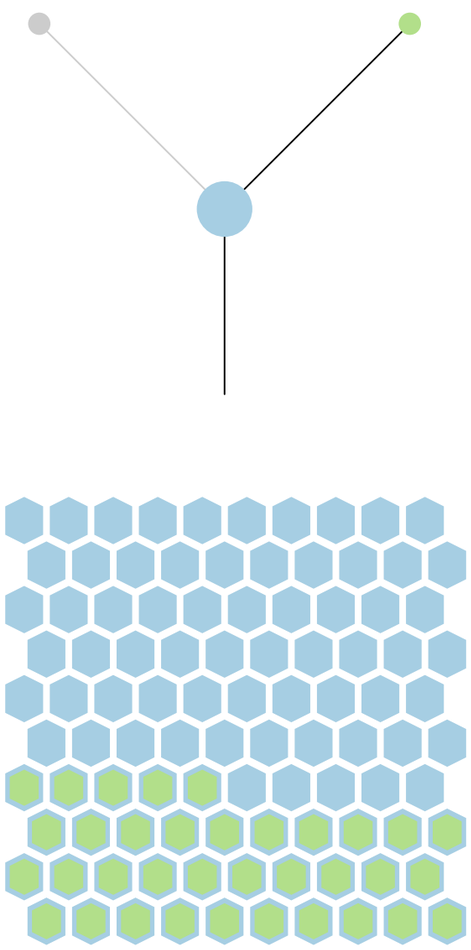
CRUK0050



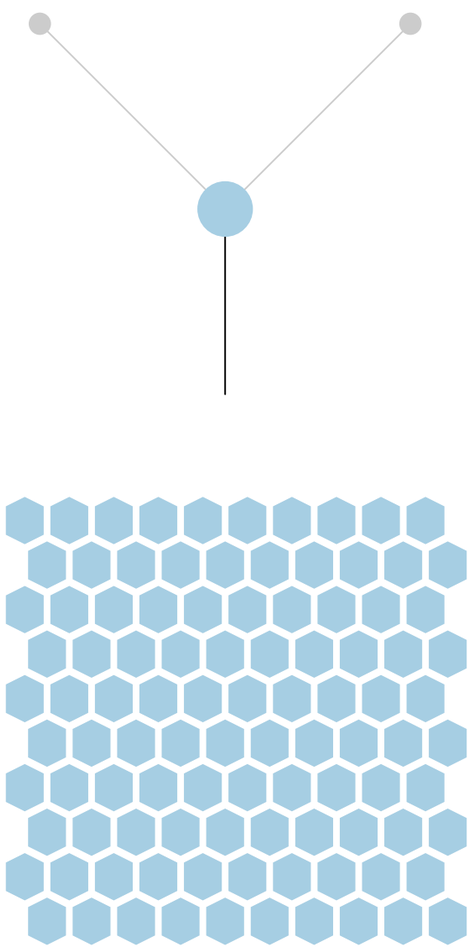
Histology:Adeno, Age:57, PackYears:42, Size:120
Stage:3a, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
NCOA2	1	8q13.3	Amp
HEY1	1	8q21.13	Amp
MYC	1	8q24.21	Amp
KRAS	1	12p12.1	SNV
STK11	1	19p13.3	SNV
CCND1	3	11q13.3	Amp
NDRG1	?	8q24.22	Amp

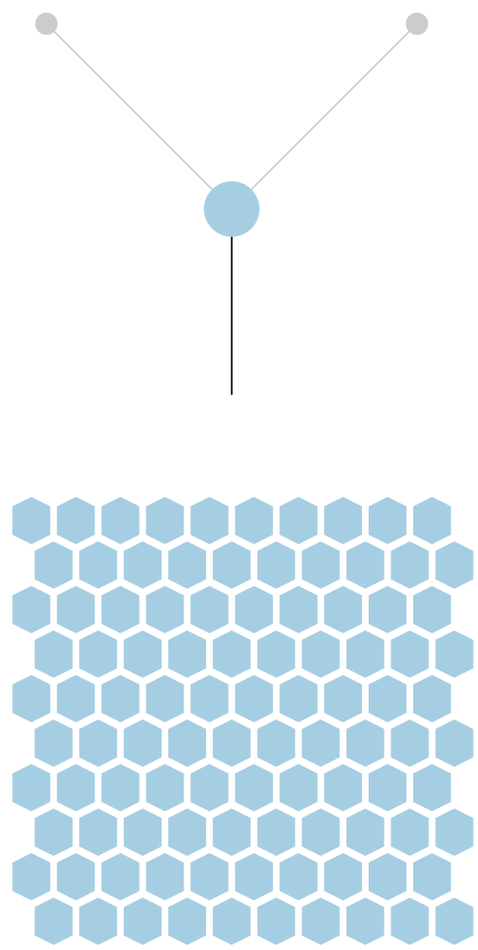
R1



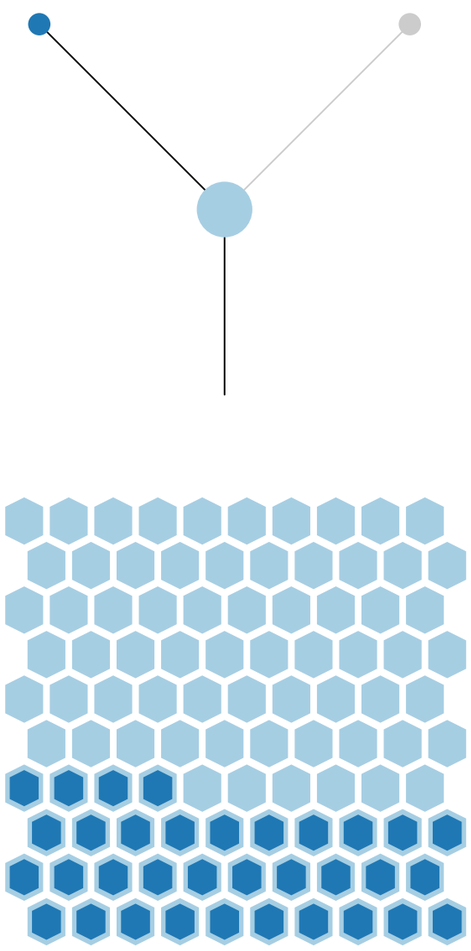
R2



R3



R4



R5

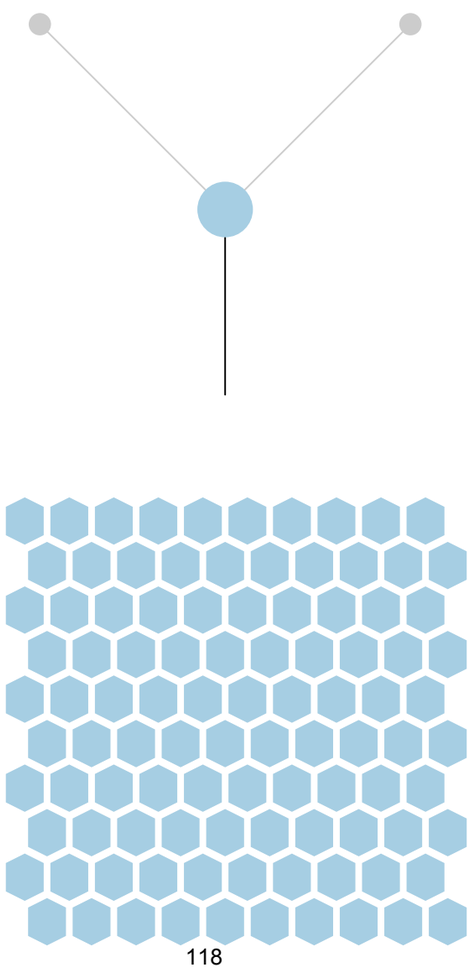
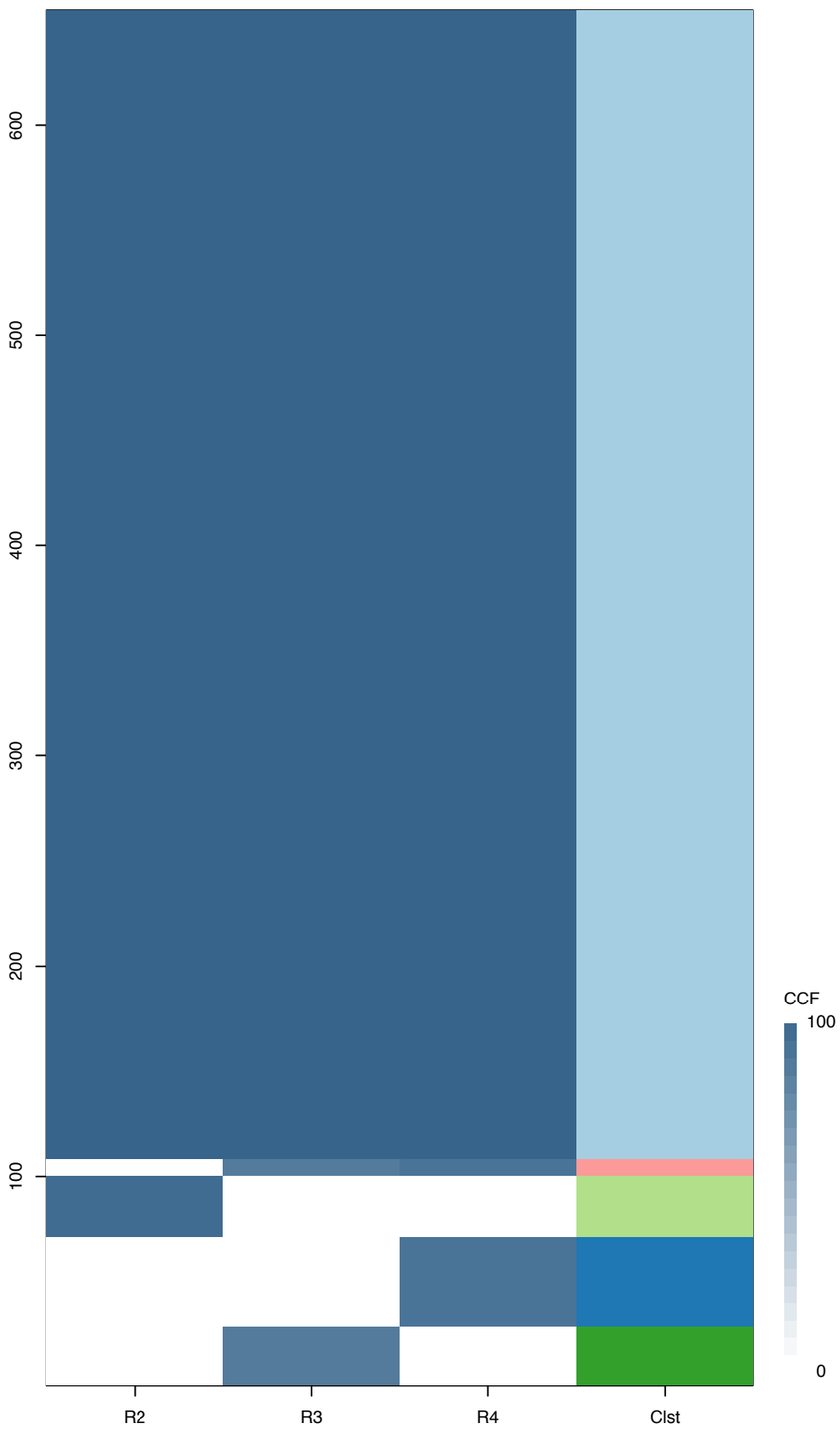
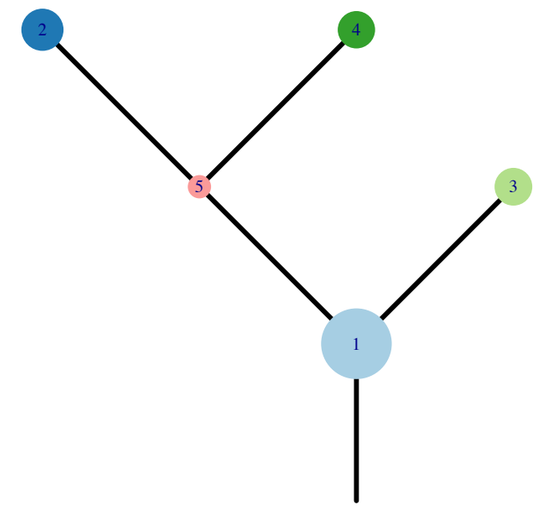


Fig.S12AX



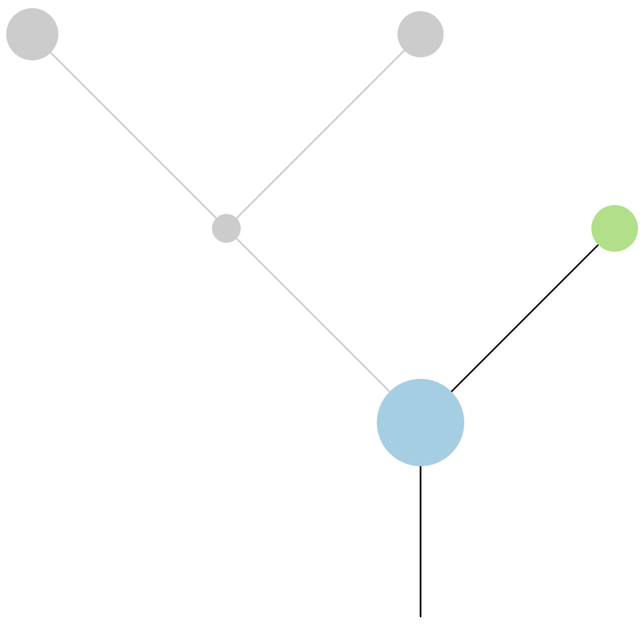
CRUK0051

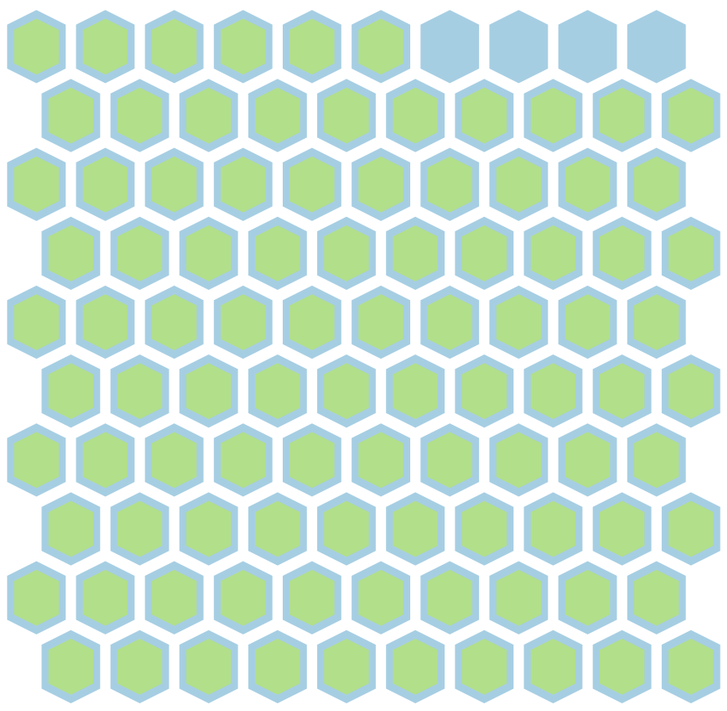


Histology:Adeno, Age:77, PackYears:73.5, Size:50
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

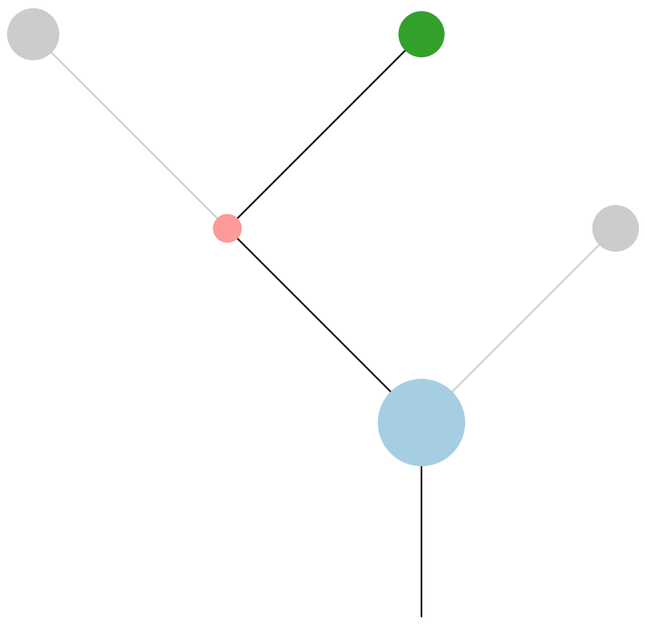
Gene	Cluster	Cytoband	Type
FBXW7	1	4q31.3	SNV
CARD11	1	7p22.2	Amp
RAC1	1	7p22.1	Amp
ETV1	1	7p21.2	Amp
HNRNPA2B1	1	7p15.2	Amp
HOXA9	1	7p15.2	Amp
HOXA11	1	7p15.2	Amp
HOXA13	1	7p15.2	Amp
JAZF1	1	7p15.2	Amp
EGFR	1	7p11.2	Amp
KRAS	1	12p12.1	SNV
ERCC5	1	13q33.1	SNV
TP53	1	17p13.1	SNV
EP300	3	22q13.2	SNV

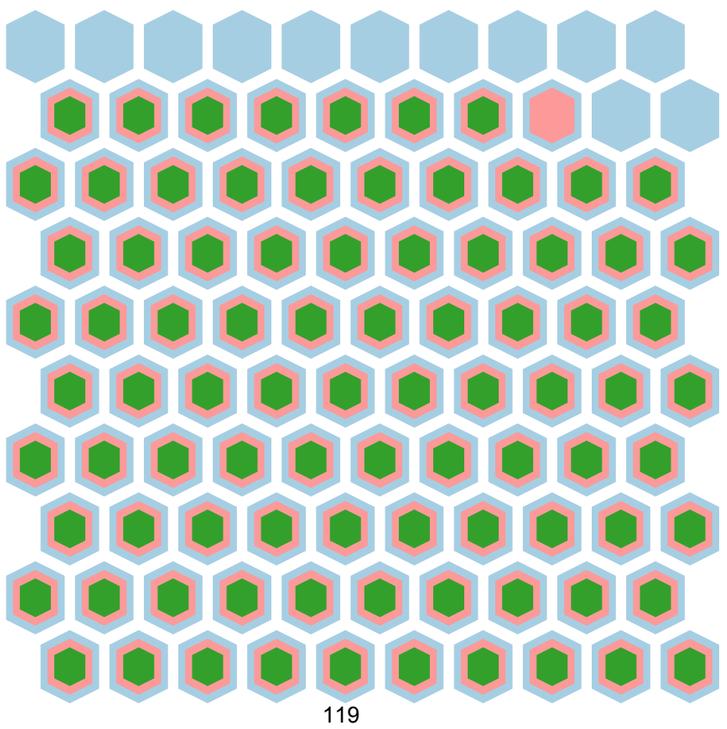
R2



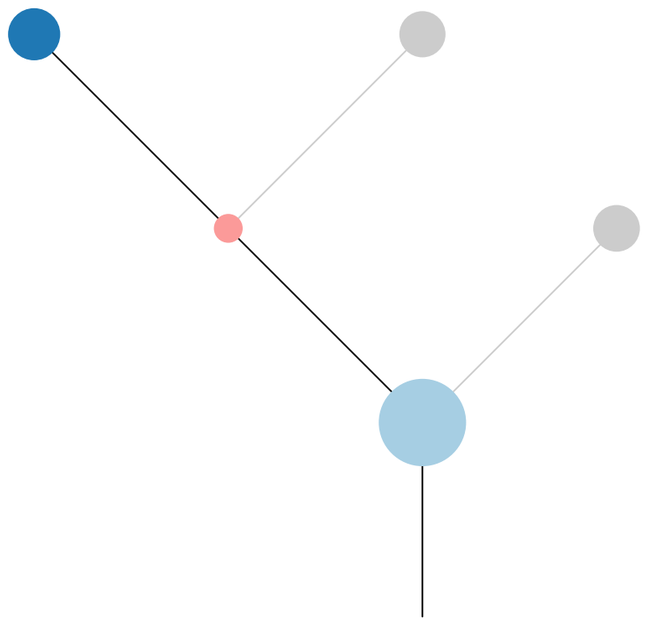


R3





R4



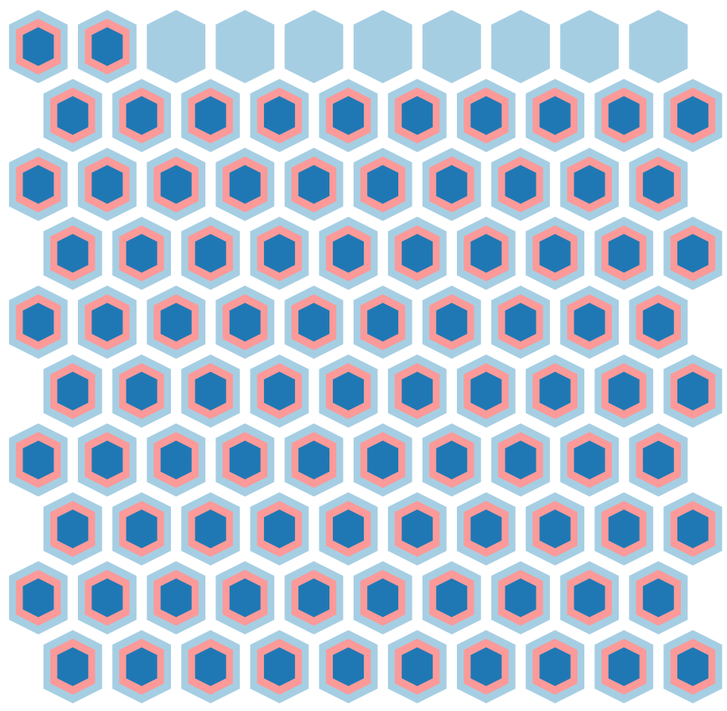
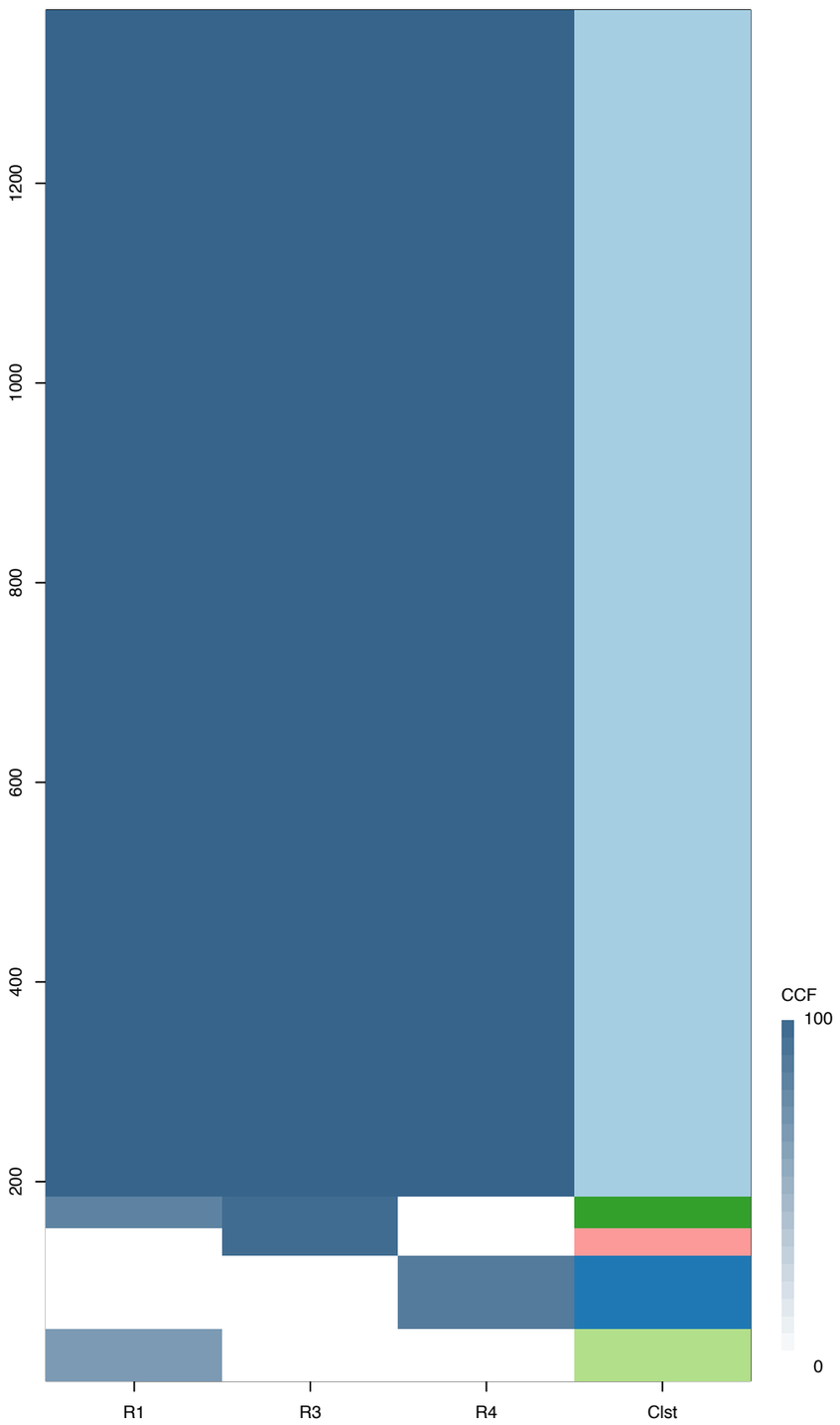
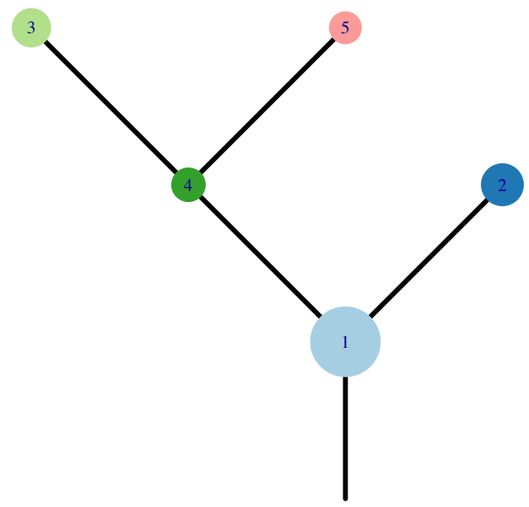


Fig.S12AY

CRUK0052

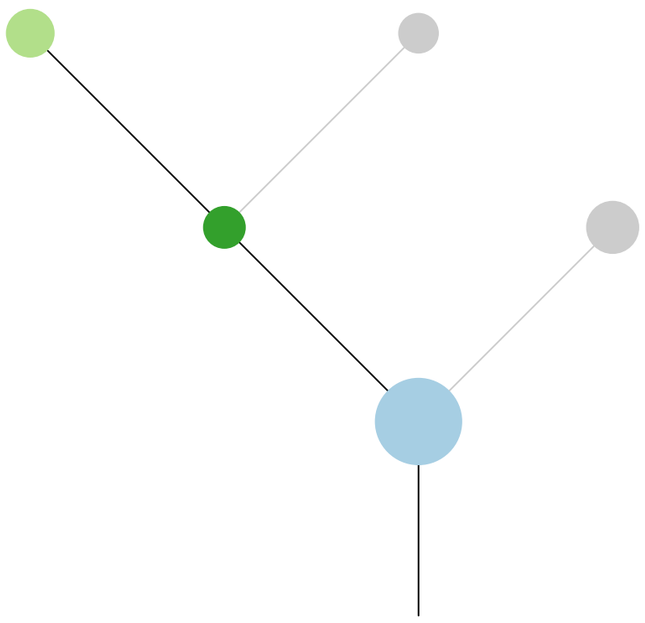




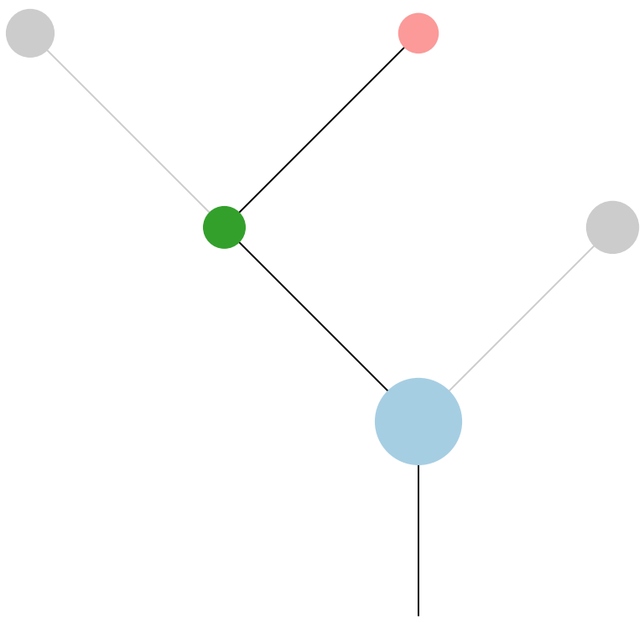
Histology:Adeno, Age:60, PackYears:35, Size:37
Stage:3a, Gender:Female, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
ARID1A	1	1p36.11	SNV
NOTCH2	1	1p12	SNV
SGK223	1	8p23.1	SNV
KRAS	1	12p12.1	SNV
KMT2D	1	12q13.12	SNV
TSHR	1	14q31.1	SNV
MGA	1	15q15.1	SNV
B2M	1	15q21.1	SNV
TP53	1	17p13.1	SNV
NF1	1	17q11.2	SNV
KEAP1	1	19p13.2	SNV
UBR5	2	8q22.3	SNV

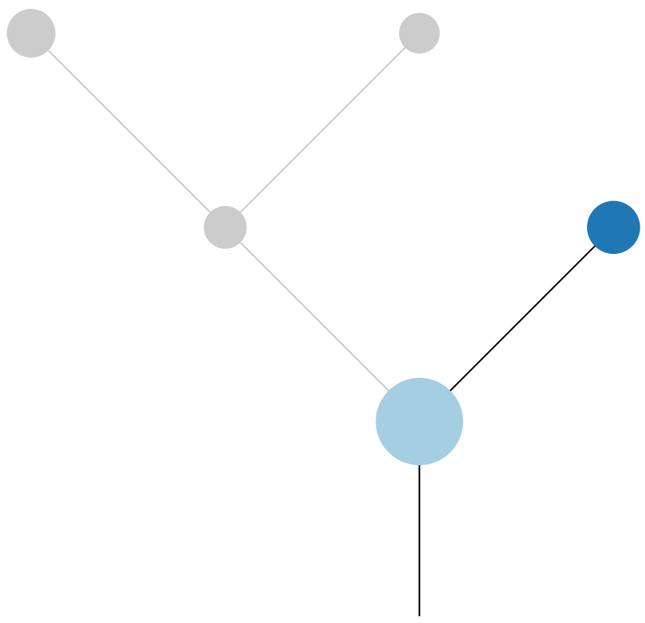
R1

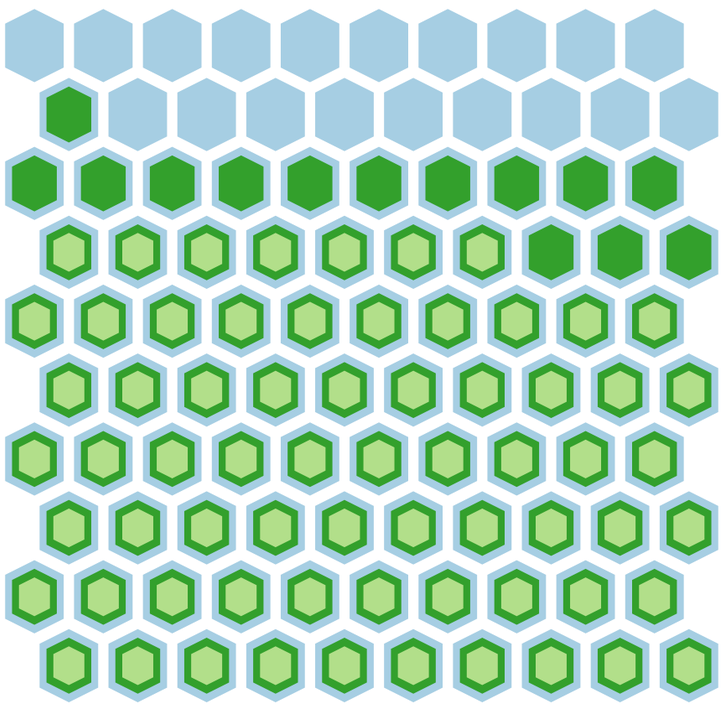


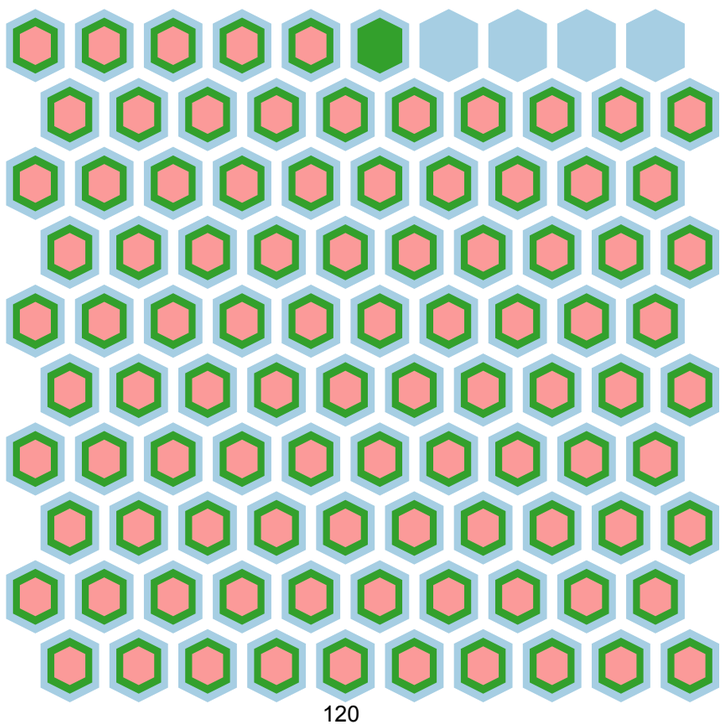
R3



R4







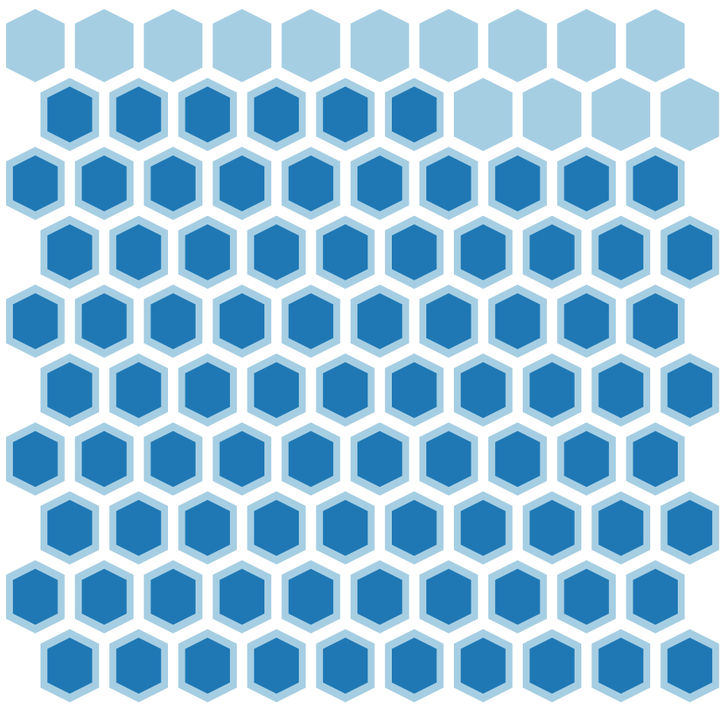
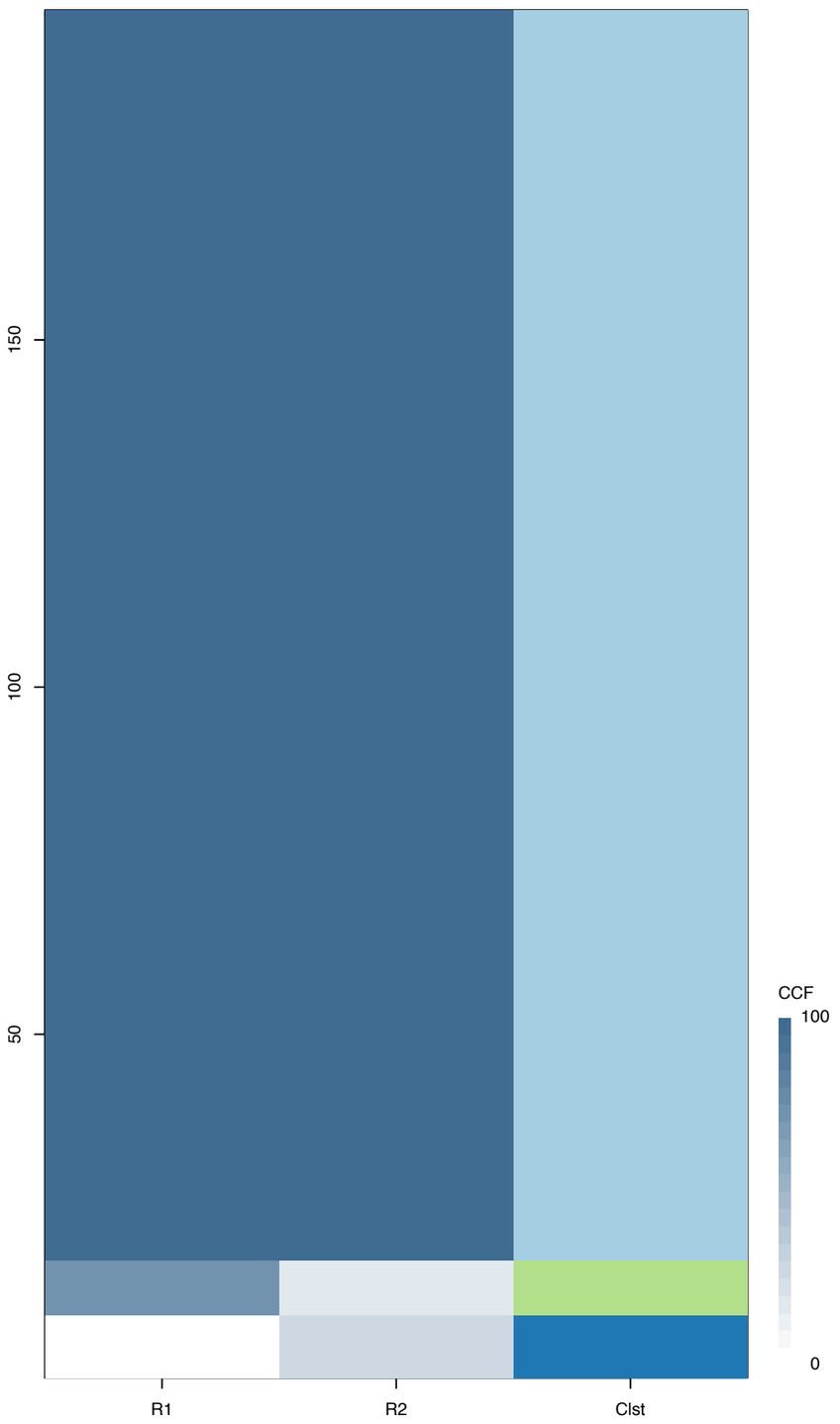
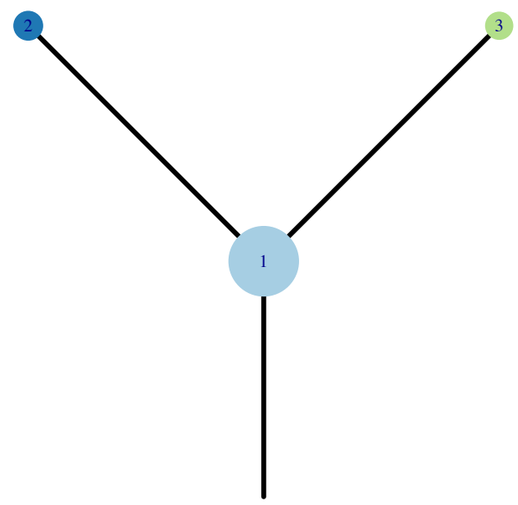


Fig.S12AZ



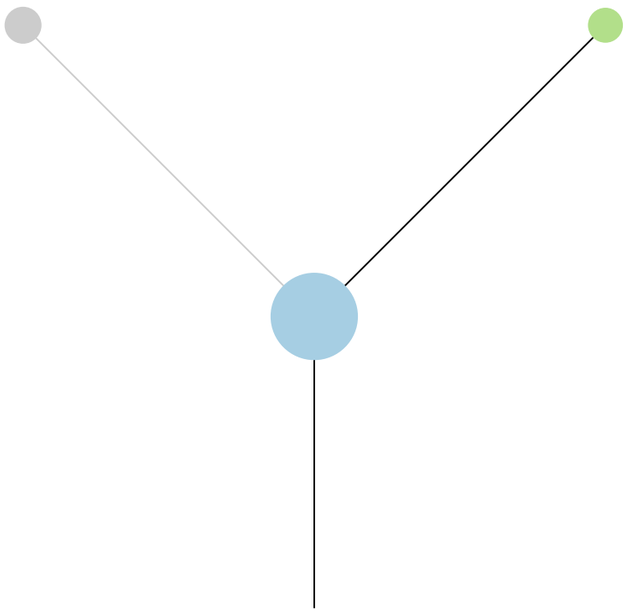
CRUK0054



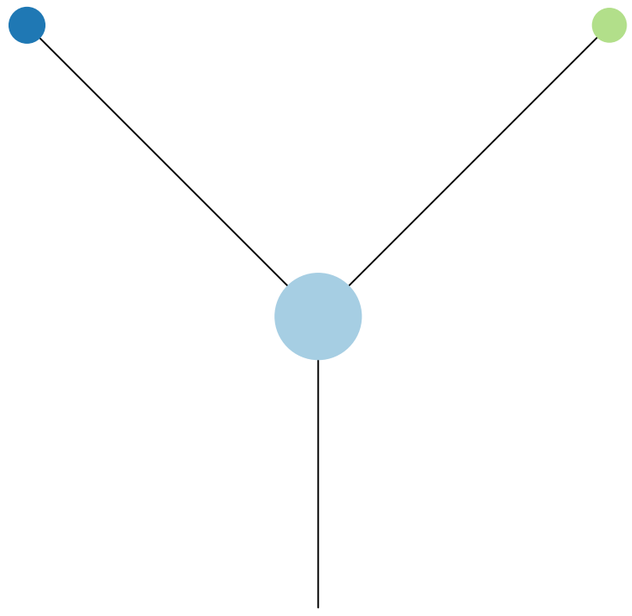
Histology:Adeno, Age:77, PackYears:NA, Size:14
Stage:1a, Gender:Female, GD:Clonal GD, Recur:no

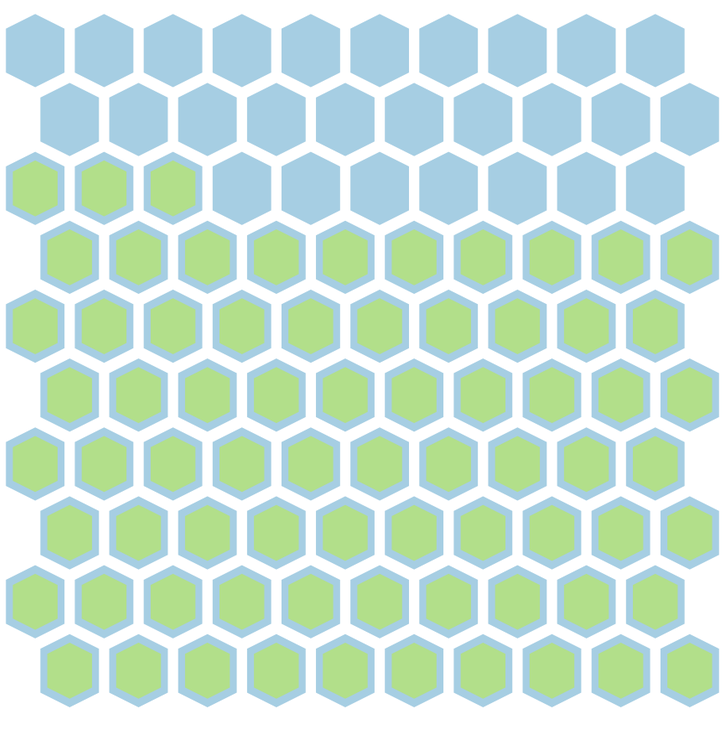
Gene	Cluster	Cytoband	Type
ARNT	1	1q21.3	Amp
SETDB1	1	1q21.3	Amp
MLLT11	1	1q21.3	Amp
TPR	1	1q31.1	Amp
MDM4	1	1q32.1	Amp
SLC45A3	1	1q32.1	Amp
IRF6	1	1q32.2	Amp
EGFR	1	7p11.2	SNV

R1



R2





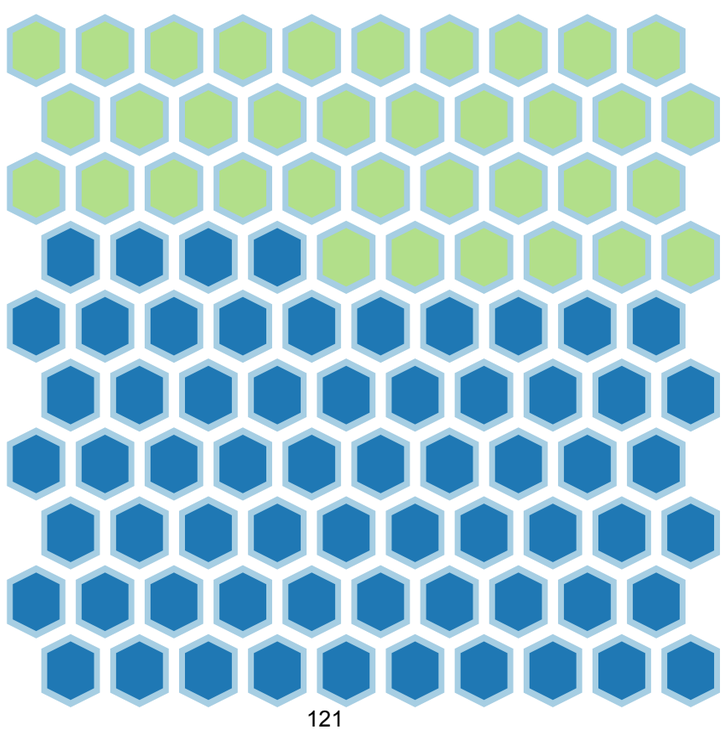
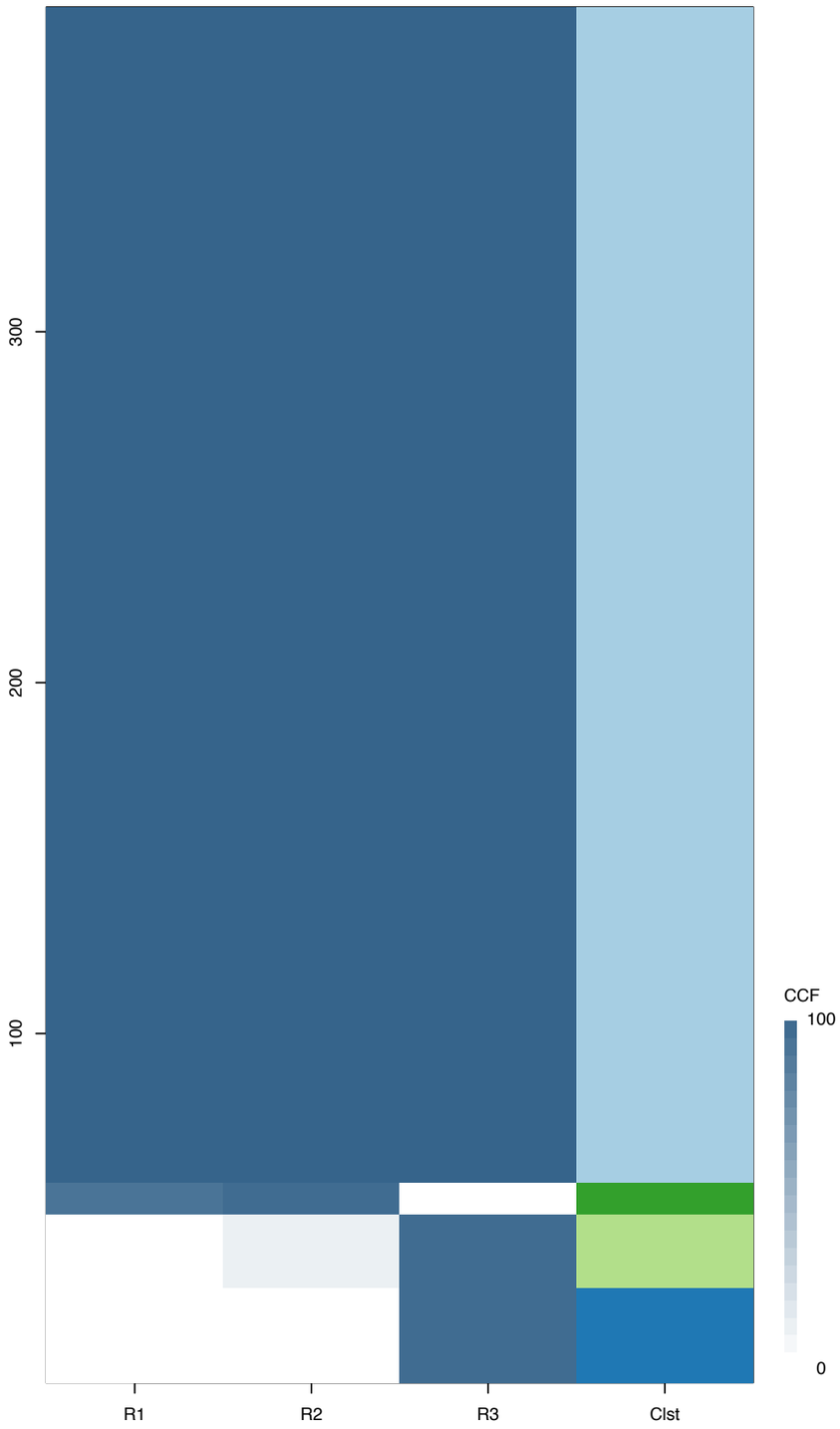
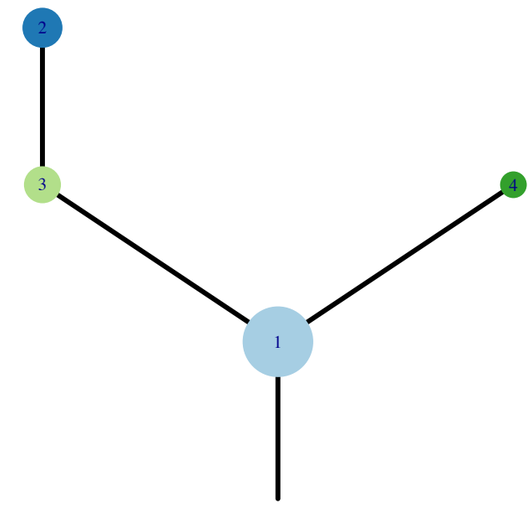


Fig.S12BA



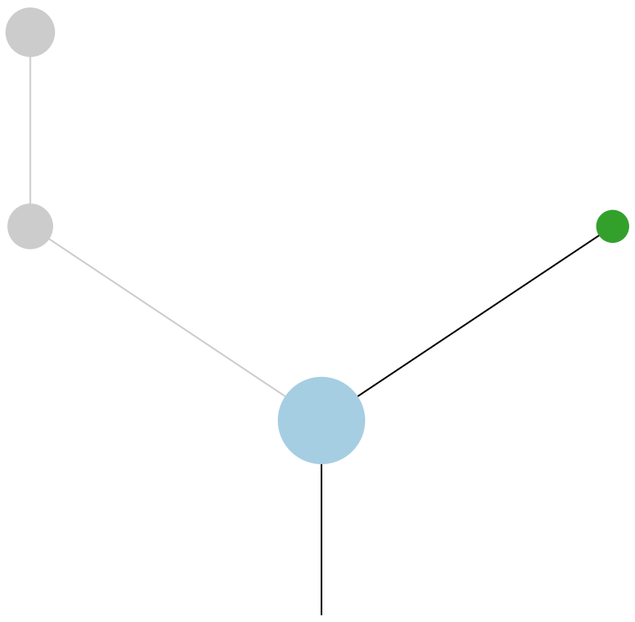
CRUK0056



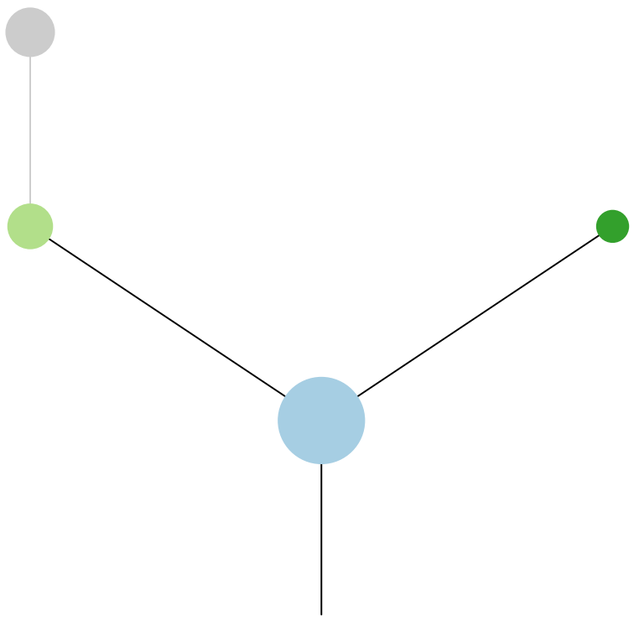
Histology:Adeno, Age:72, PackYears:90, Size:25
Stage:1b, Gender:Female, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
RASA1	1	5q14.3	SNV
CREBBP	1	16p13.3	SNV
ZFP36L1	2	14q24.1	SNV
TP53	3	17p13.1	SNV

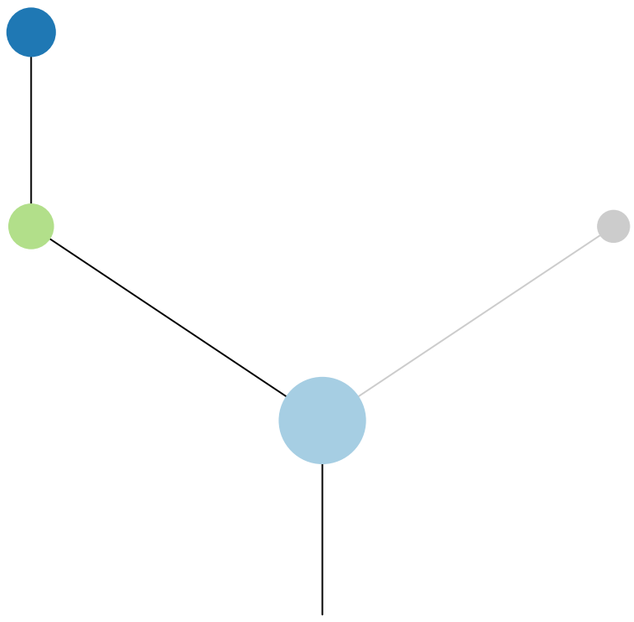
R1

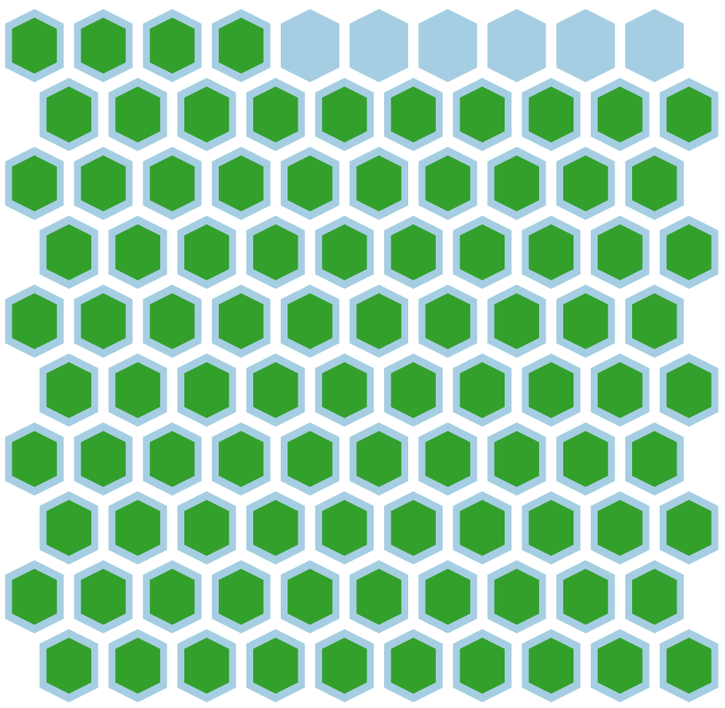


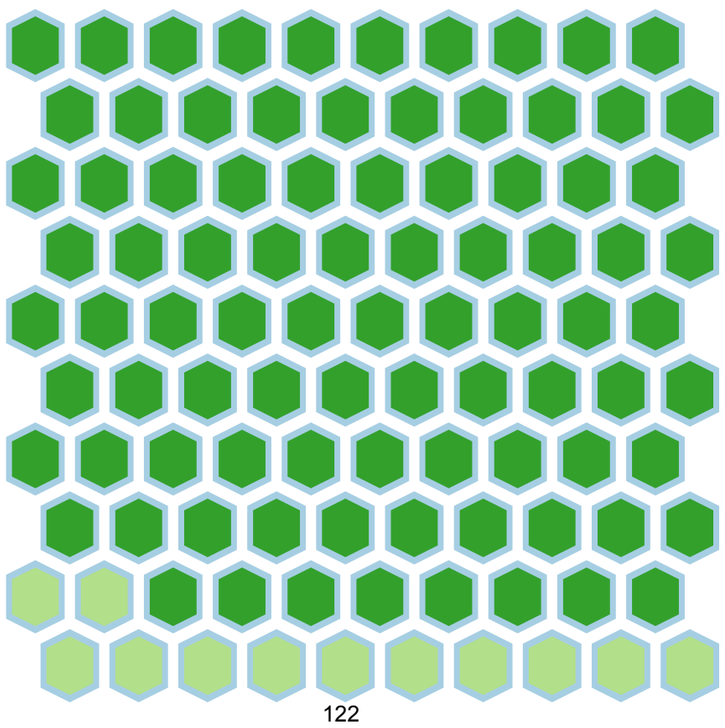
R2



R3







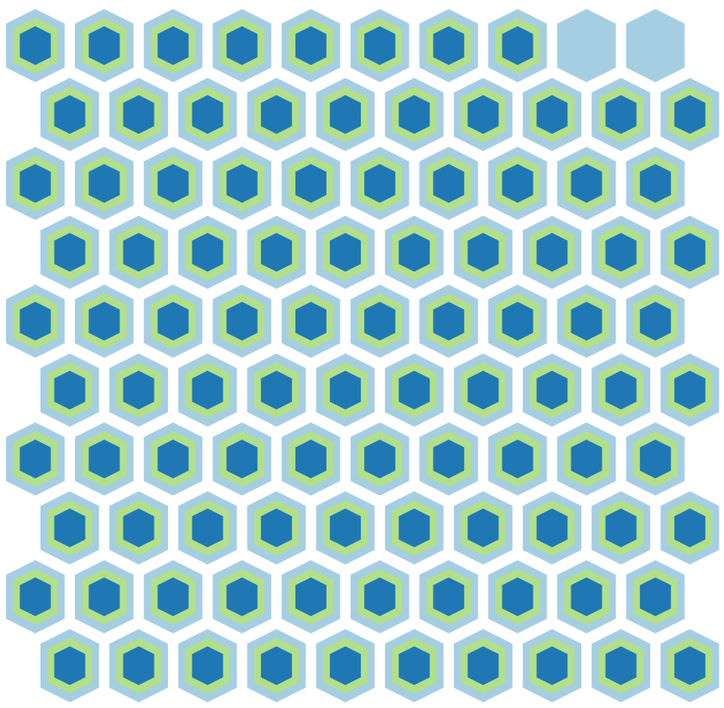
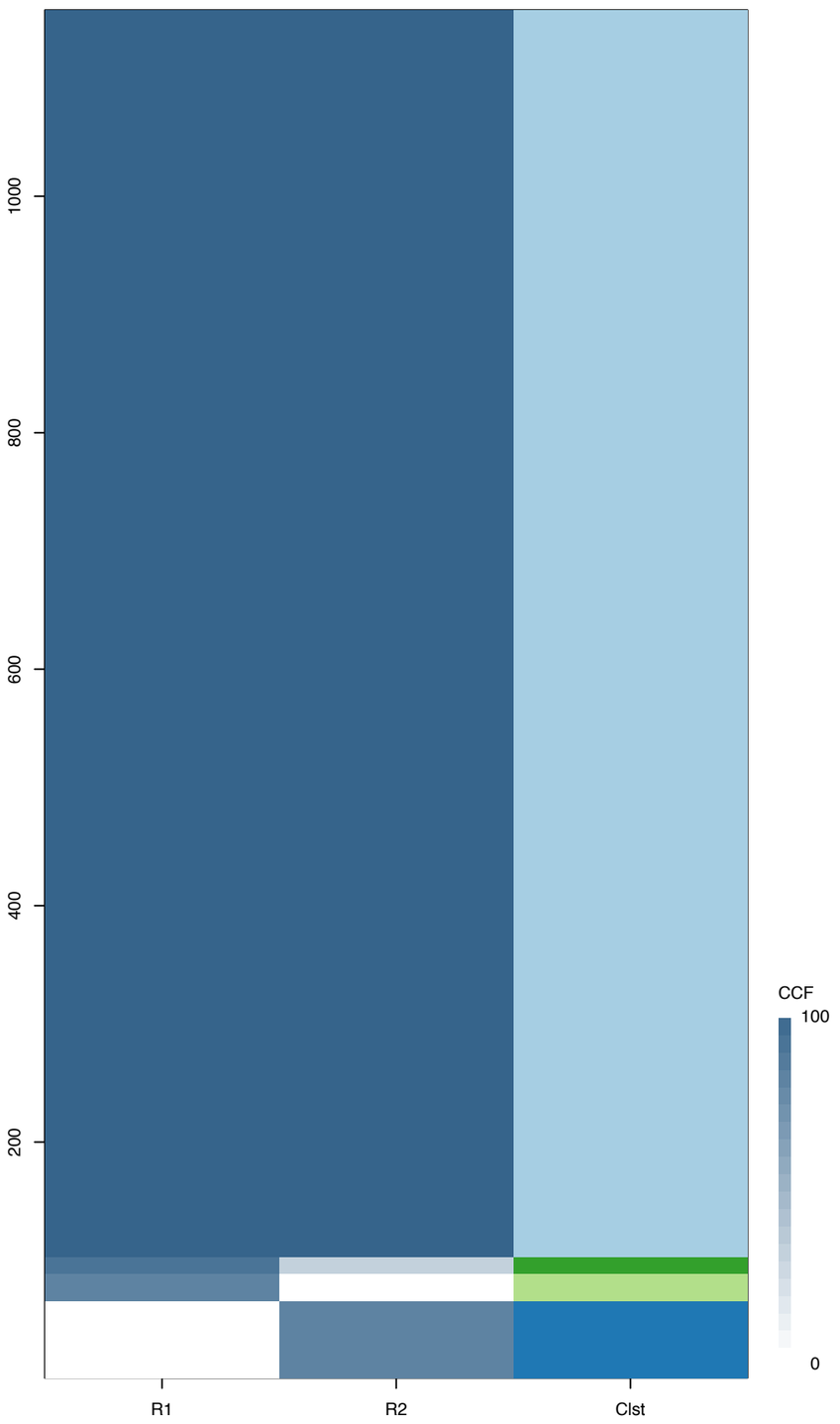
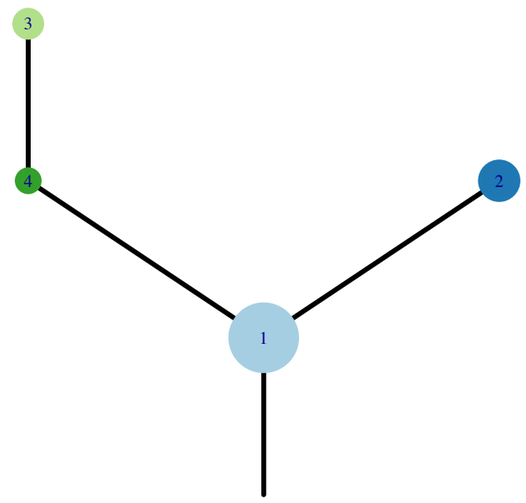


Fig.S12BB



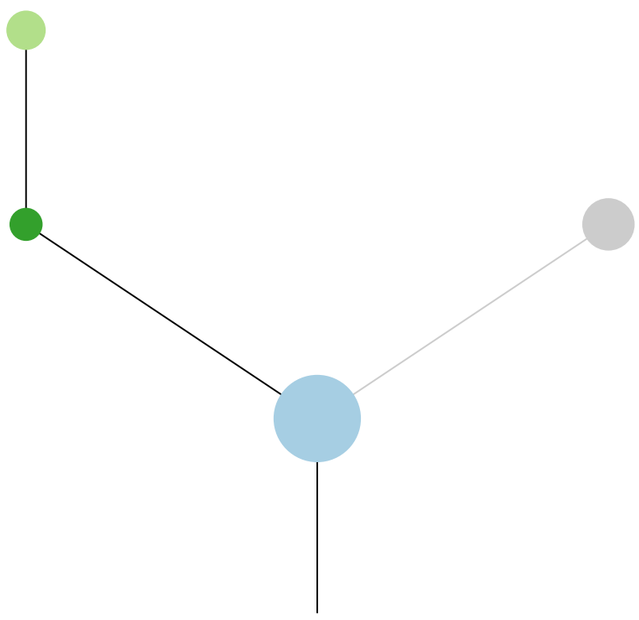
CRUK0057



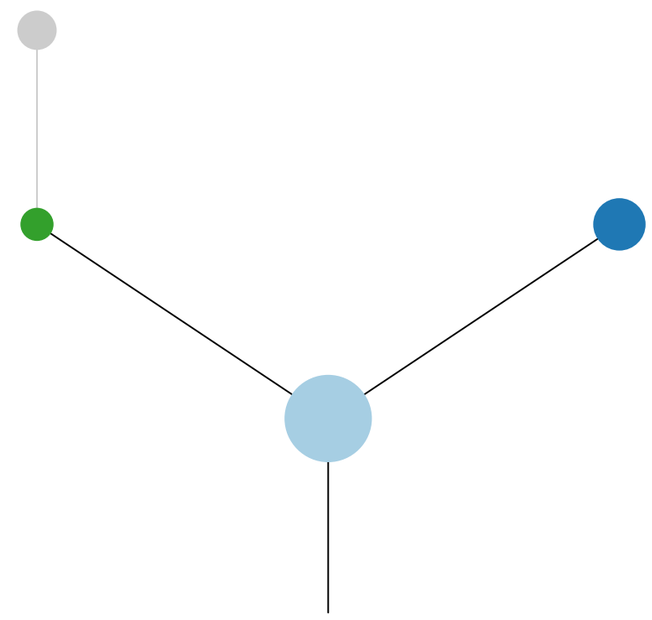
Histology:Adeno, Age:69, PackYears:53, Size:22
Stage:1b, Gender:Female, GD:Clonal GD, Recur:no

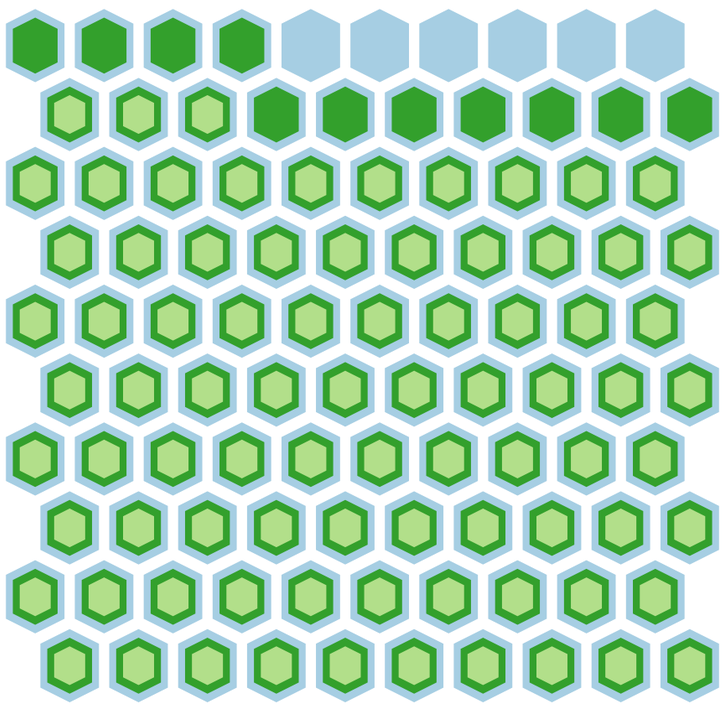
Gene	Cluster	Cytoband	Type
BCL9	1	1q21.2	Amp
ARNT	1	1q21.3	Amp
SETDB1	1	1q21.3	Amp
MLLT11	1	1q21.3	Amp
TERT	1	5p15.33	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
KRAS	1	12p12.1	SNV
TSC2	1	16p13.3	SNV
MYH11	1	16p13.11	Amp
DNM2	1	19p13.2	SNV
SMARCA4	1	19p13.2	SNV
WAS	?		SNV
ATRX	?		SNV

R1



R2





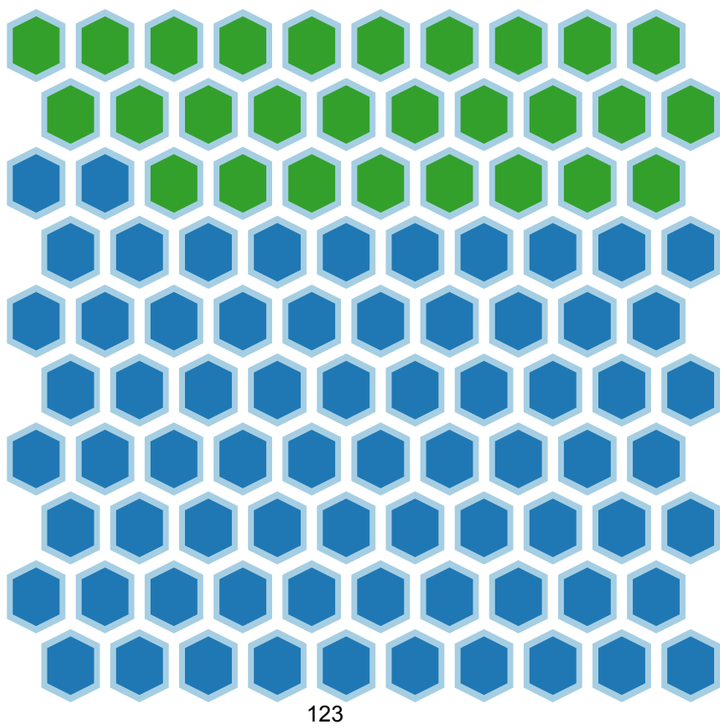
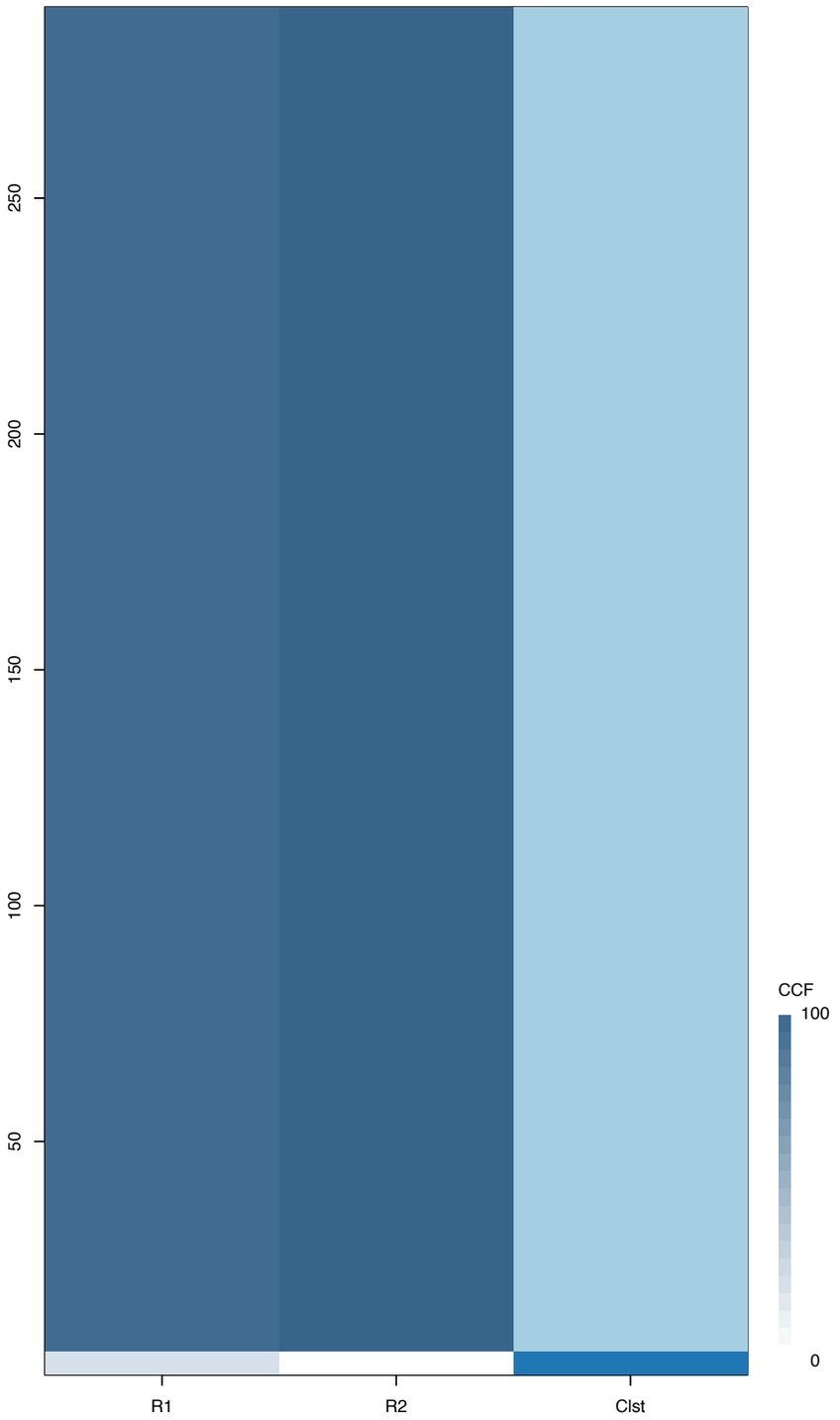
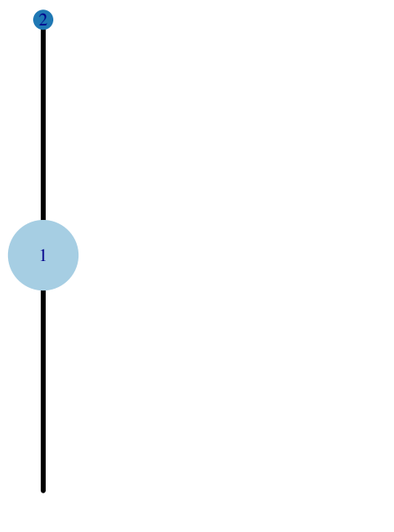


Fig.S12BC



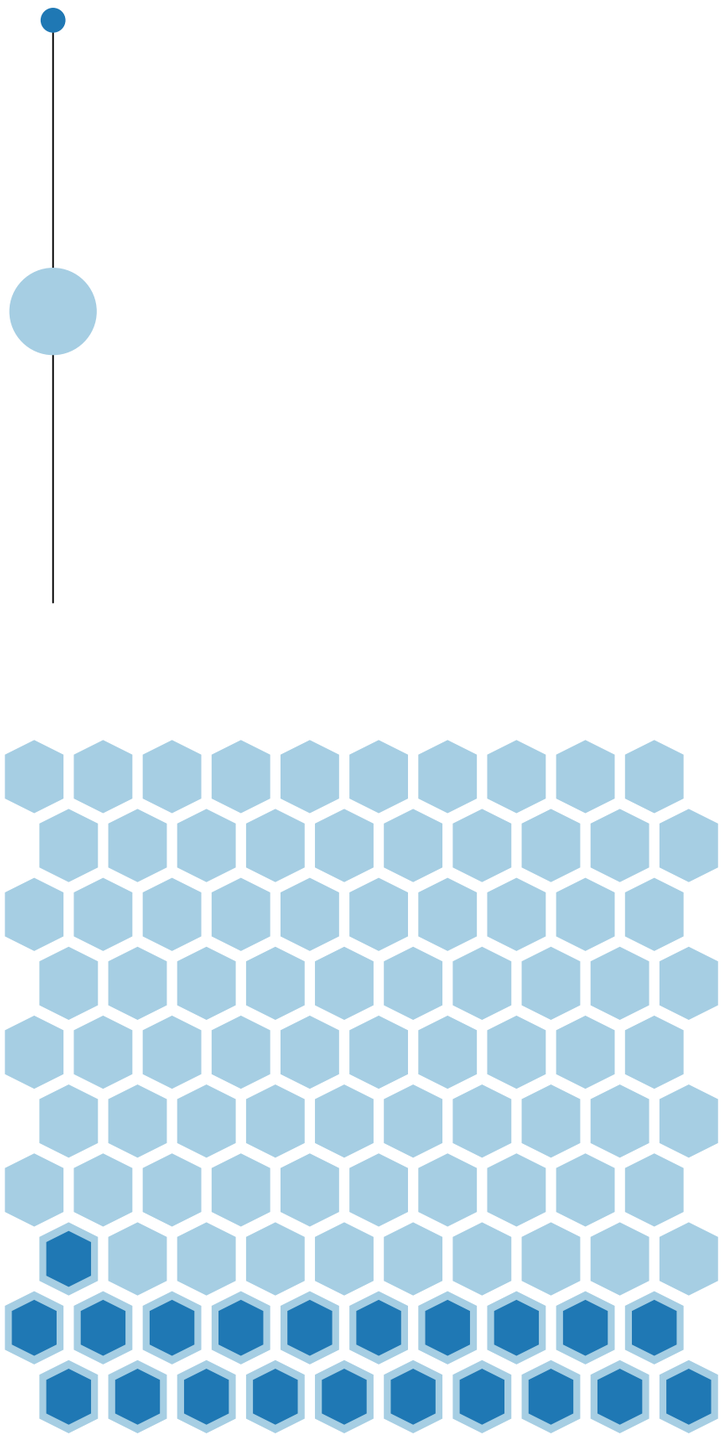
CRUK0061



Histology:Adeno, Age:83, PackYears:40, Size:30
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
CBFA2T3	1	16q24.3	Amp
STK11	1	19p13.3	SNV

R1



R2

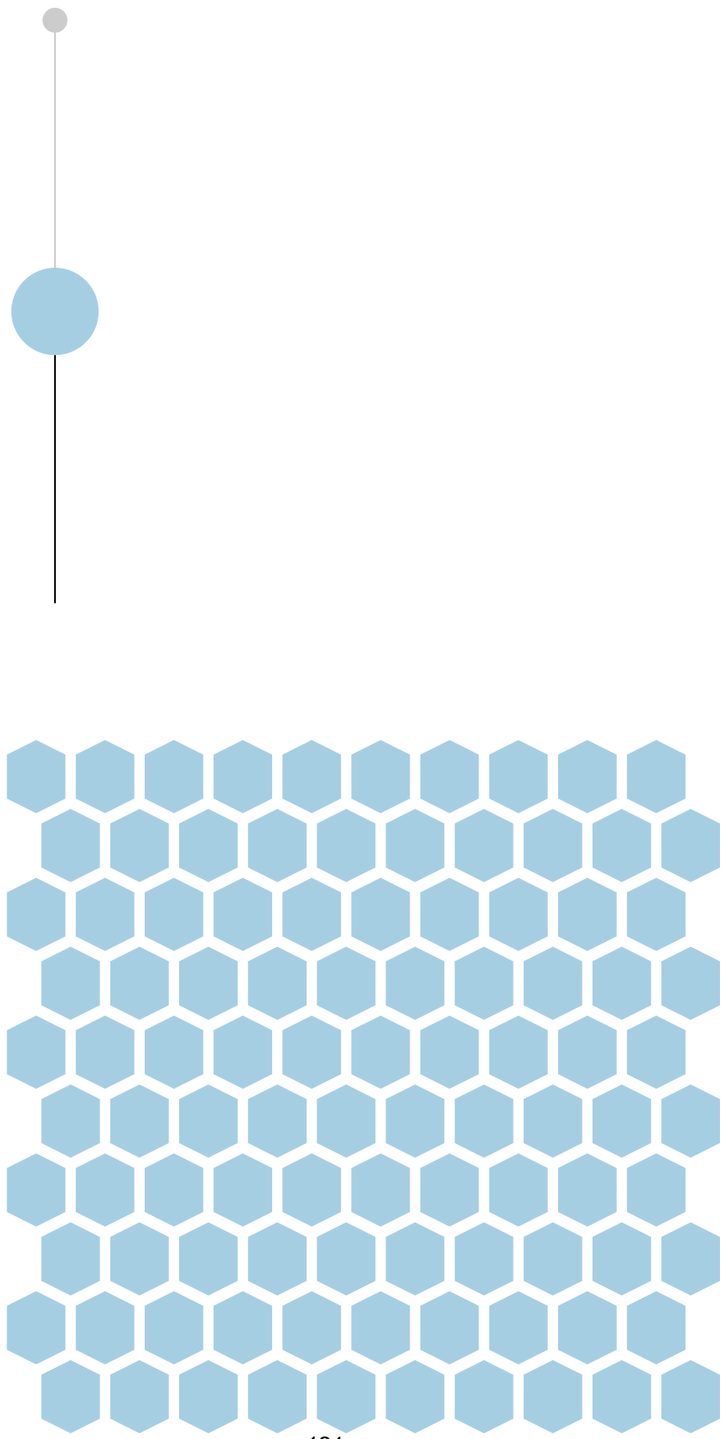
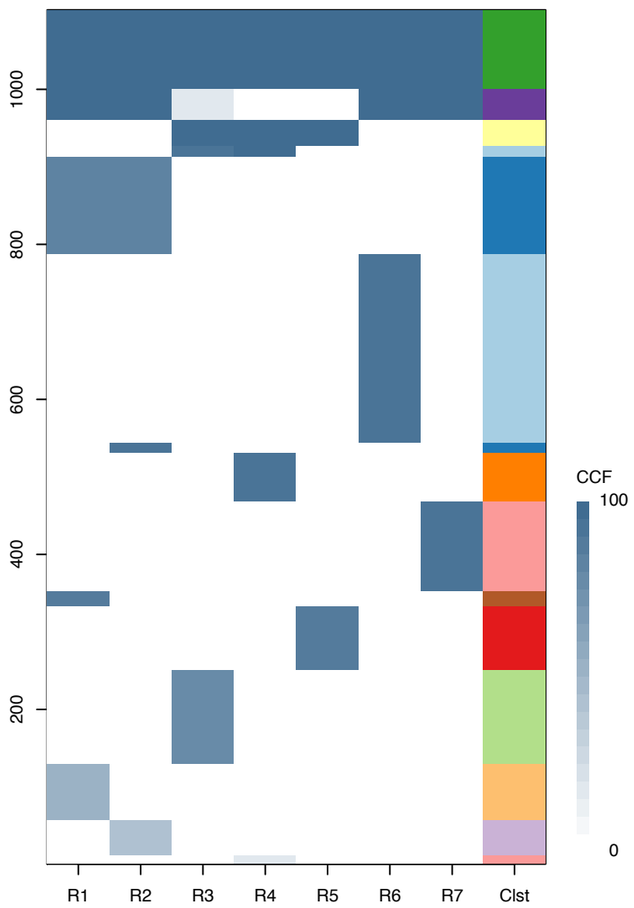
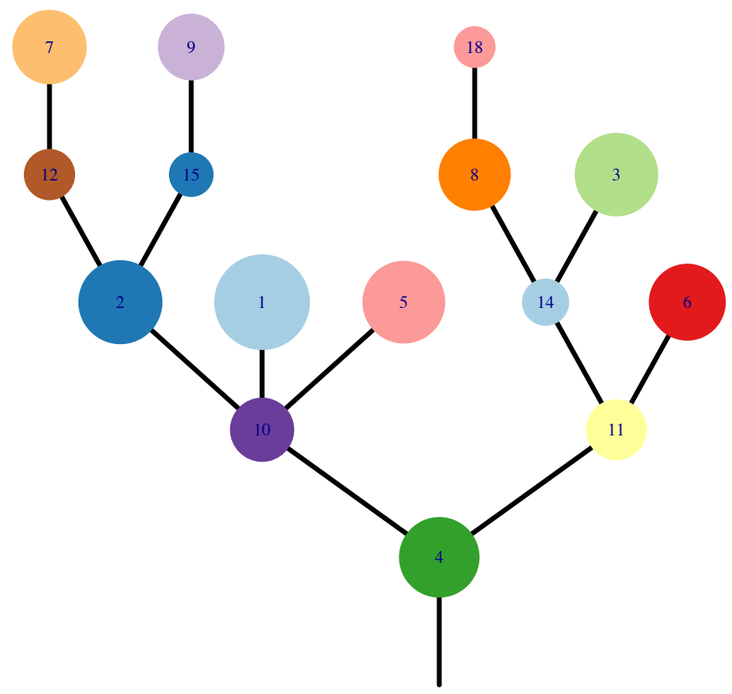


Fig.S12BD

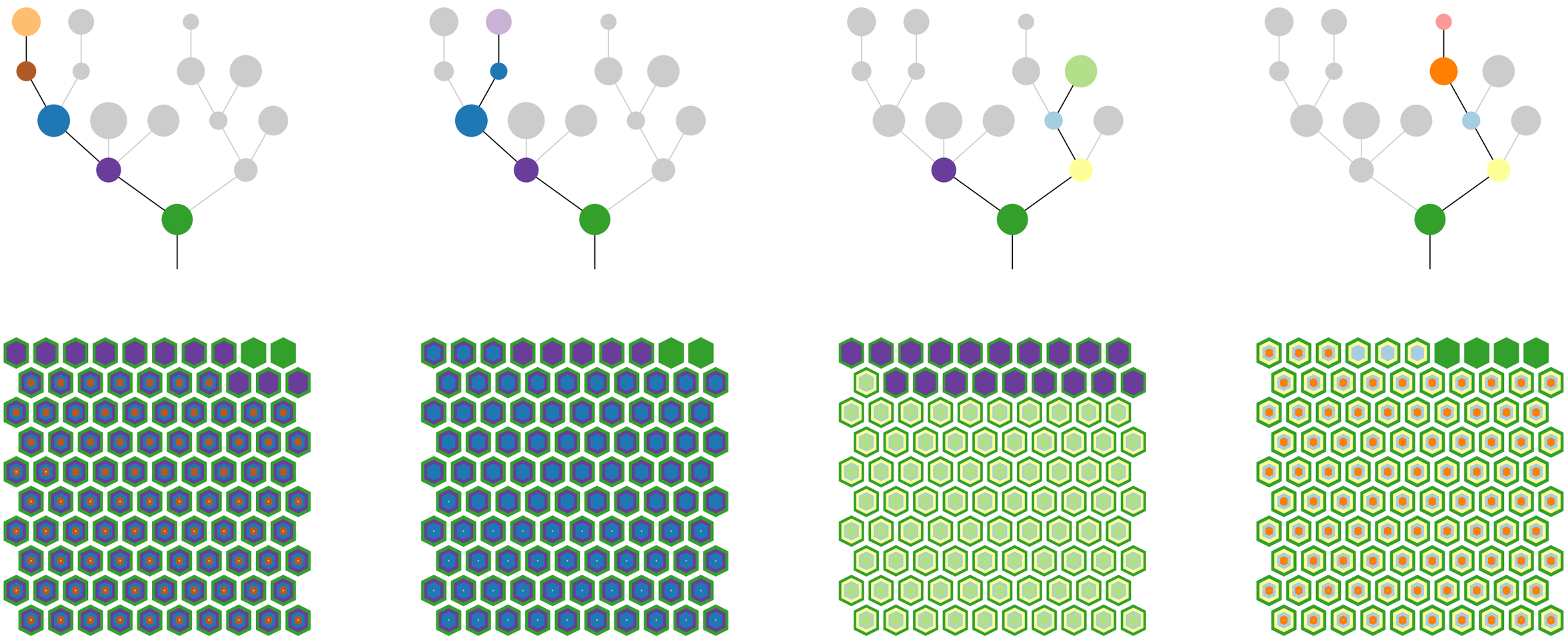


CRUK0062



Histology:Squamous, Age:58, PackYears:19, Size:55
Stage:2b, Gender:Male, GD:Subclonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
FAS	1	10q23.31	SNV
RNF217−AS1	2	6q22.31	Amp
IL21R	2	16p12.1	Amp
PLXNB2	2	22q13.33	SNV
KAT6A	3	8p11.21	Amp
IKBKB	3	8p11.21	Amp
HOOK3	3	8p11.21	Amp
FOXL2	4	3q22.3	Amp
WWTR1	4	3q25.1	Amp
GMPS	4	3q25.31	Amp
MLF1	4	3q25.32	Amp
MECOM	4	3q26.2	Amp
PIK3CA	4	3q26.32	Amp
SOX2	4	3q26.33	Amp
ETV5	4	3q27.2	Amp
EIF4A2	4	3q27.3	Amp
BCL6	4	3q27.3	Amp
LPP	4	3q27.3	Amp
TFRC	4	3q29	Amp
CCND1	4	11q13.3	Amp
TP53	4	17p13.1	SNV
BLM	6	15q26.1	SNV
BCR	8	22q11.23	Amp
PAX5	11	9p13.2	Amp
FIP1L1	?	4q12	Amp
NSD1	?	5q35.3	SNV
UBR5	?	8q22.3	SNV
ATRX	?		SNV



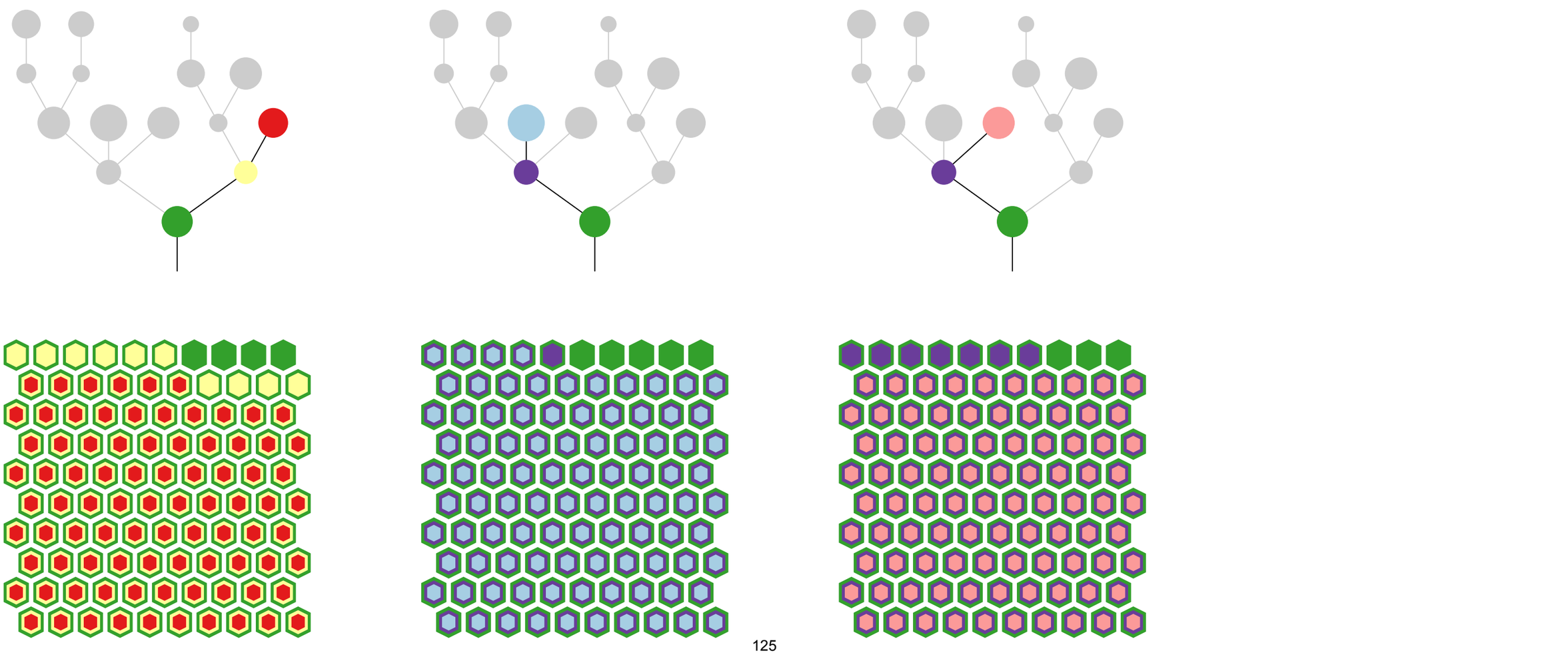
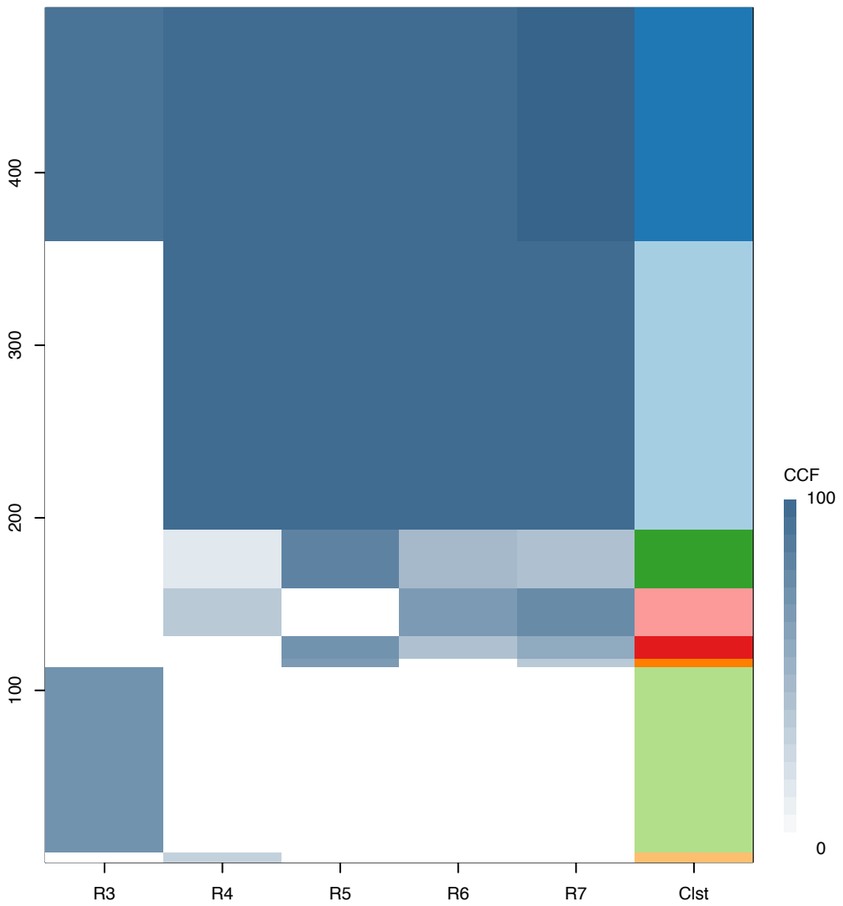
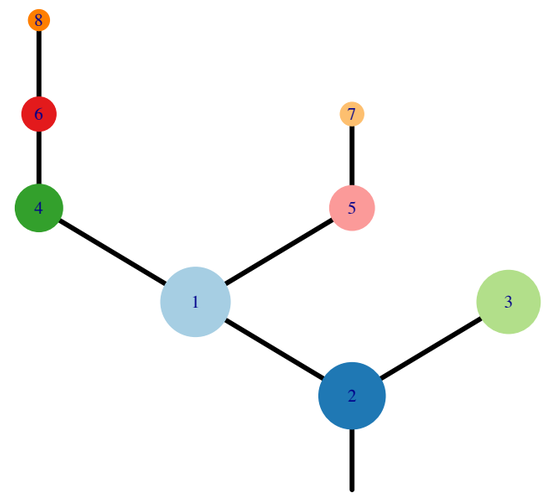


Fig.S12BE



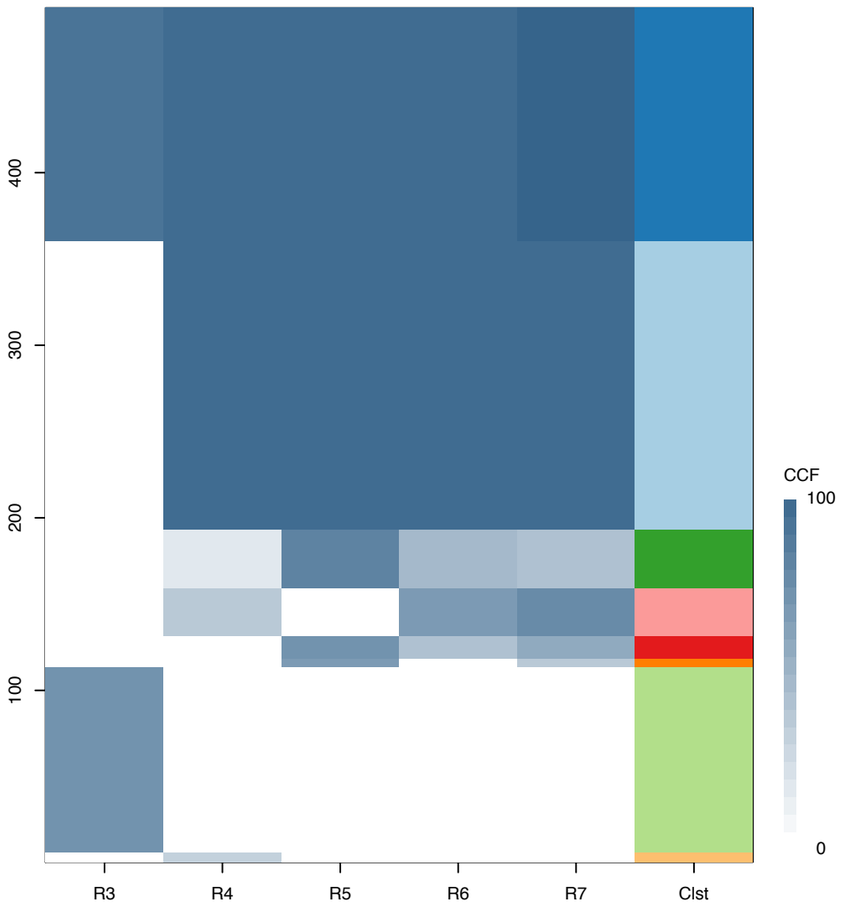
CRUK0063_A



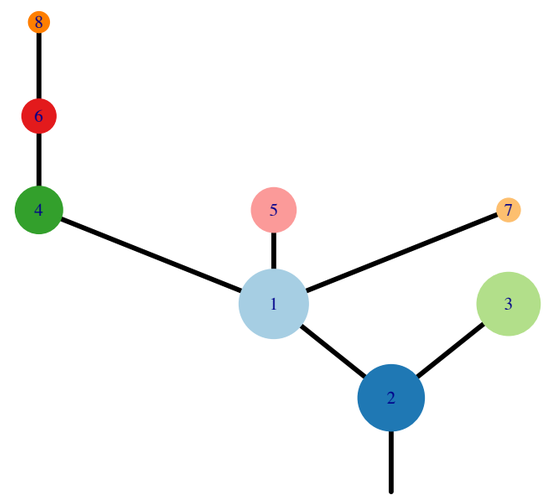
Histology:Squamous, Age:66, PackYears:57.5, Size:50
Stage:2a, Gender:Male, GD:Subclonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
RET	1	10q11.21	Amp
CYLD	1	16q12.1	SNV
NF1	1	17q11.2	SNV
MSH2	2	2p21	SNV
WWTR1	2	3q25.1	Amp
GMPS	2	3q25.31	Amp
MLF1	2	3q25.32	Amp
MECOM	2	3q26.2	Amp
PIK3CA	2	3q26.32	Amp
PIK3CA	2	3q26.32	SNV
SOX2	2	3q26.33	Amp
ETV5	2	3q27.2	Amp
EIF4A2	2	3q27.3	Amp
BCL6	2	3q27.3	Amp
LPP	2	3q27.3	Amp
TFRC	2	3q29	Amp
FBXW7	2	4q31.3	SNV
TERT	2	5p15.33	Amp
IL7R	2	5p13.2	Amp
CDKN2A	2	9p21.3	SNV
PRF1	2	10q22.1	Del
KLF5	2	13q22.1	SNV
TP53	2	17p13.1	SNV
LMO2	3	11p13	Amp
FANCM	6	14q21.2	SNV
CDKN2A	?	9p21.3	SNV
EP300	?	22q13.2	SNV

Fig.S12BE



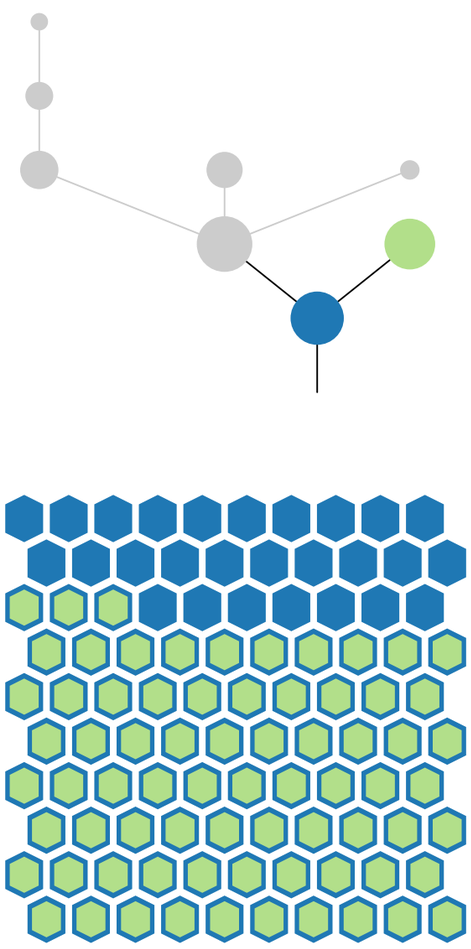
CRUK0063_B



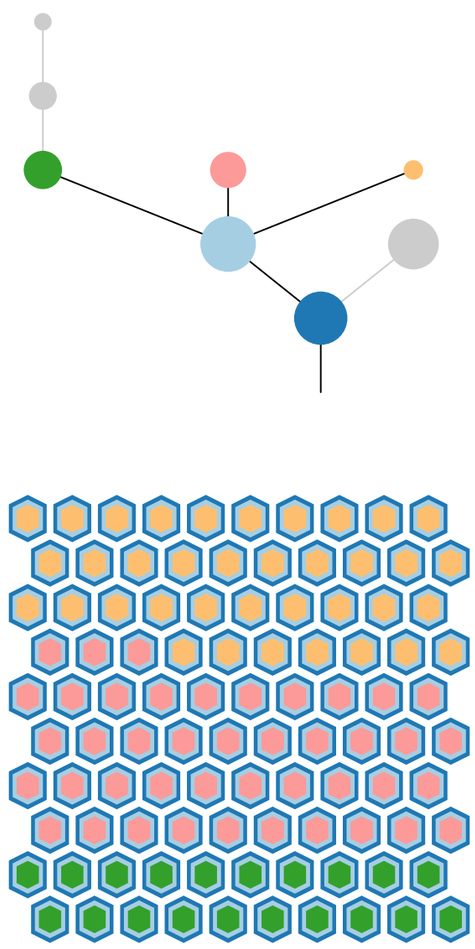
Histology:Squamous, Age:66, PackYears:57.5, Size:50
Stage:2a, Gender:Male, GD:Subclonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
RET	1	10q11.21	Amp
CYLD	1	16q12.1	SNV
NF1	1	17q11.2	SNV
MSH2	2	2p21	SNV
WWTR1	2	3q25.1	Amp
GMPS	2	3q25.31	Amp
MLF1	2	3q25.32	Amp
MECOM	2	3q26.2	Amp
PIK3CA	2	3q26.32	Amp
PIK3CA	2	3q26.32	SNV
SOX2	2	3q26.33	Amp
ETV5	2	3q27.2	Amp
EIF4A2	2	3q27.3	Amp
BCL6	2	3q27.3	Amp
LPP	2	3q27.3	Amp
TFRC	2	3q29	Amp
FBXW7	2	4q31.3	SNV
TERT	2	5p15.33	Amp
IL7R	2	5p13.2	Amp
CDKN2A	2	9p21.3	SNV
PRF1	2	10q22.1	Del
KLF5	2	13q22.1	SNV
TP53	2	17p13.1	SNV
LMO2	3	11p13	Amp
FANCM	6	14q21.2	SNV
CDKN2A	?	9p21.3	SNV
EP300	?	22q13.2	SNV

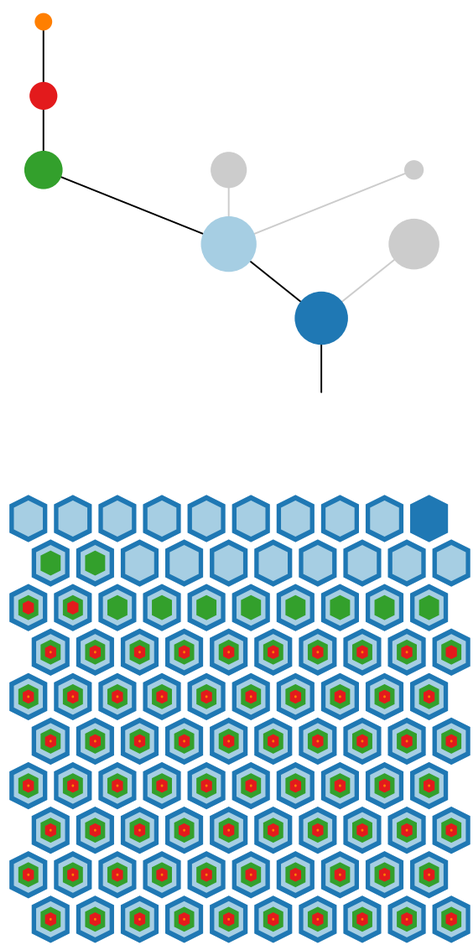
R3



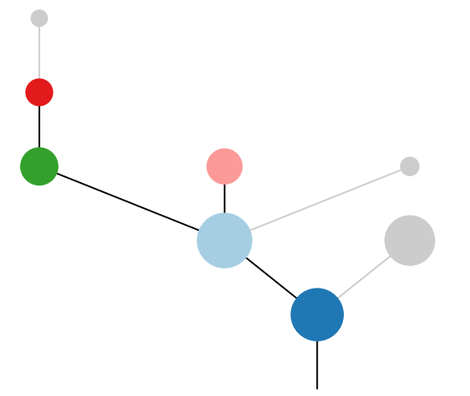
R4



R5



R6



R7

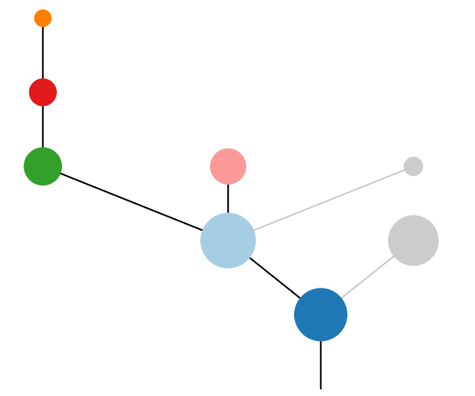
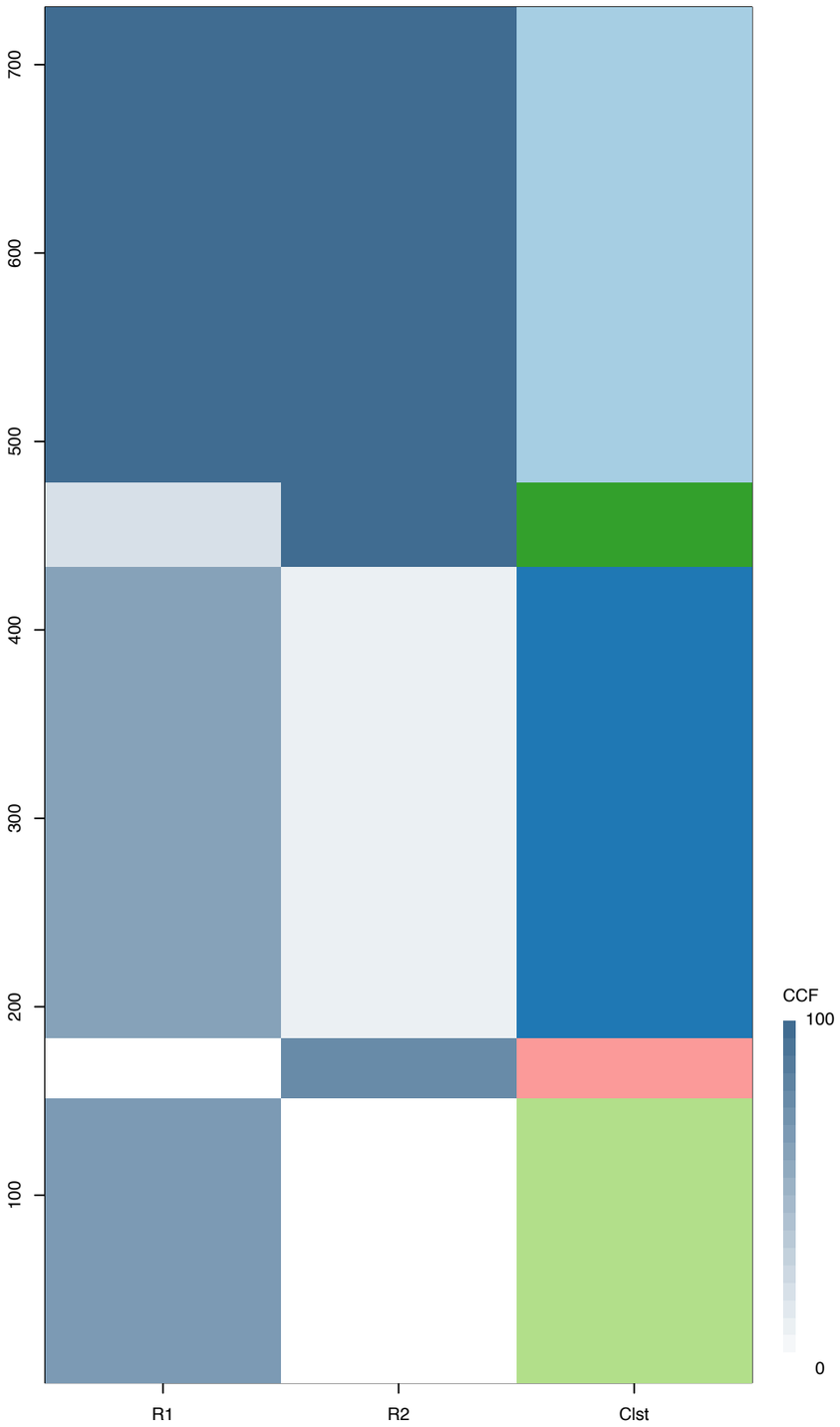
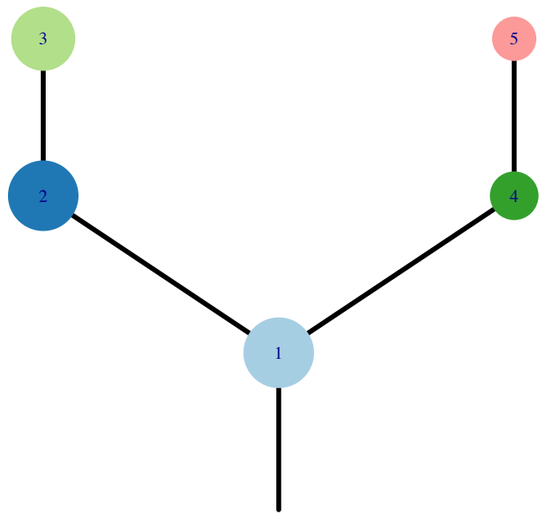


Fig.S12BF



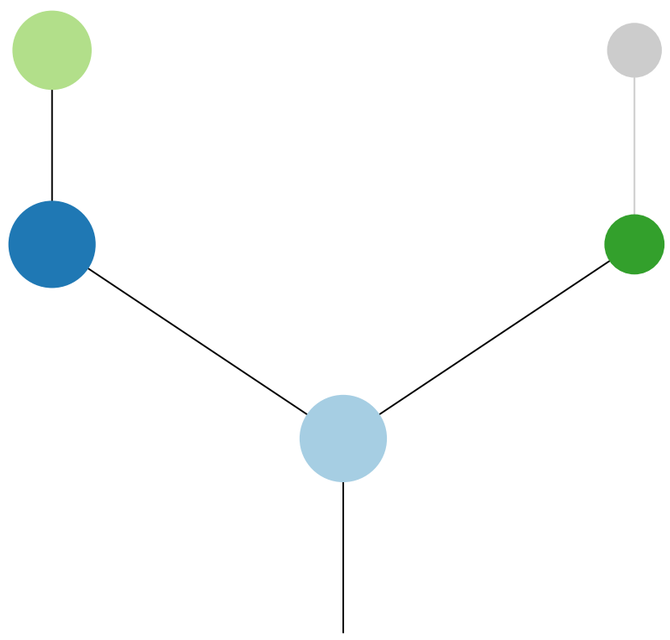
CRUK0064



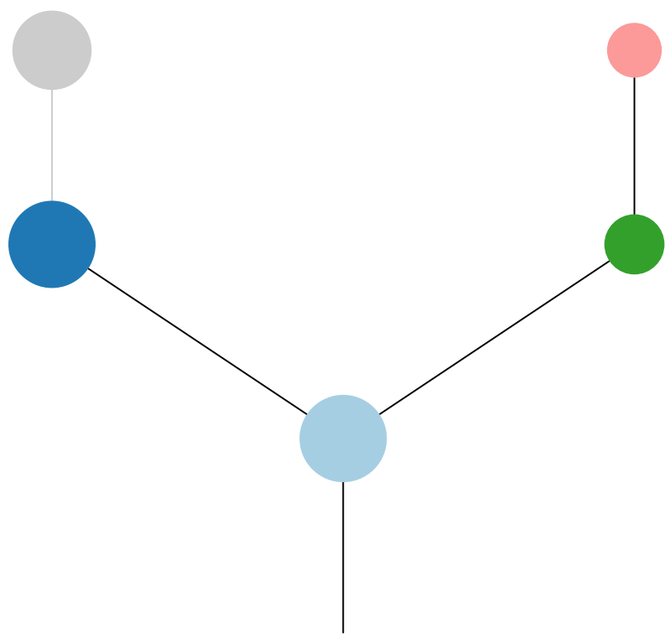
Gene	Cluster	Cytoband	Type
TP53	1	17p13.1	SNV
MLH1	2	3p22.2	SNV
FAT1	2	4q35.2	SNV
FAT1	3	4q35.2	SNV

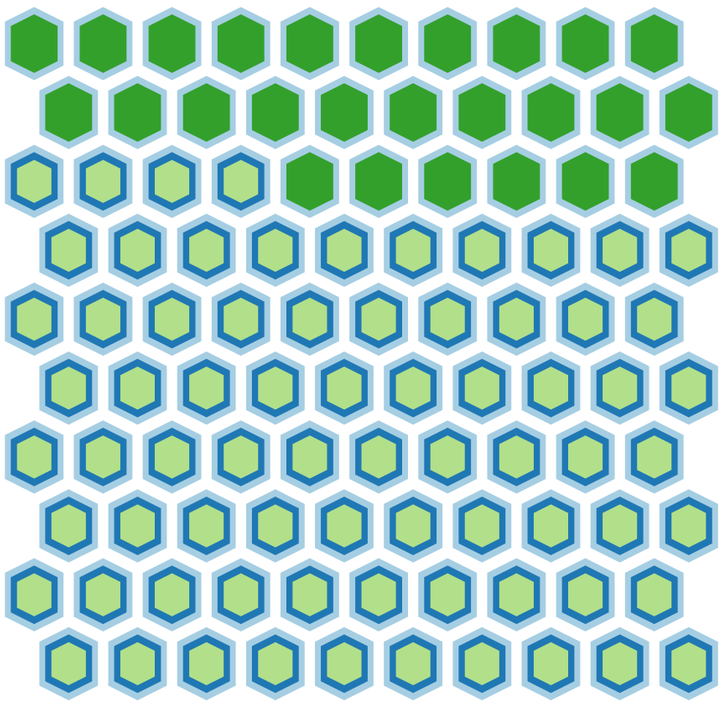
Histology:Squamous, Age:69, PackYears:26.5, Size:38
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

R1



R2





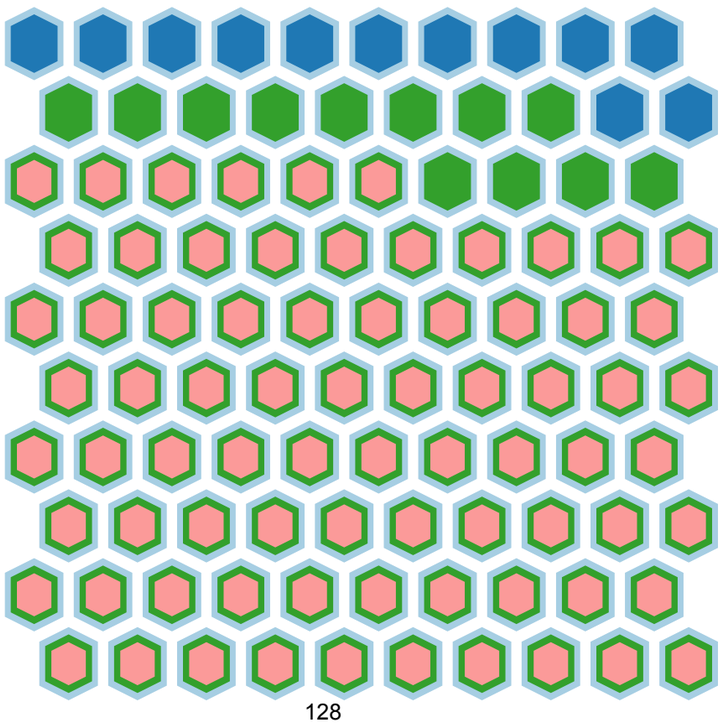
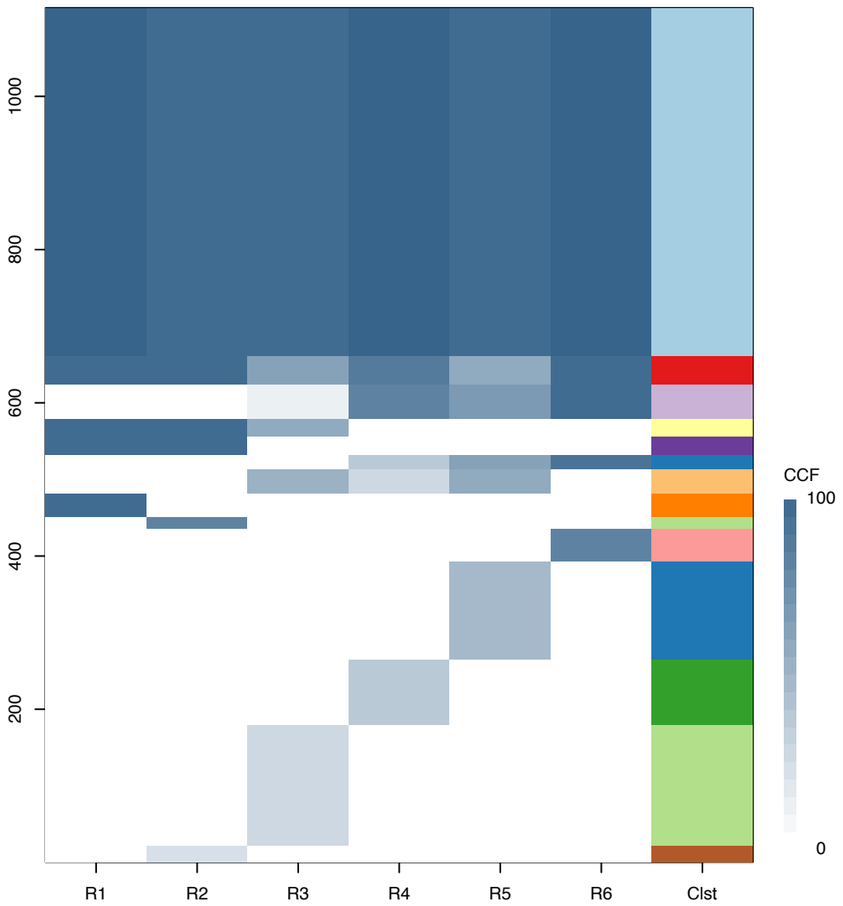
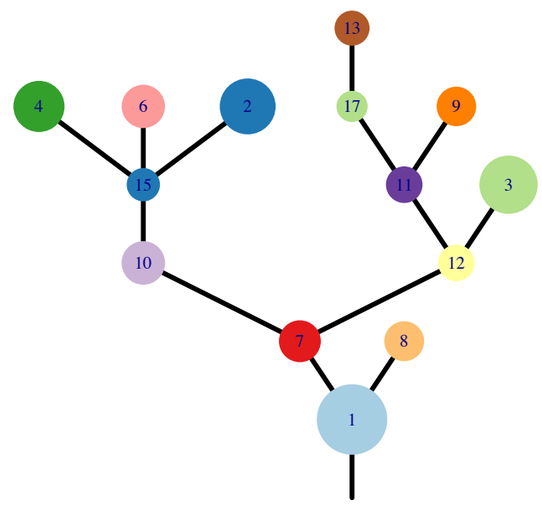


Fig.S12BG



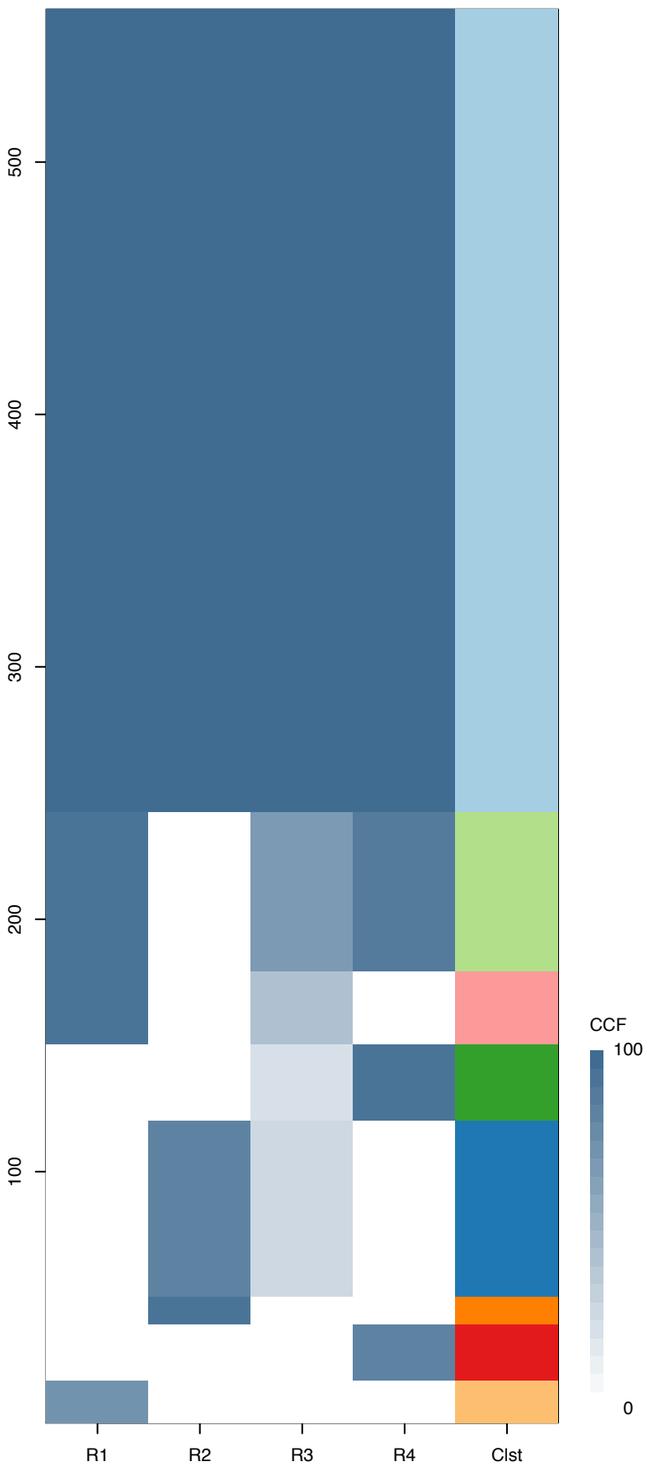
CRUK0065



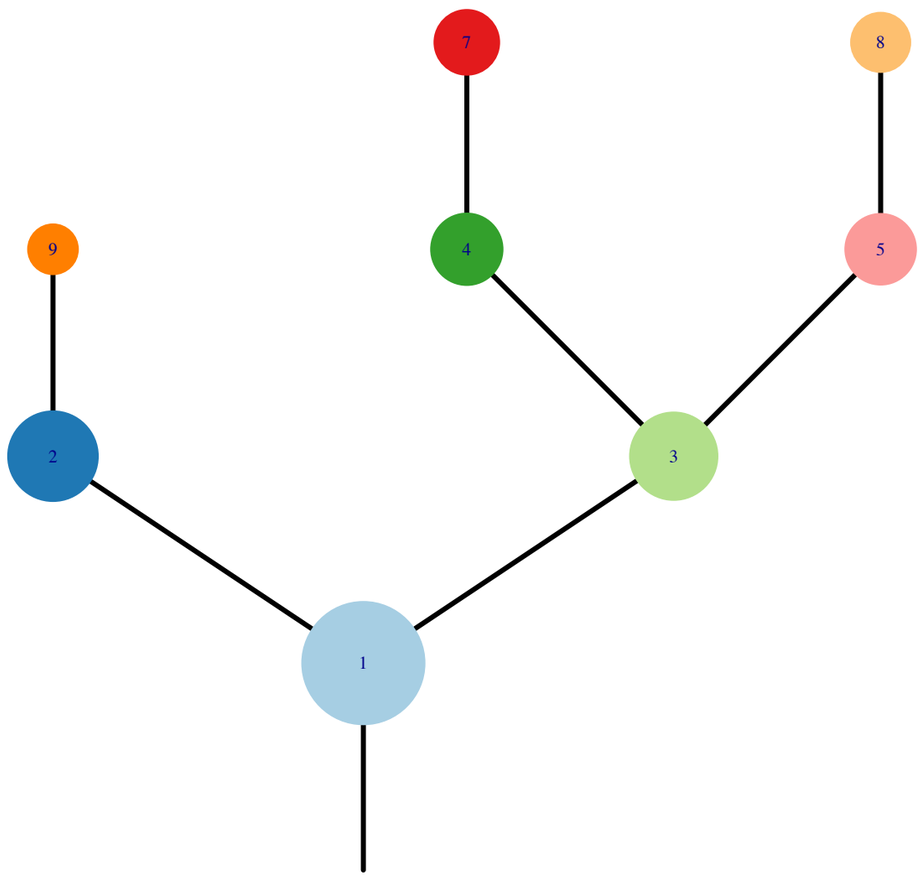
Histology:Squamous, Age:78, PackYears:22.5, Size:75
Stage:2b, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
ARID1A	1	1p36.11	SNV
NFE2L2	1	2q31.2	SNV
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
STX2	1	12q24.33	SNV
NIN	1	14q22.1	SNV
TP53	1	17p13.1	SNV
PTPRC	2	1q32.1	SNV
MLH1	2	3p22.2	SNV
PIK3CA	2	3q26.32	SNV
UBR5	2	8q22.3	SNV
DSN1	3	20q11.23	SNV
NOTCH1	6	9q34.3	SNV
NCOA6	6	20q11.22	SNV
NCOA6	15	20q11.22	SNV
PTEN	?	10q23.31	Del
FAS	?	10q23.31	Del

Fig.S12BH



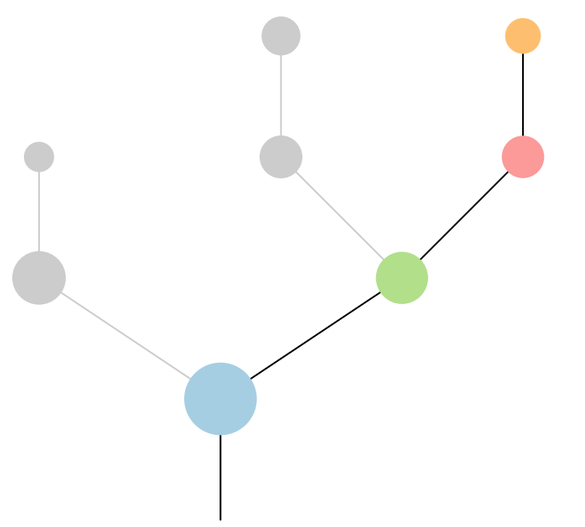
CRUK0066

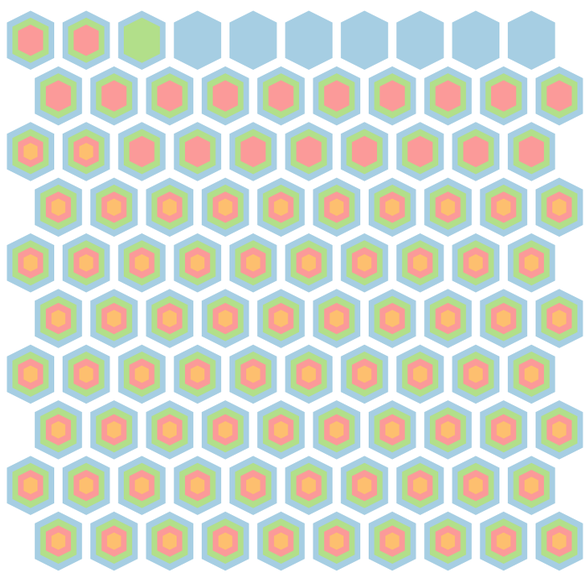


Histology:Squamous, Age:80, PackYears:60, Size:42
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

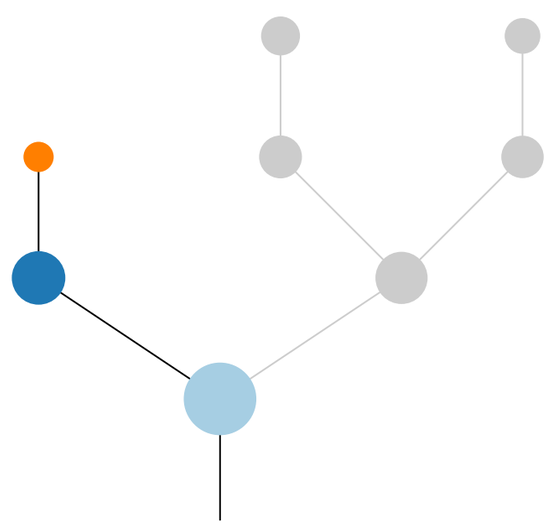
Gene	Cluster	Cytoband	Type
FIP1L1	1	4q12	Amp
CHIC2	1	4q12	Amp
PDGFRA	1	4q12	Amp
KIT	1	4q12	Amp
KDR	1	4q12	Amp
TERT	1	5p15.33	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
WRN	1	8p12	SNV
CDKN2A	1	9p21.3	SNV
NOTCH1	1	9q34.3	SNV
MDM2	1	12q15	Amp
TP53	1	17p13.1	SNV
BRIP1	1	17q23.2	SNV
AKT2	1	19q13.2	Amp
NCOA6	1	20q11.22	Amp
DSN1	1	20q11.23	Amp
MAFB	1	20q12	Amp
TOP1	1	20q12	Amp
PLCG1	1	20q12	Amp
SDC4	1	20q13.12	Amp
NFATC2	1	20q13.2	Amp
AURKA	1	20q13.2	Amp
GNAS	1	20q13.32	Amp
COL5A2	9	2q32.2	SNV

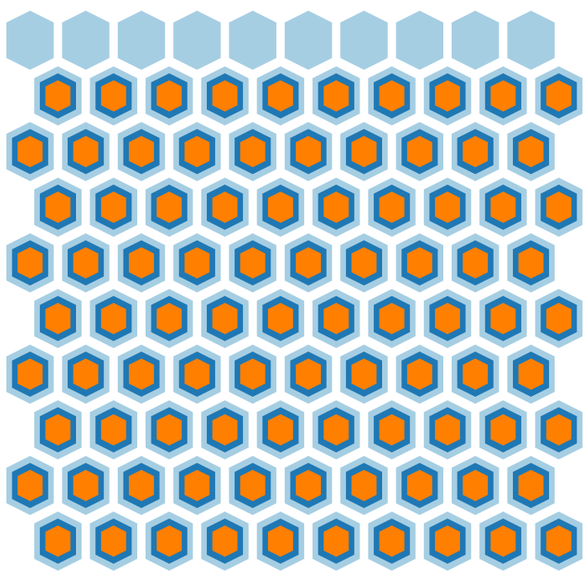
R1



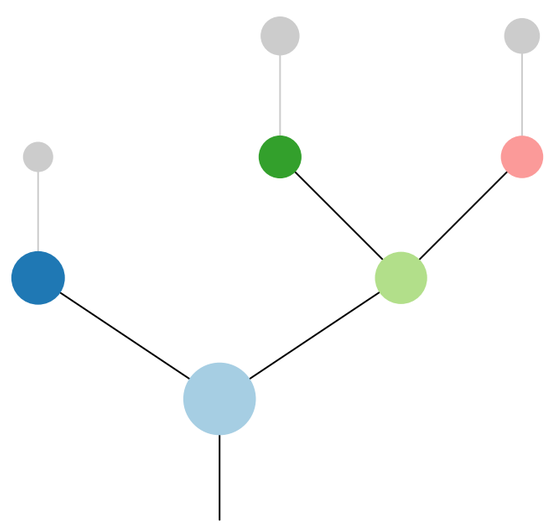


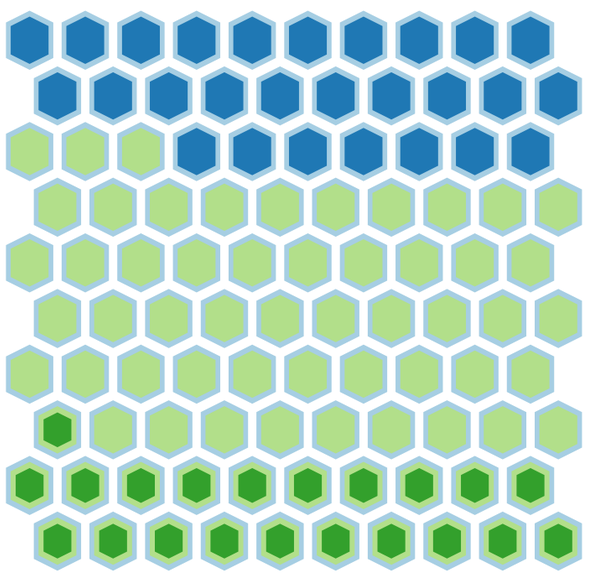
R2



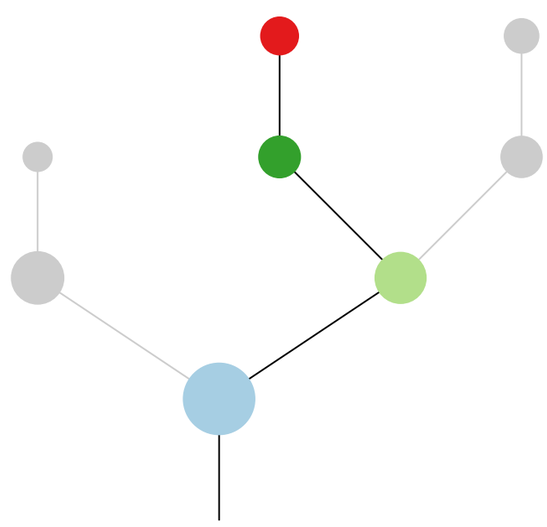


R3





R4



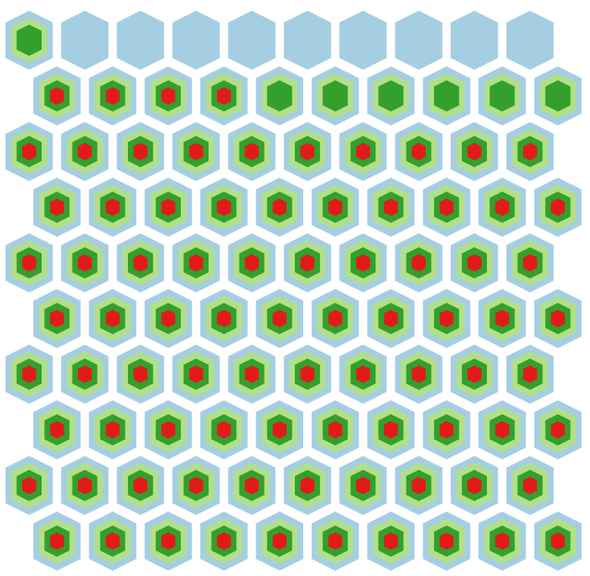
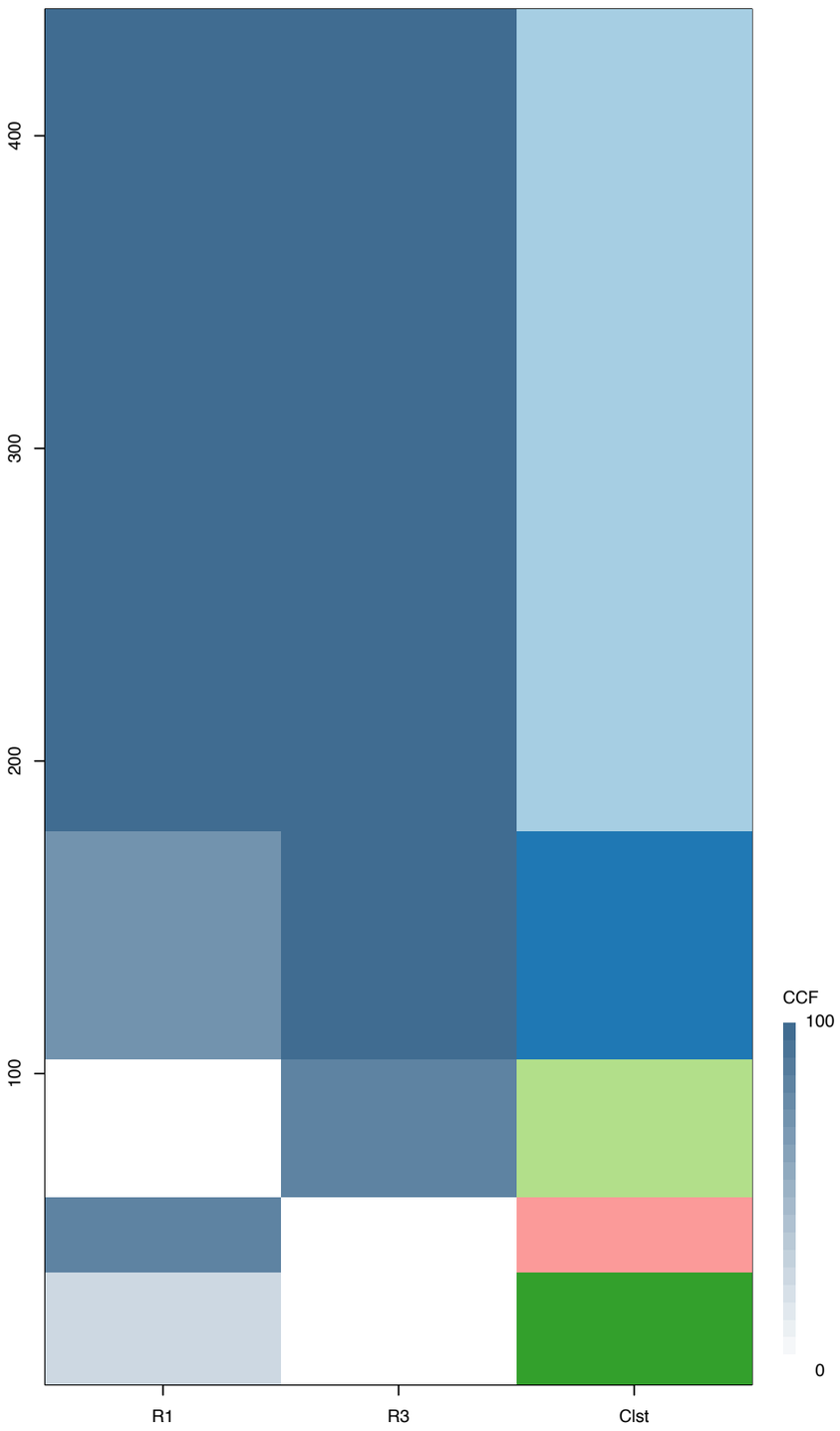
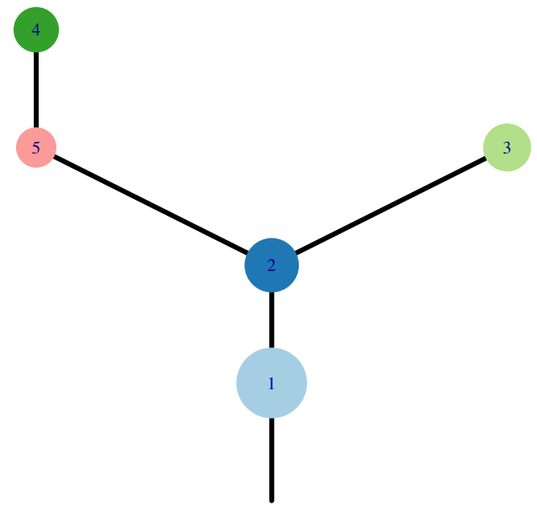


Fig.S12BI



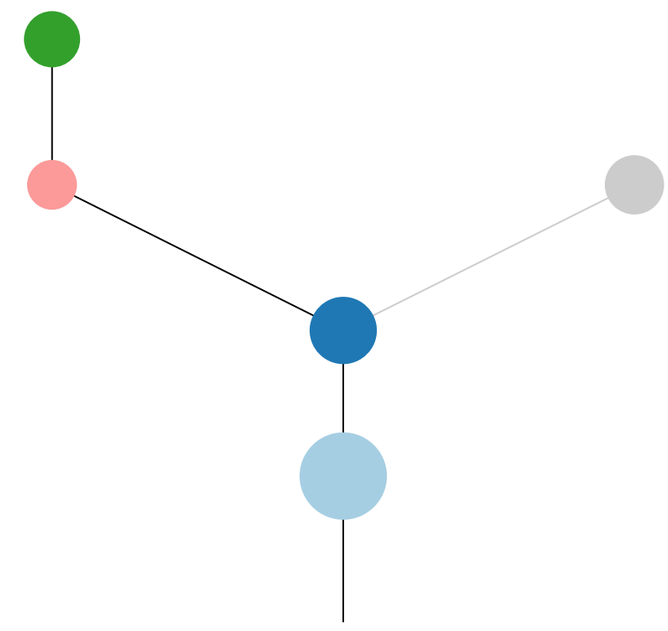
CRUK0067_A



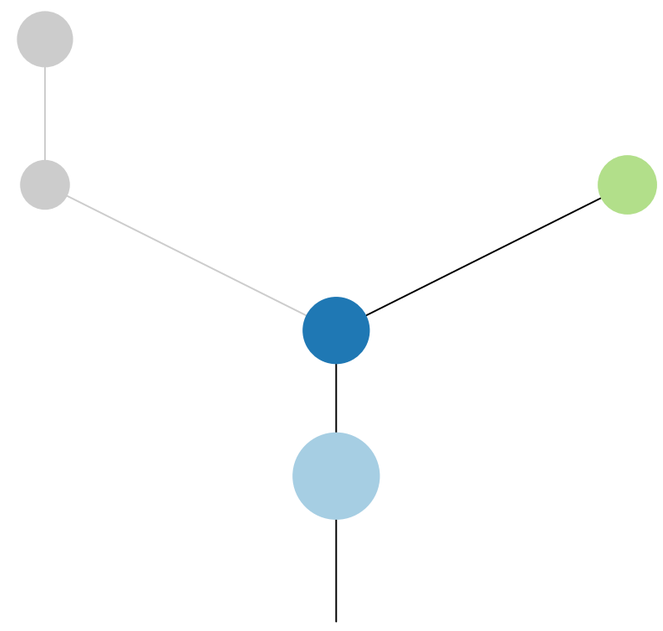
Histology:Squamous, Age:64, PackYears:50, Size:10
Stage:1a, Gender:Male, GD:Not GD, Recur:no

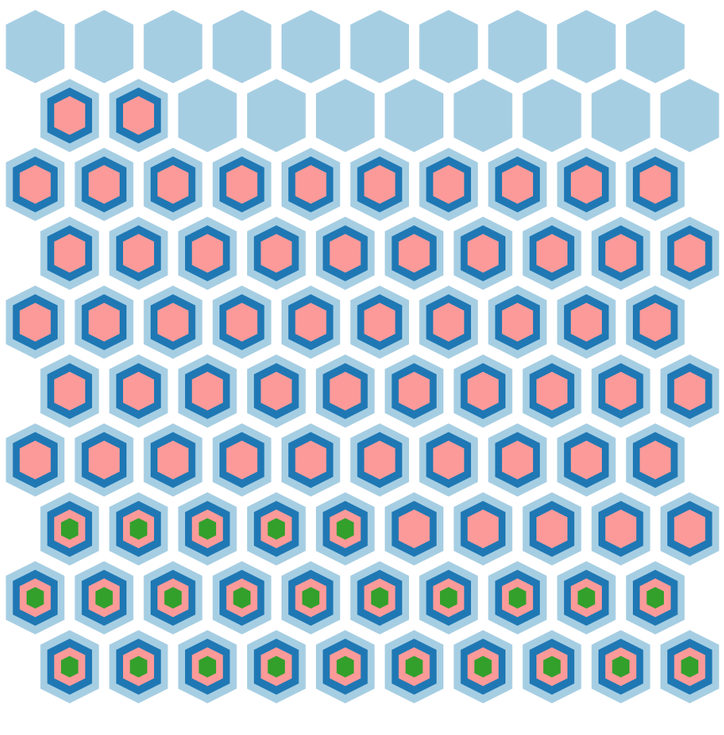
Gene	Cluster	Cytoband	Type
WWTR1	1	3q25.1	Amp
GMPS	1	3q25.31	Amp
MLF1	1	3q25.32	Amp
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
TFEB	1	6p21.1	Amp
CCND3	1	6p21.1	Amp
HSP90AB1	1	6p21.1	Amp
CDKN2A	1	9p21.3	Del
NOTCH1	1	9q34.3	SNV
TP53	1	17p13.1	SNV
NFE2L2	2	2q31.2	SNV
CUL3	?	2q36.2	SNV
KRAS	?	12p12.1	SNV

R1



R3





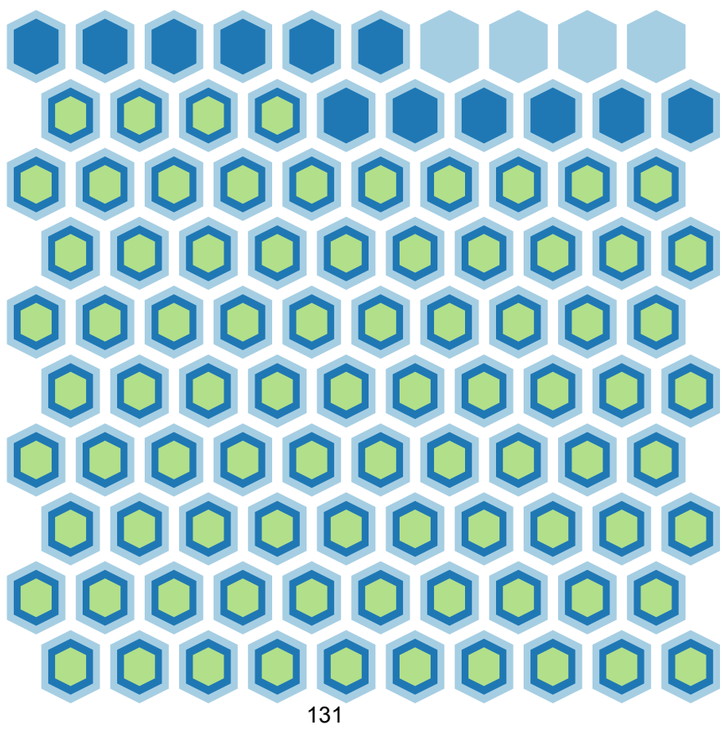
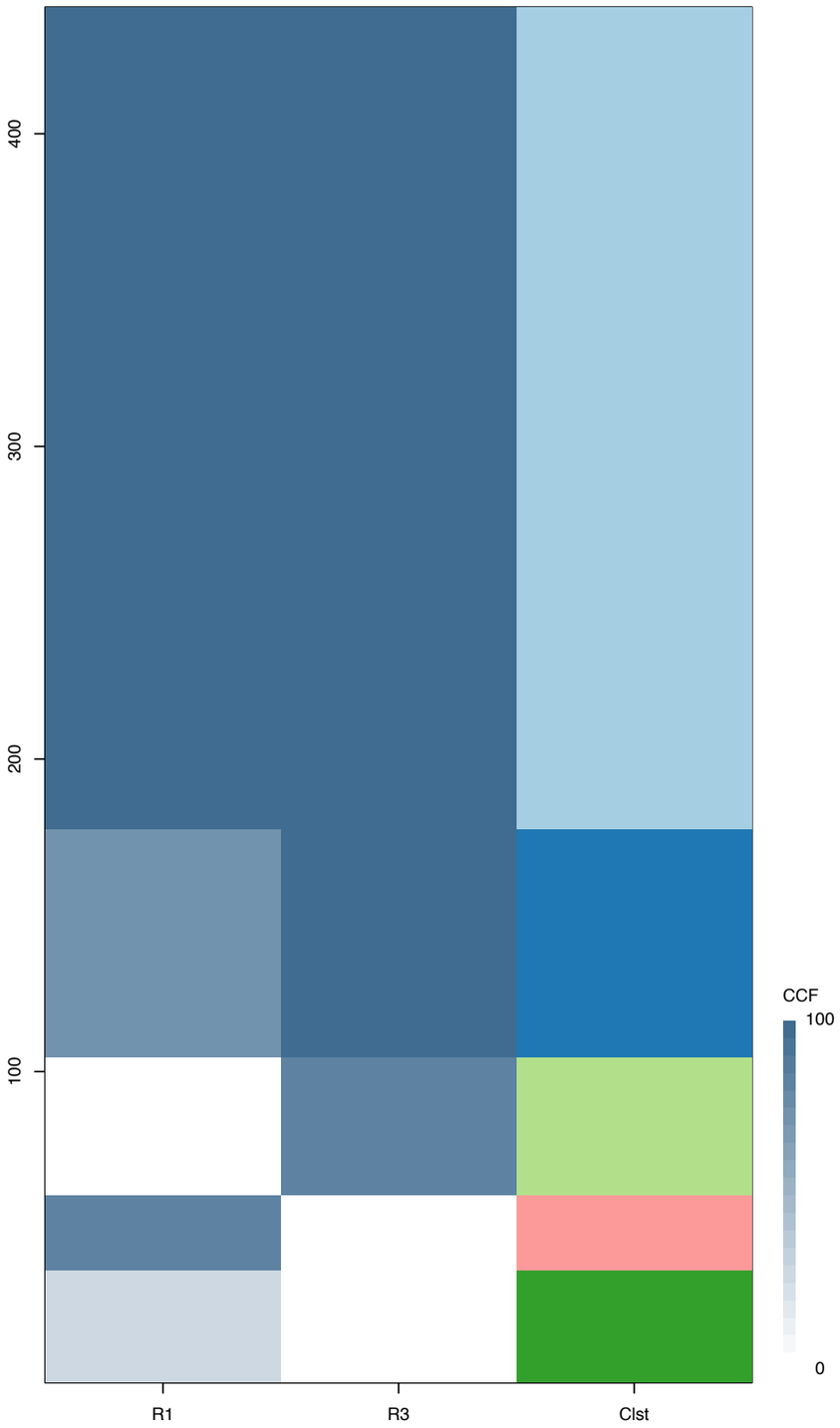
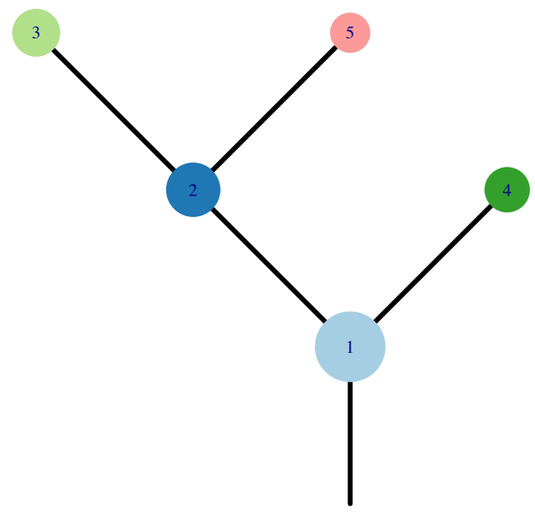


Fig.S12BI



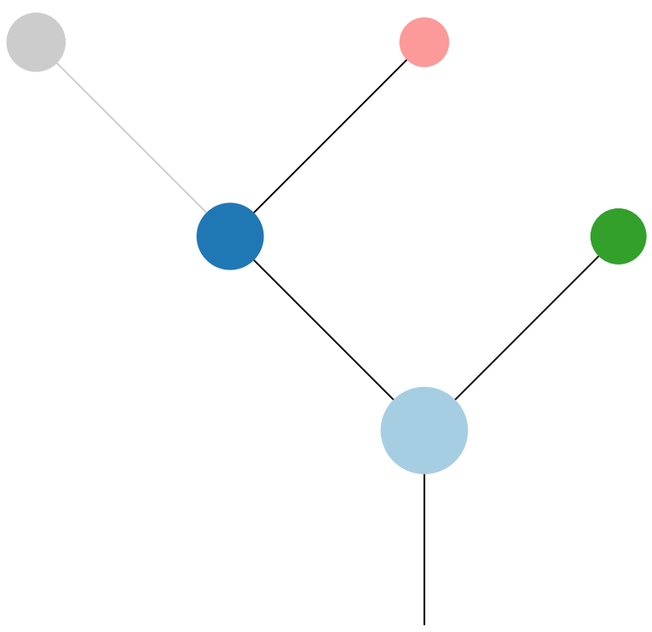
CRUK0067_B



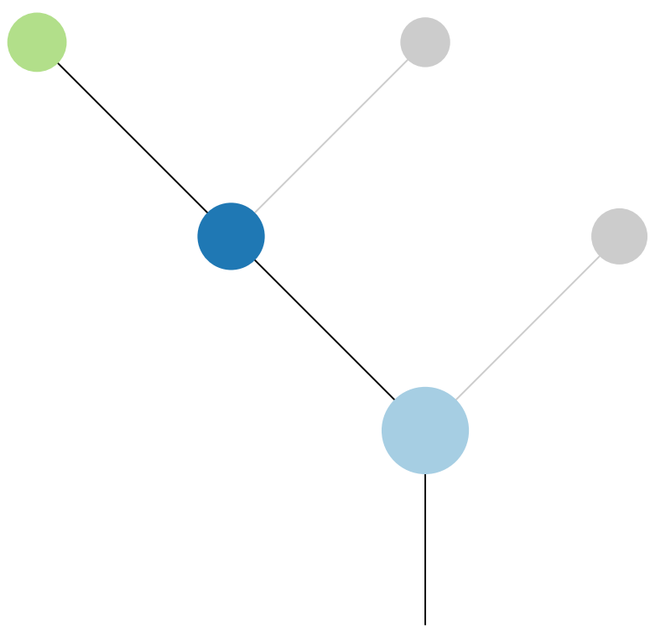
Histology:Squamous, Age:64, PackYears:50, Size:10
Stage:1a, Gender:Male, GD:Not GD, Recur:no

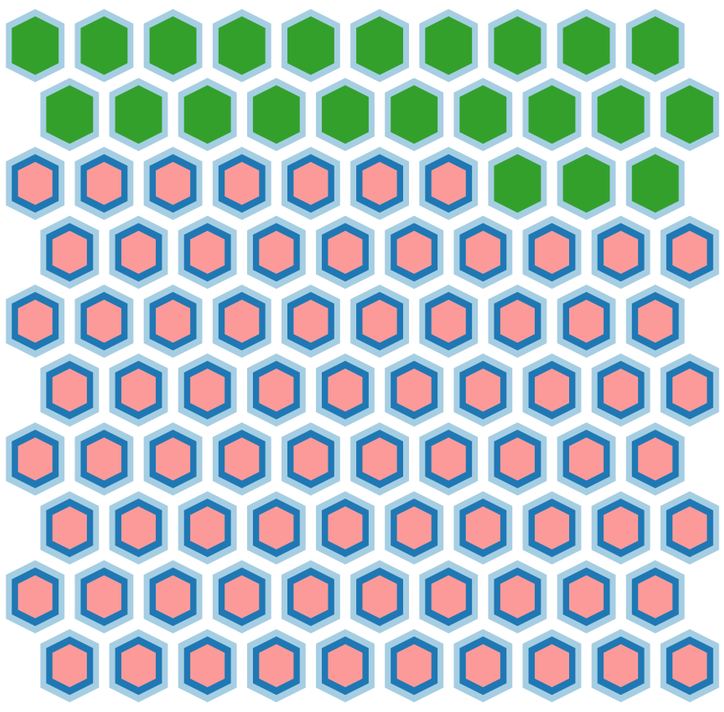
Gene	Cluster	Cytoband	Type
WWTR1	1	3q25.1	Amp
GMPS	1	3q25.31	Amp
MLF1	1	3q25.32	Amp
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
TFEB	1	6p21.1	Amp
CCND3	1	6p21.1	Amp
HSP90AB1	1	6p21.1	Amp
CDKN2A	1	9p21.3	Del
NOTCH1	1	9q34.3	SNV
TP53	1	17p13.1	SNV
NFE2L2	2	2q31.2	SNV
CUL3	?	2q36.2	SNV
KRAS	?	12p12.1	SNV

R1



R3





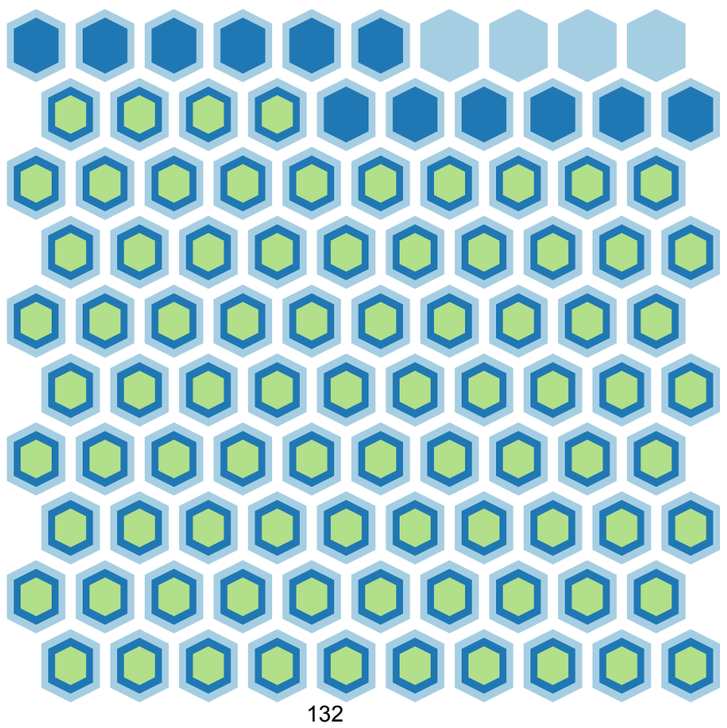
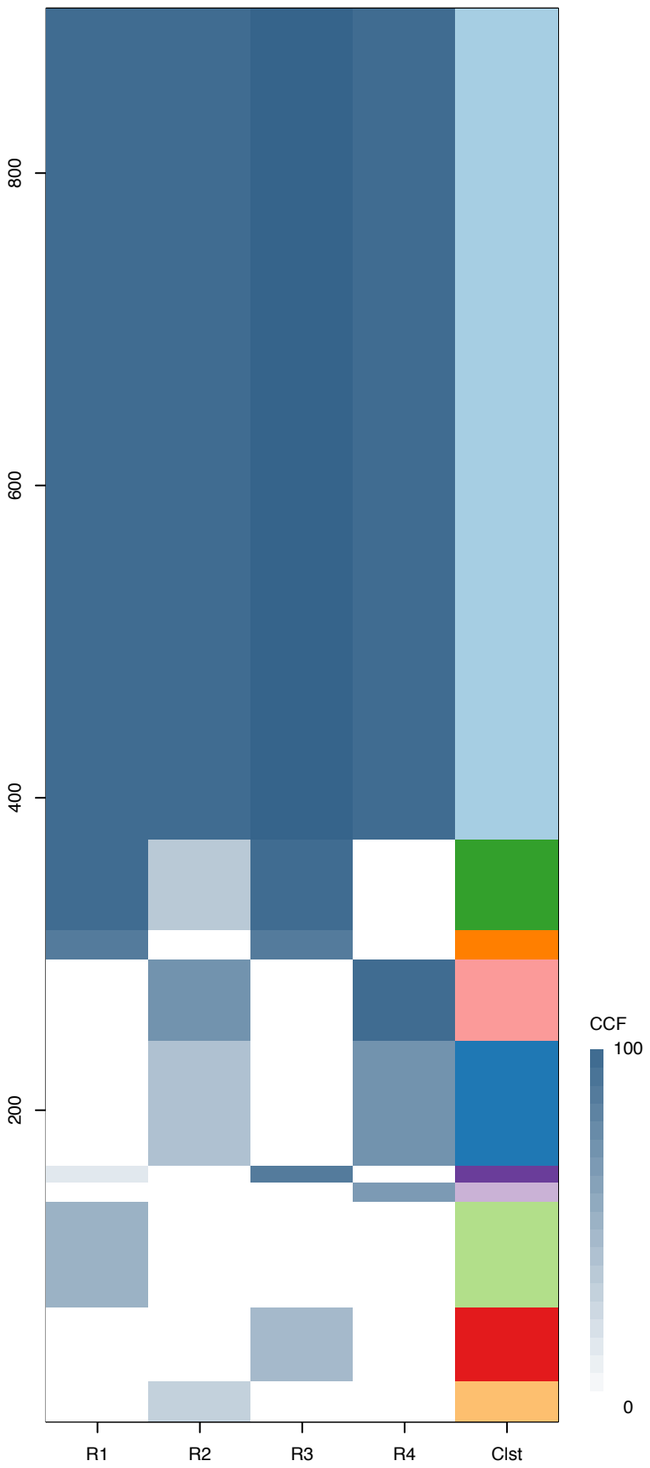
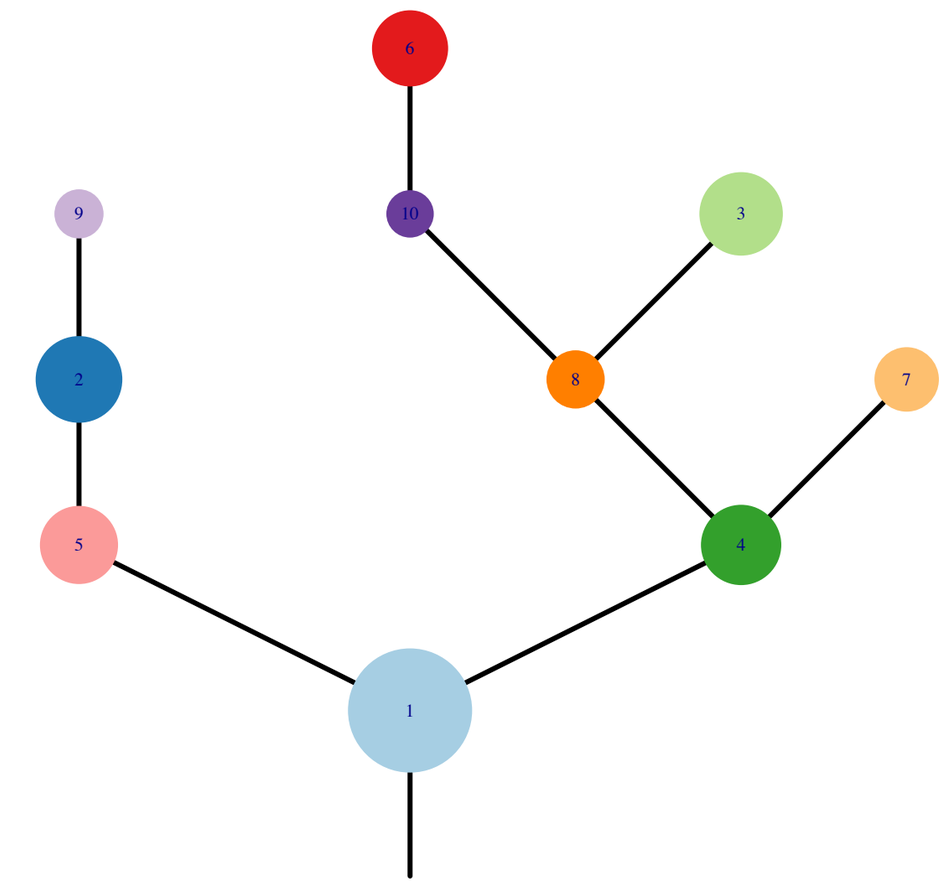


Fig.S12BJ



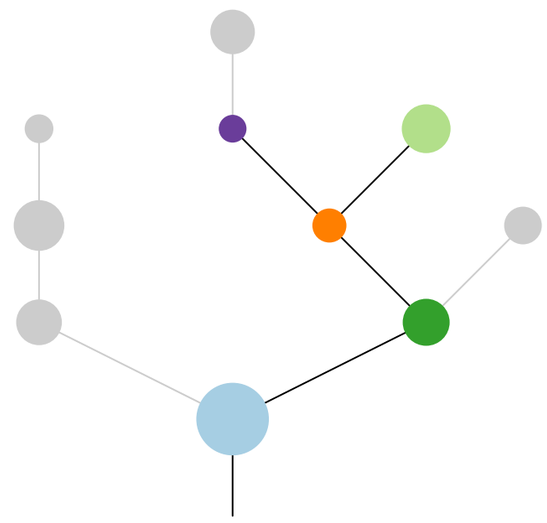
CRUK0068_A



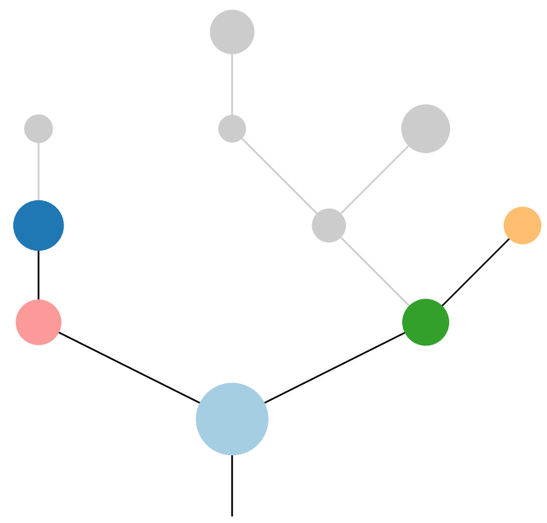
Histology:Squamous, Age:70, PackYears:106, Size:38
Stage:1b, Gender:Female, GD:Not GD, Recur:no

Gene	Cluster	Cytoband	Type
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
PTEN	1	10q23.31	SNV
KMT2D	1	12q13.12	SNV
TP53	1	17p13.1	SNV
MGA	3	15q15.1	SNV
SETD2	4	3p21.31	SNV
TERT	9	5p15.33	Amp
RASA1	?	5q14.3	SNV

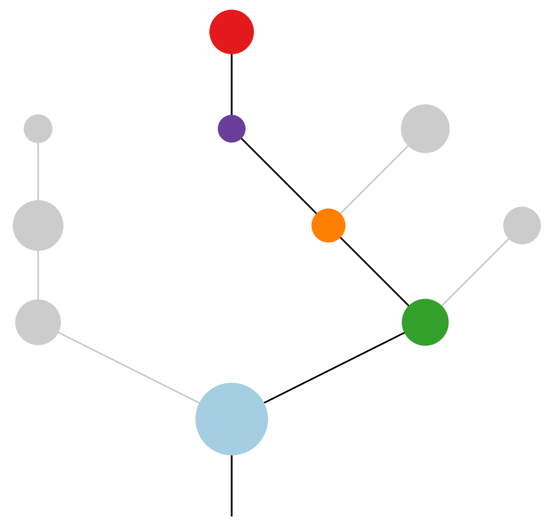
R1



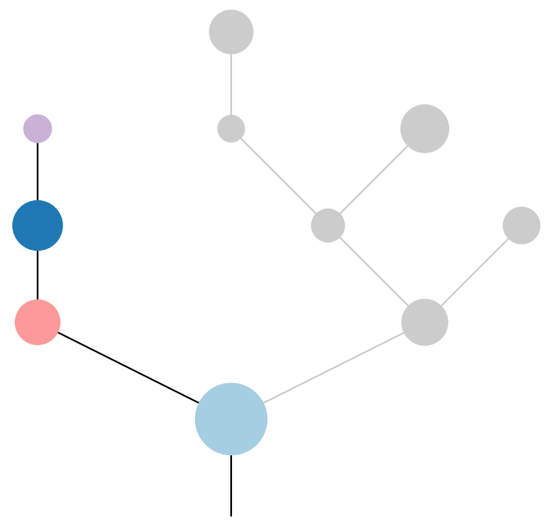
R2



R3



R4



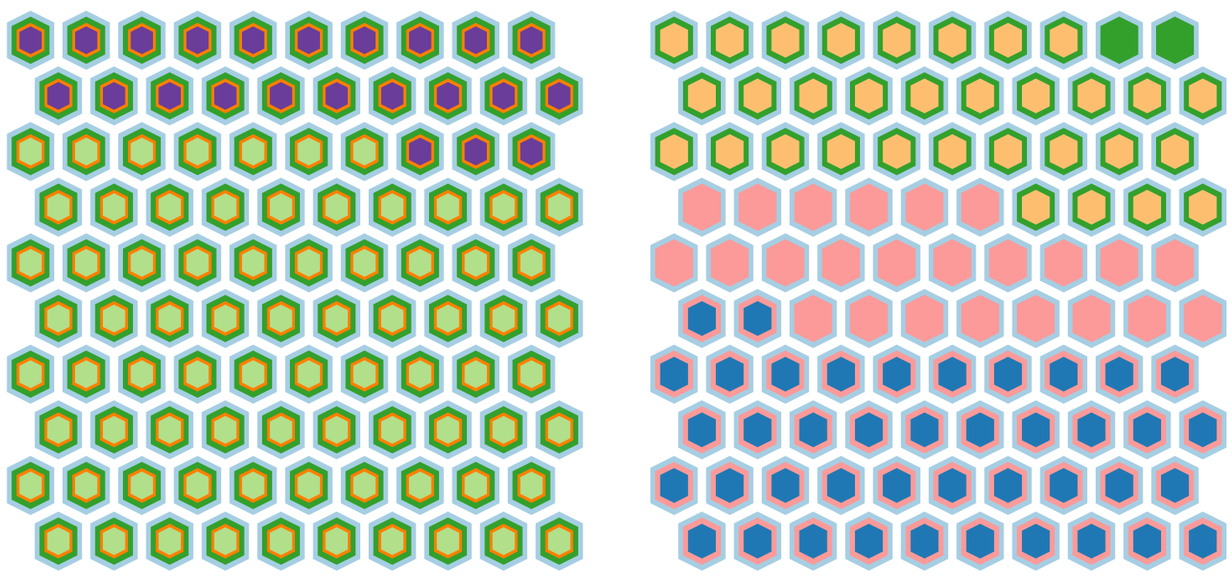
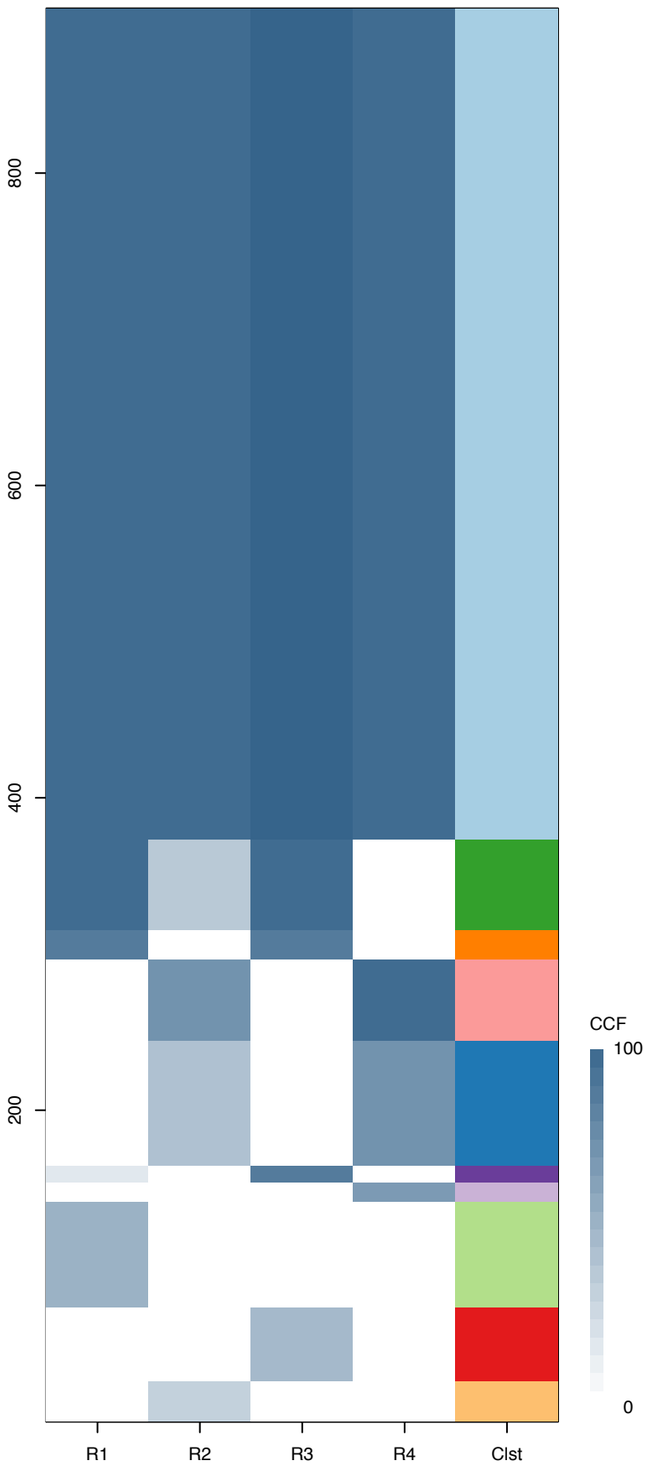
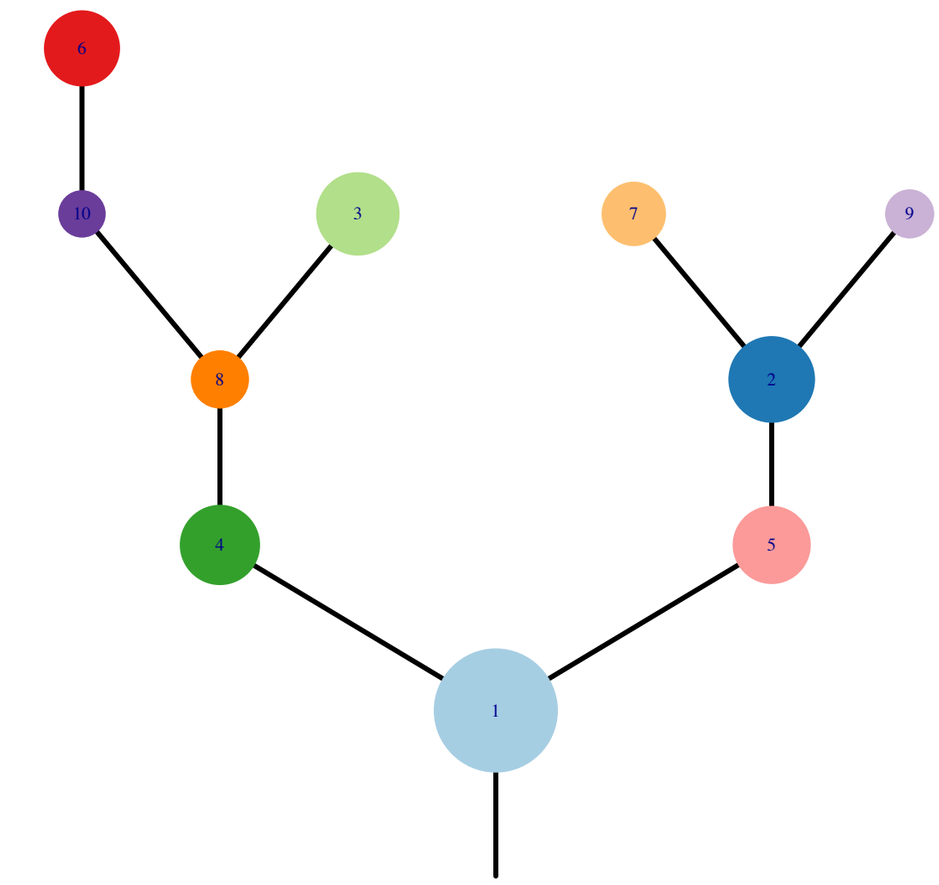


Fig.S12BJ



CRUK0068_B



Histology:Squamous, Age:70, PackYears:106, Size:38
Stage:1b, Gender:Female, GD:Not GD, Recur:no

Gene	Cluster	Cytoband	Type
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
PTEN	1	10q23.31	SNV
KMT2D	1	12q13.12	SNV
TP53	1	17p13.1	SNV
MGA	3	15q15.1	SNV
SETD2	4	3p21.31	SNV
TERT	9	5p15.33	Amp
RASA1	?	5q14.3	SNV

R1 **R2** **R3** **R4**

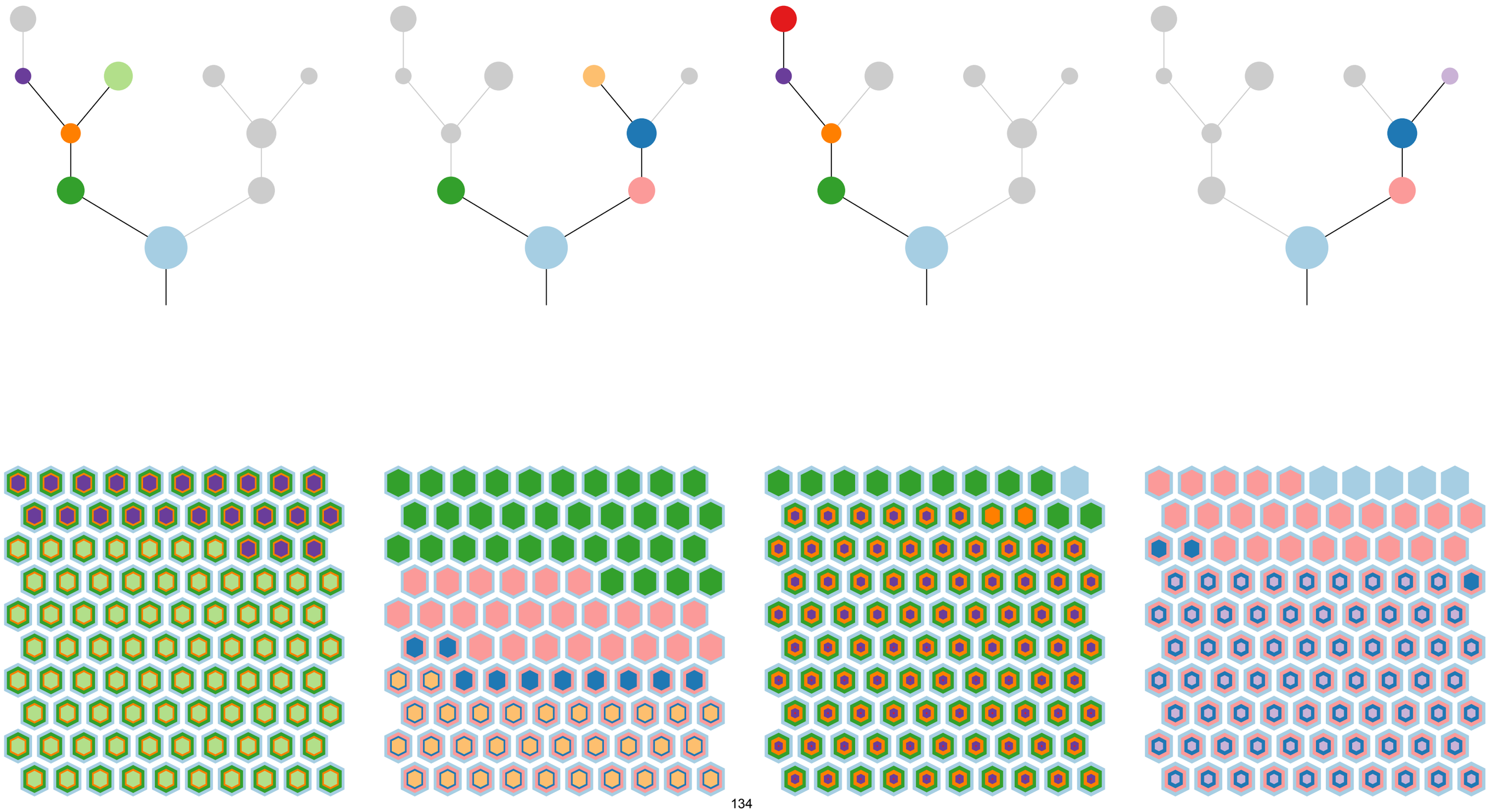
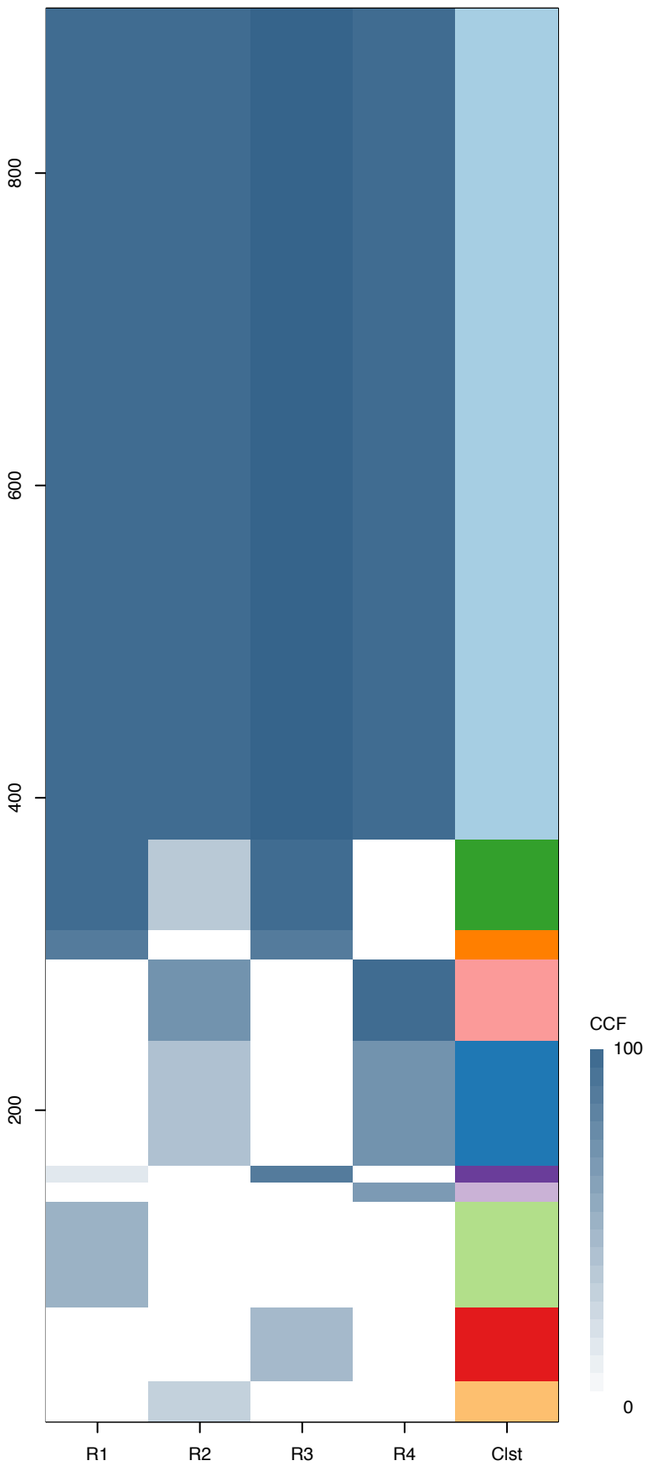
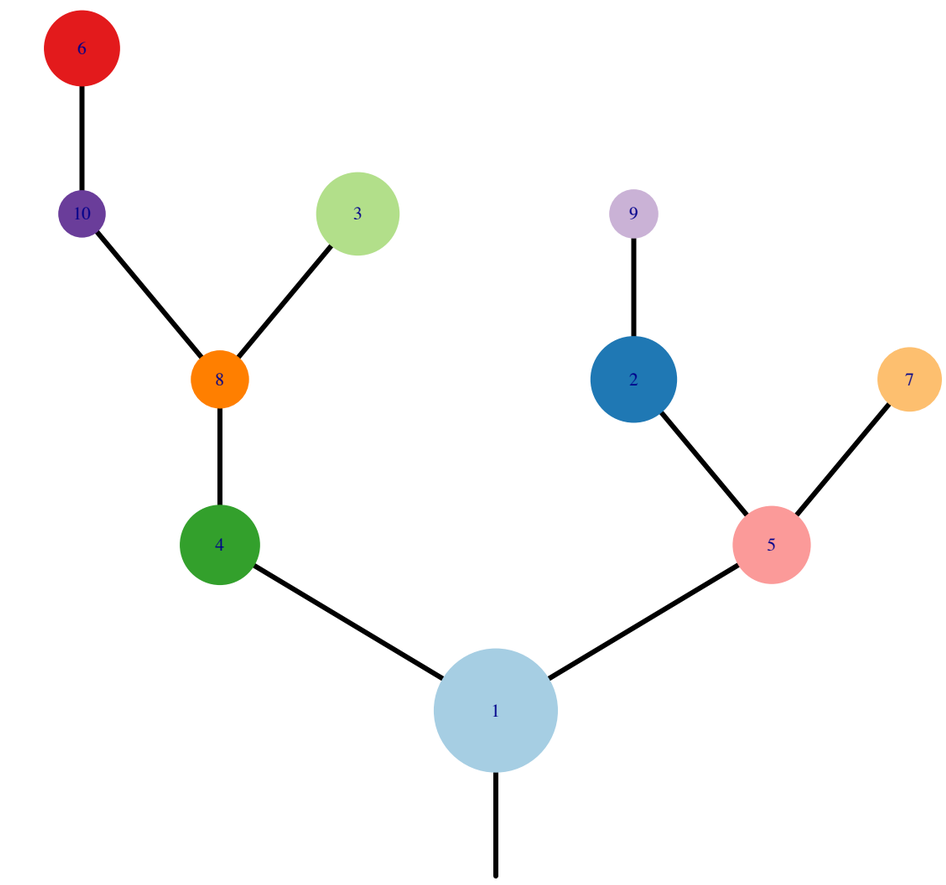


Fig.S12BJ



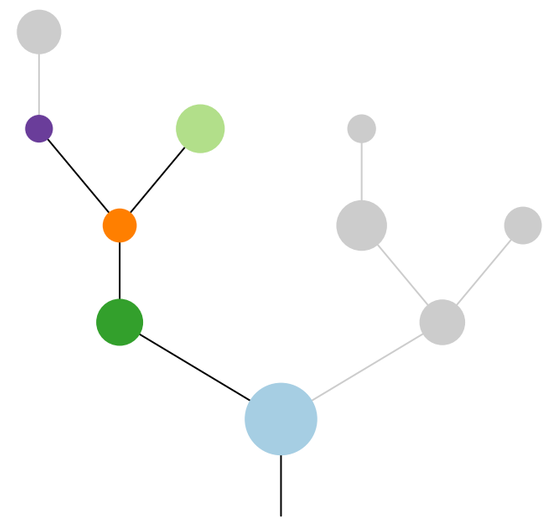
CRUK0068_C



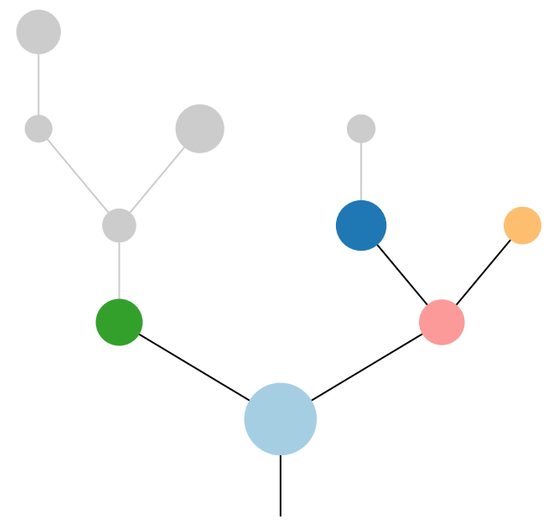
Histology:Squamous, Age:70, PackYears:106, Size:38
Stage:1b, Gender:Female, GD:Not GD, Recur:no

Gene	Cluster	Cytoband	Type
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
PTEN	1	10q23.31	SNV
KMT2D	1	12q13.12	SNV
TP53	1	17p13.1	SNV
MGA	3	15q15.1	SNV
SETD2	4	3p21.31	SNV
TERT	9	5p15.33	Amp
RASA1	?	5q14.3	SNV

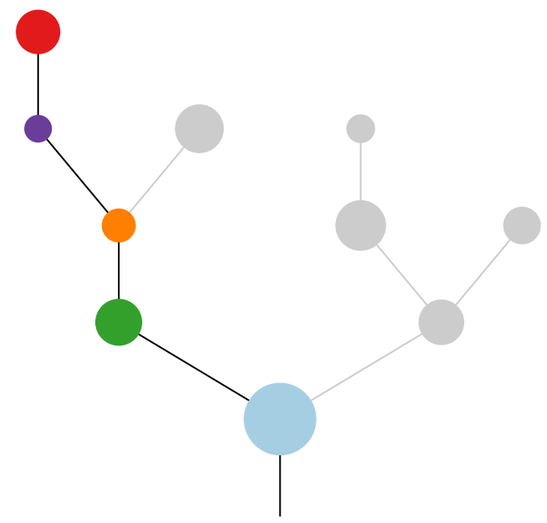
R1



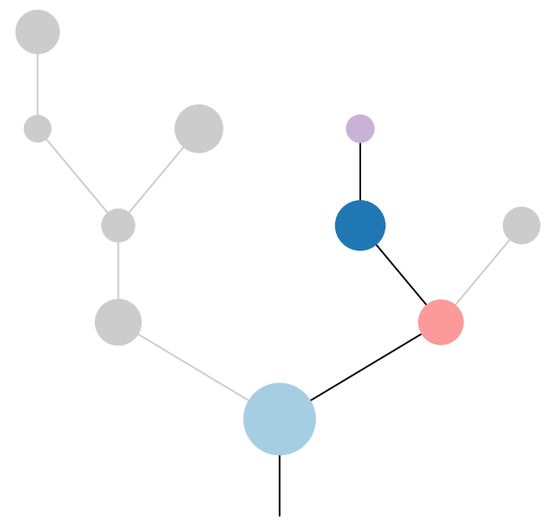
R2



R3



R4



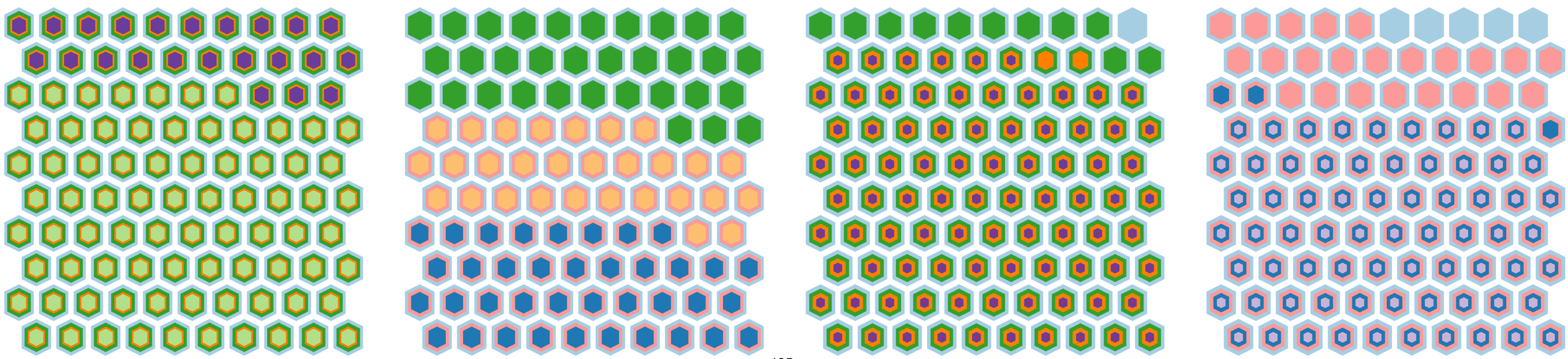
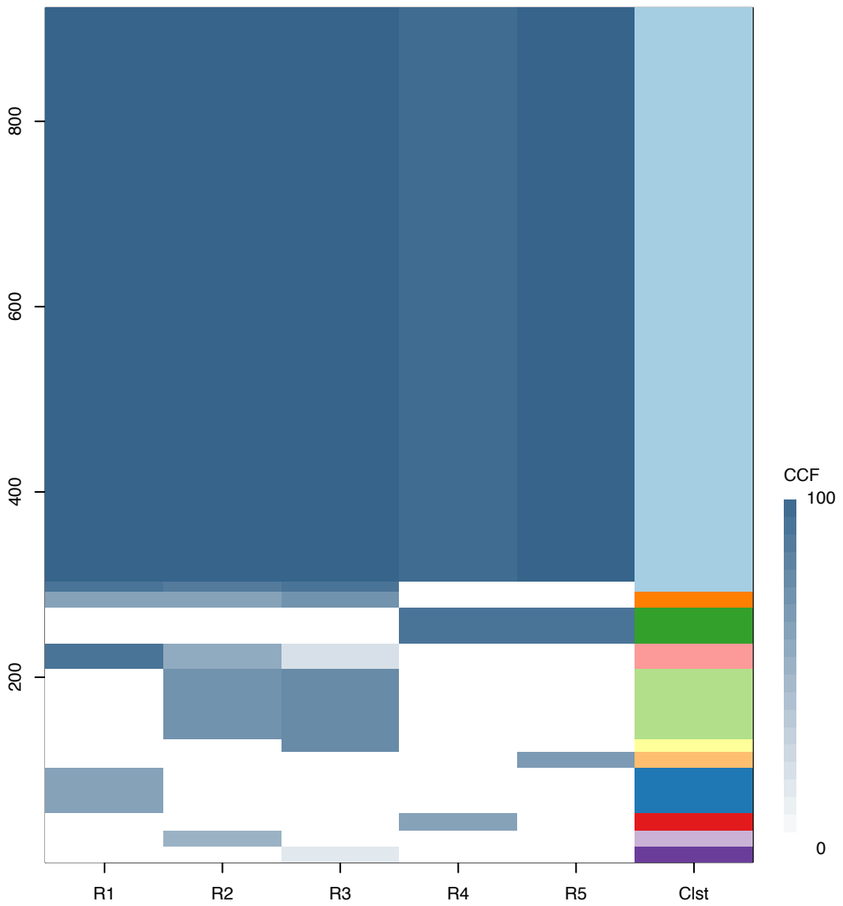
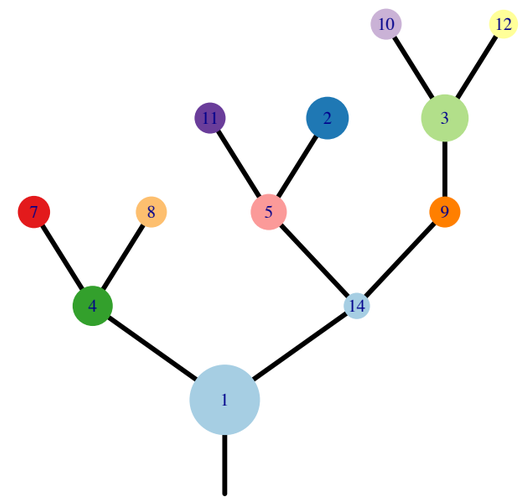


Fig.S12BK



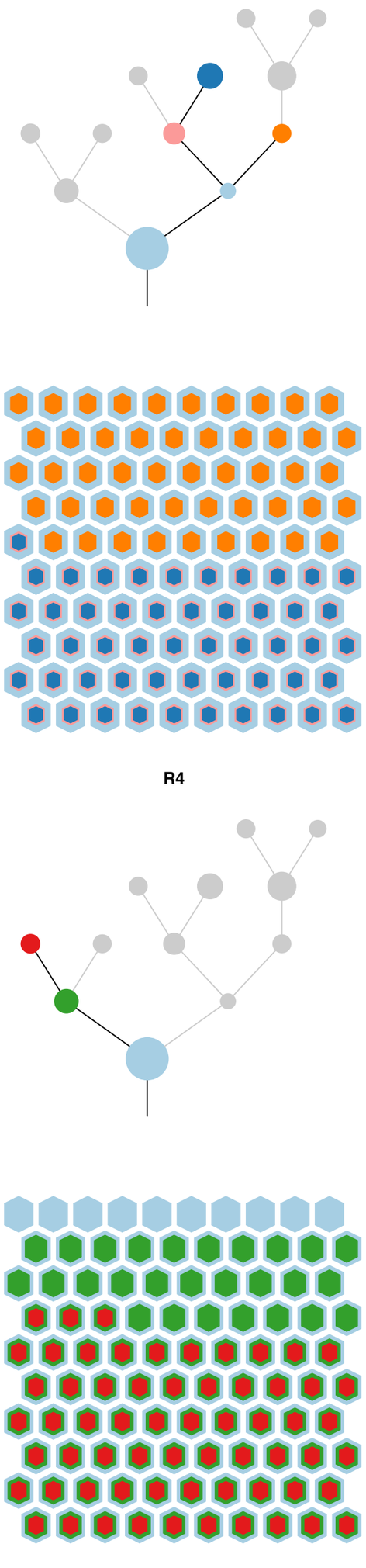
CRUK0069



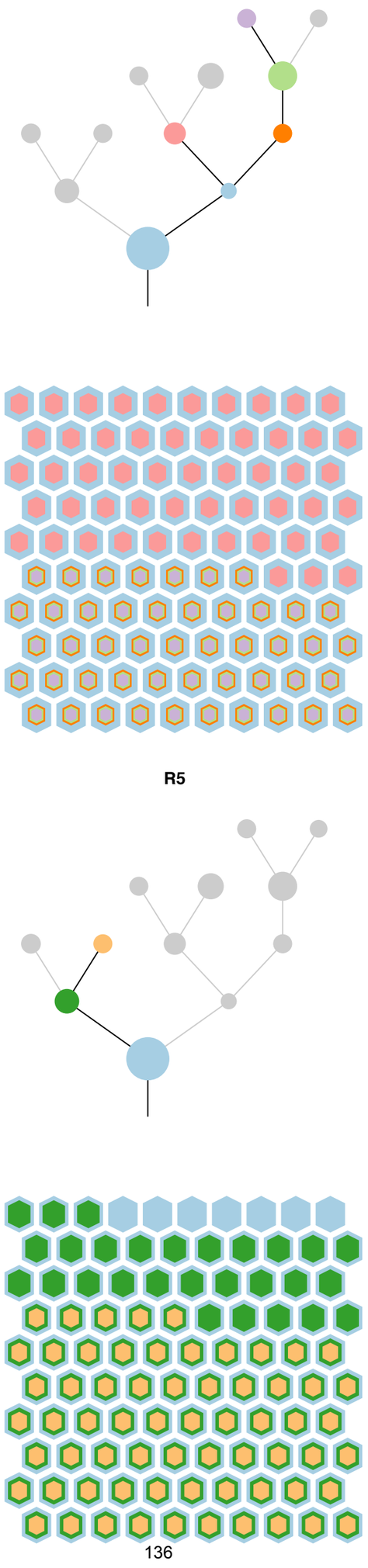
Histology:Squamous, Age:74, PackYears:44, Size:50
Stage:1b, Gender:Female, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
CHIC2	1	4q12	Amp
PDGFRA	1	4q12	Amp
KIT	1	4q12	Amp
KDR	1	4q12	Amp
FAT1	1	4q35.2	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
ERBB3	1	12q13.2	Amp
NACA	1	12q13.3	Amp
NAB2	1	12q13.3	Amp
STAT6	1	12q13.3	Amp
DDIT3	1	12q13.3	Amp
CDK4	1	12q14.1	Amp
TP53	1	17p13.1	SNV
CEP89	1	19q13.11	Amp
CEBPA	1	19q13.11	Amp
LSM14A	1	19q13.11	Amp
AKT2	1	19q13.2	Amp
IL7R	2	5p13.2	Amp
JAK3	2	19p13.11	Amp
ATP5B	3	12q13.3	Amp
FIP1L1	4	4q12	Amp
NFATC2	4	20q13.2	Amp
AURKA	4	20q13.2	Amp
GNAS	4	20q13.32	Amp
SS18L1	4	20q13.33	Amp
LIFR	11	5p13.1	Amp
KRAS	?	12p12.1	SNV
AMER1	?		SNV

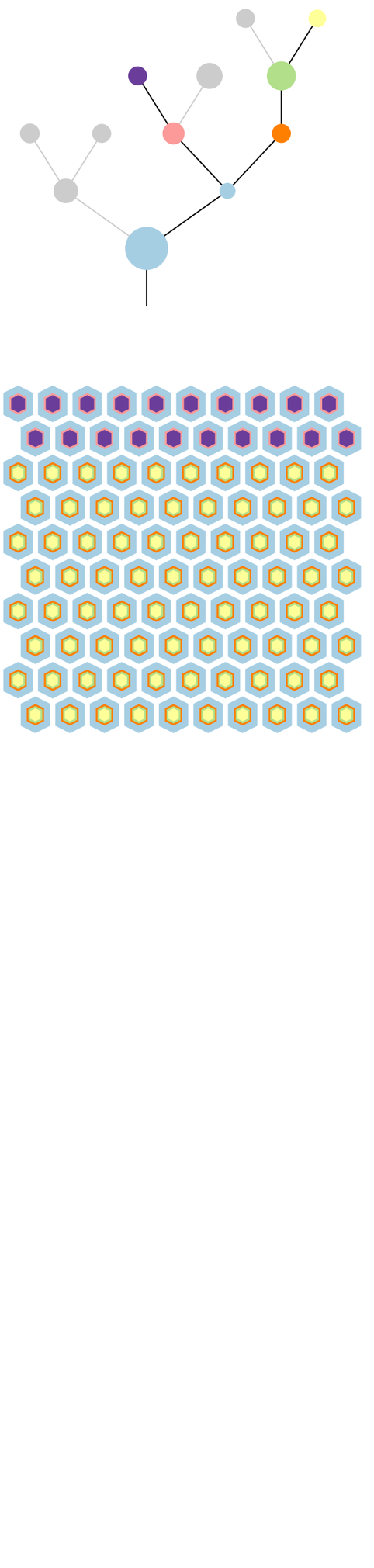
R1



R2



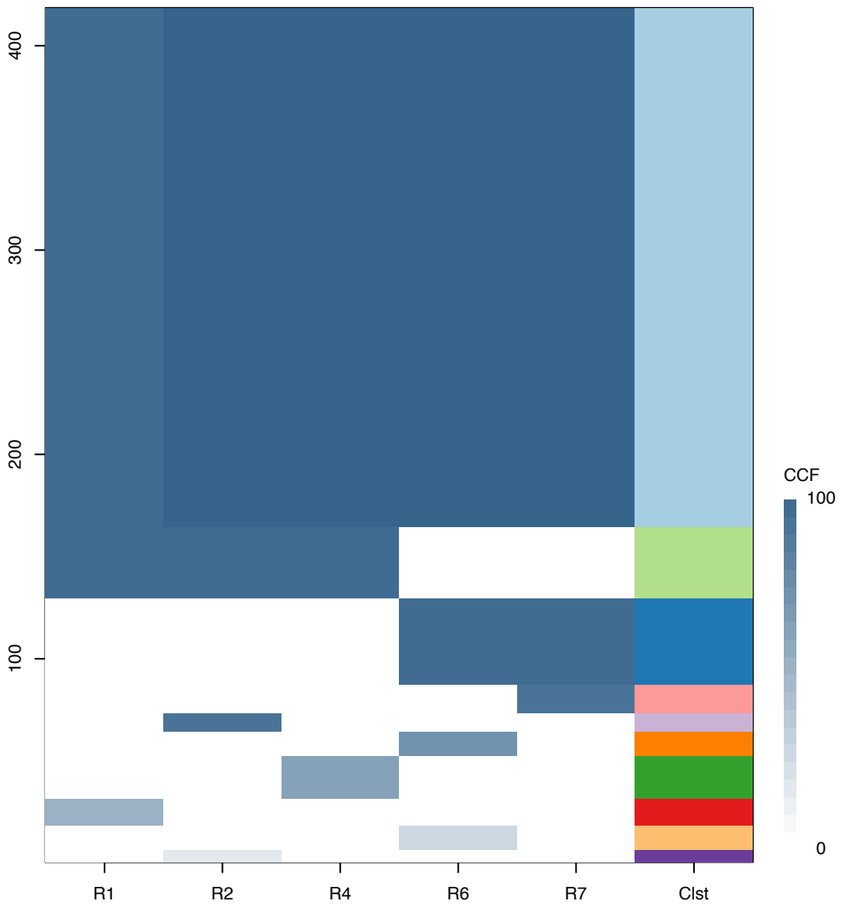
R3



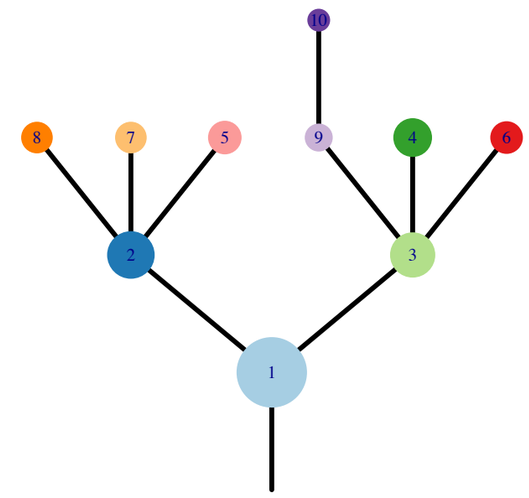
R4

R5

Fig.S12BL



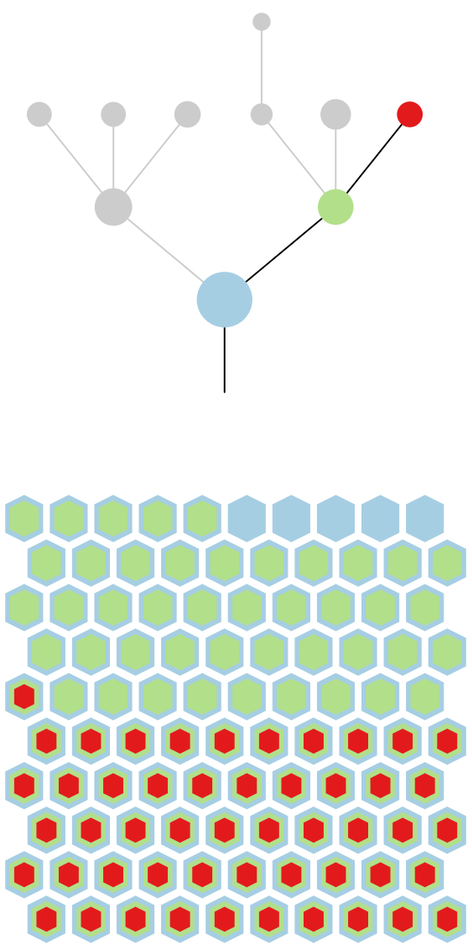
CRUK0070_A



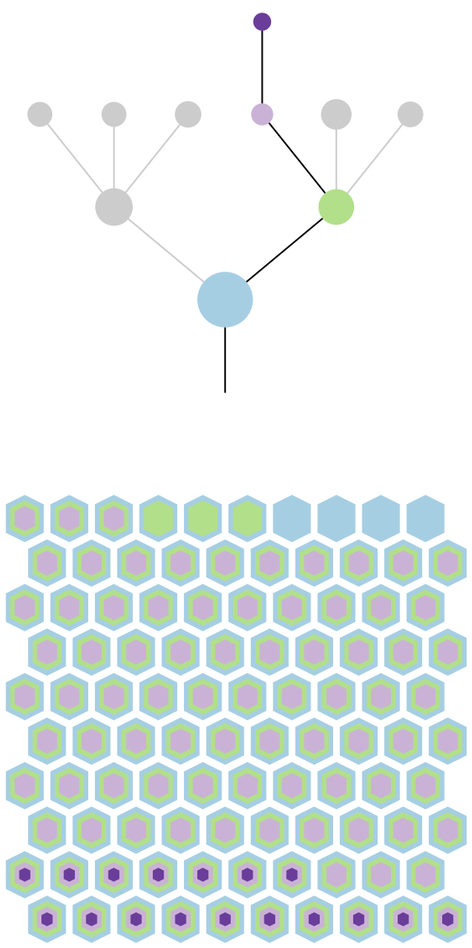
Histology:Squamous, Age:58, PackYears:41, Size:45
Stage:2a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
COL5A2	1	2q32.2	Amp
SF3B1	1	2q33.1	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TP53	1	17p13.1	SNV
DNM2	1	19p13.2	SNV
CBLB	2	3q13.11	SNV
AKT2	2	19q13.2	Amp
NFE2L2	3	2q31.2	Amp
KDM6A	?		SNV

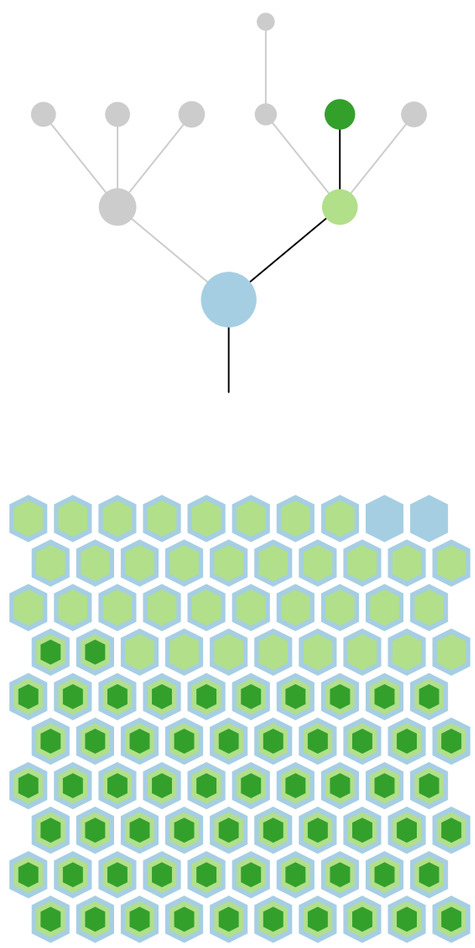
R1



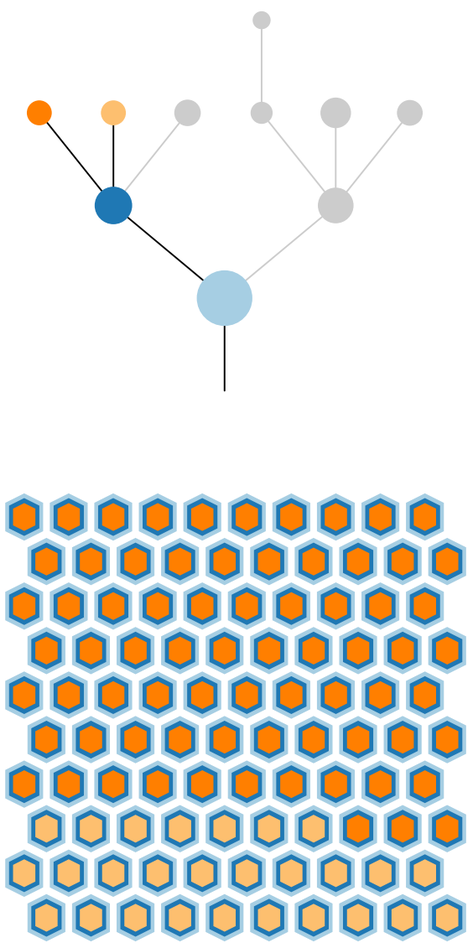
R2



R4



R6



R7

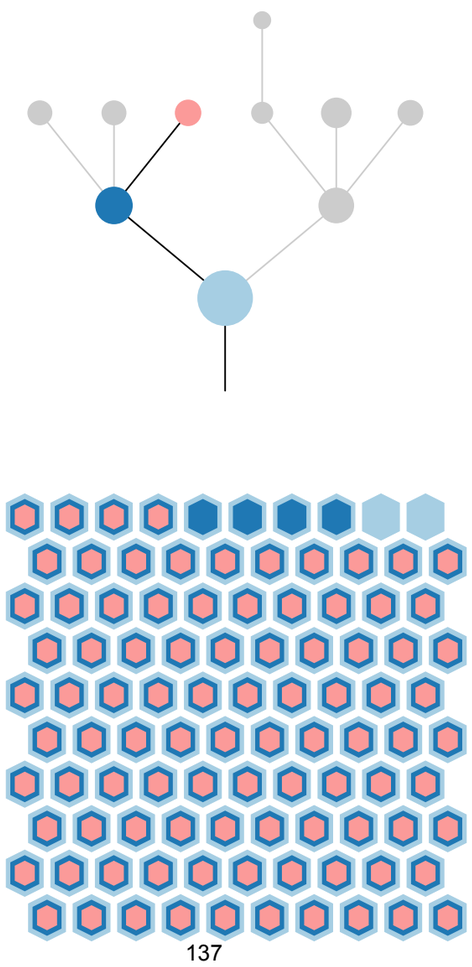
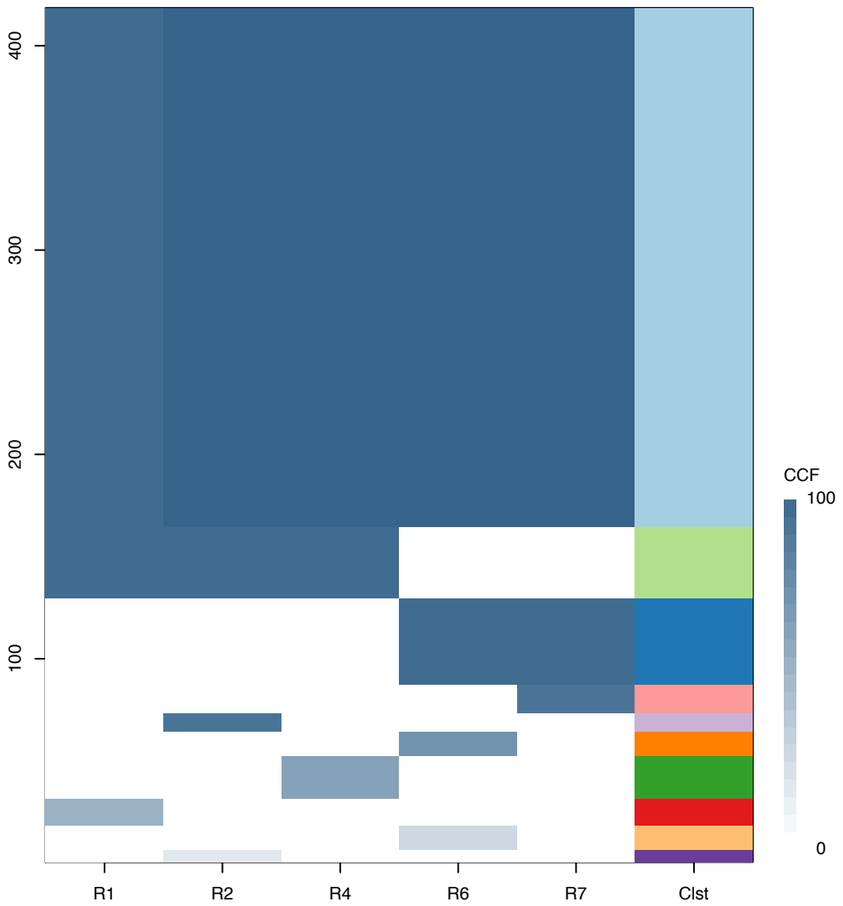
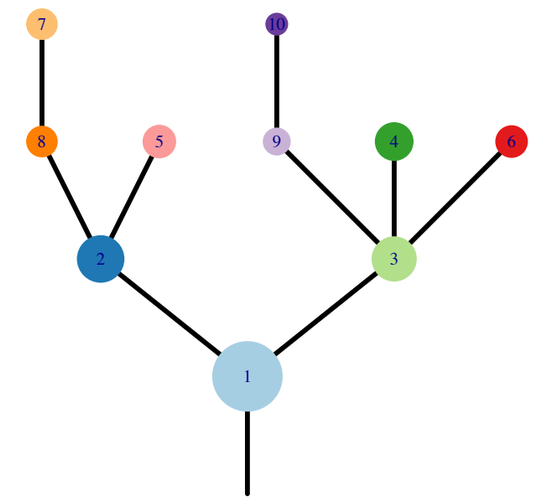


Fig.S12BL



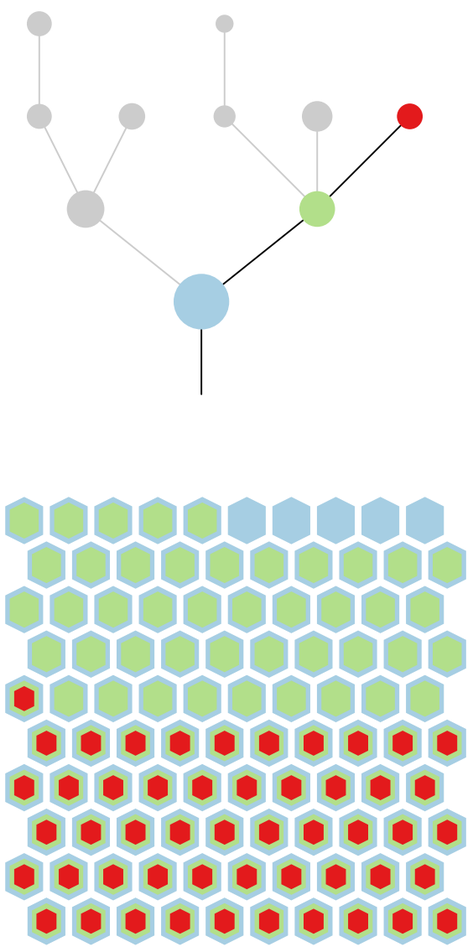
CRUK0070_B



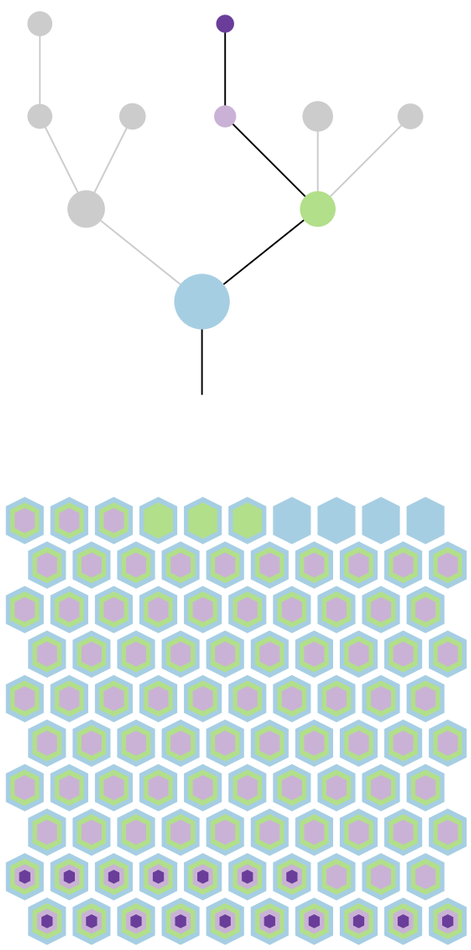
Histology:Squamous, Age:58, PackYears:41, Size:45
Stage:2a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
COL5A2	1	2q32.2	Amp
SF3B1	1	2q33.1	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TP53	1	17p13.1	SNV
DNM2	1	19p13.2	SNV
CBLB	2	3q13.11	SNV
AKT2	2	19q13.2	Amp
NFE2L2	3	2q31.2	Amp
KDM6A	?		SNV

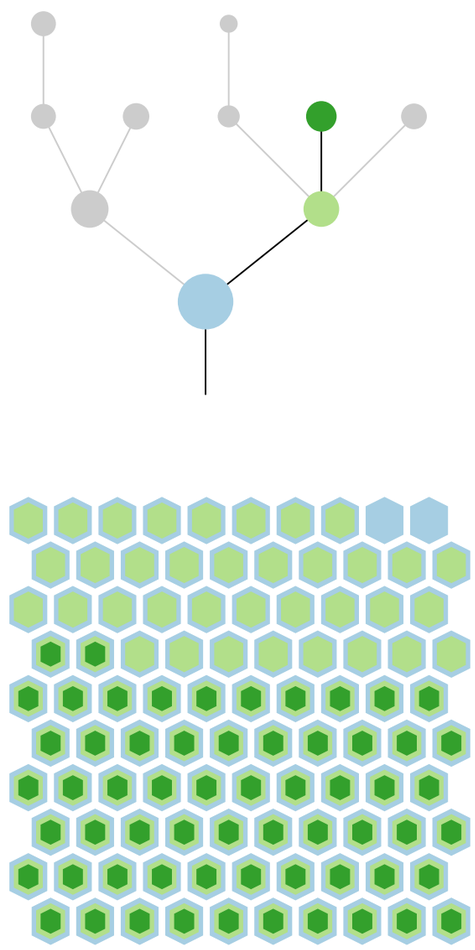
R1



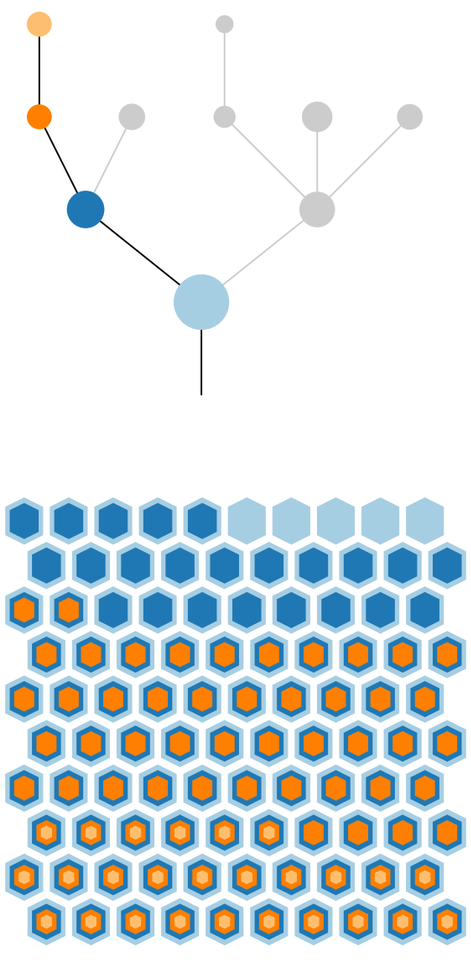
R2



R4



R6



R7

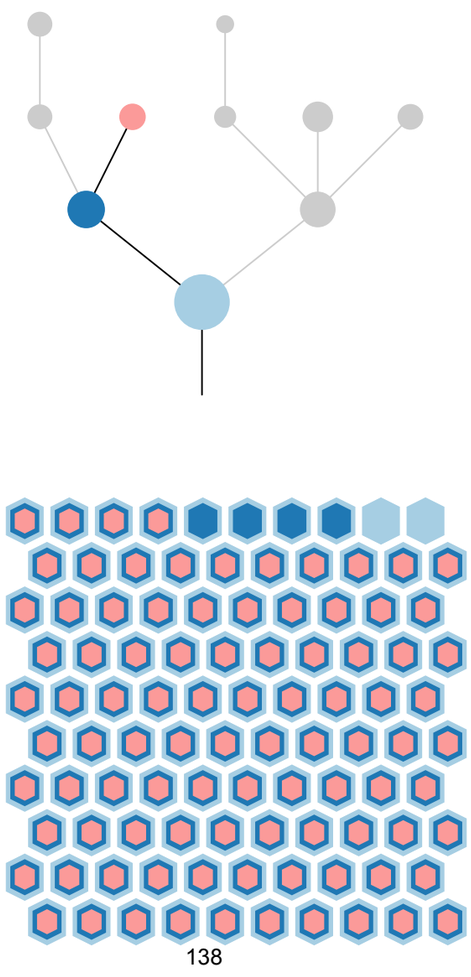
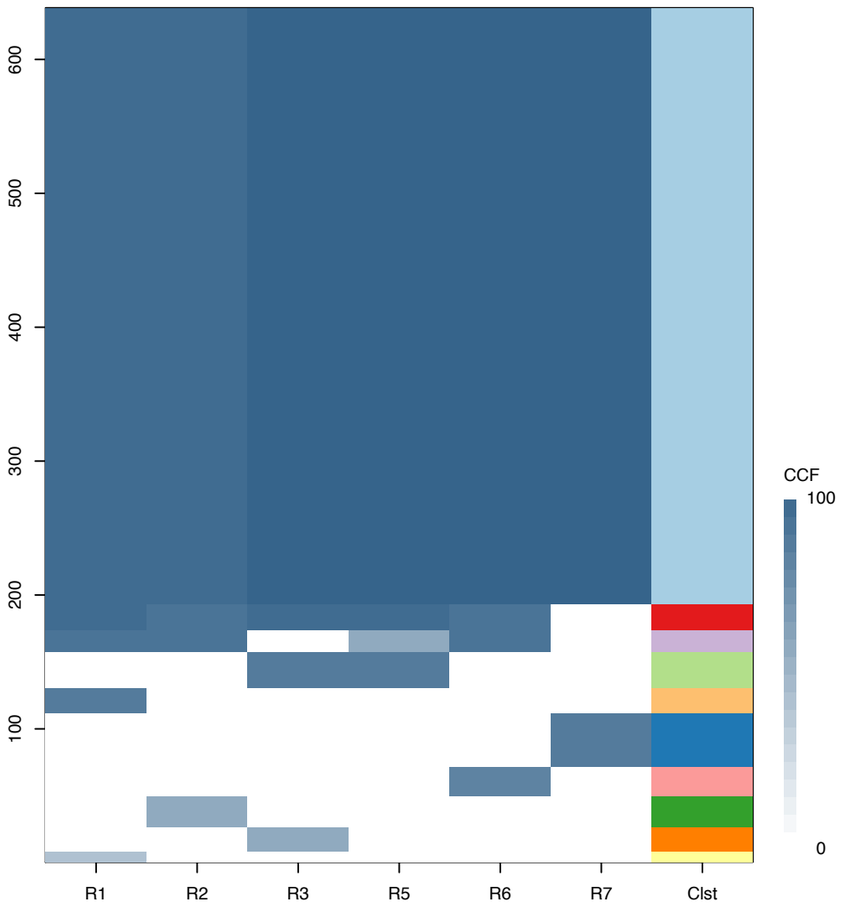
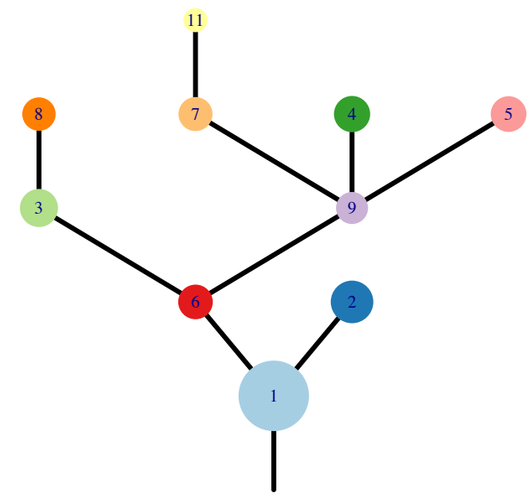


Fig.S12BM



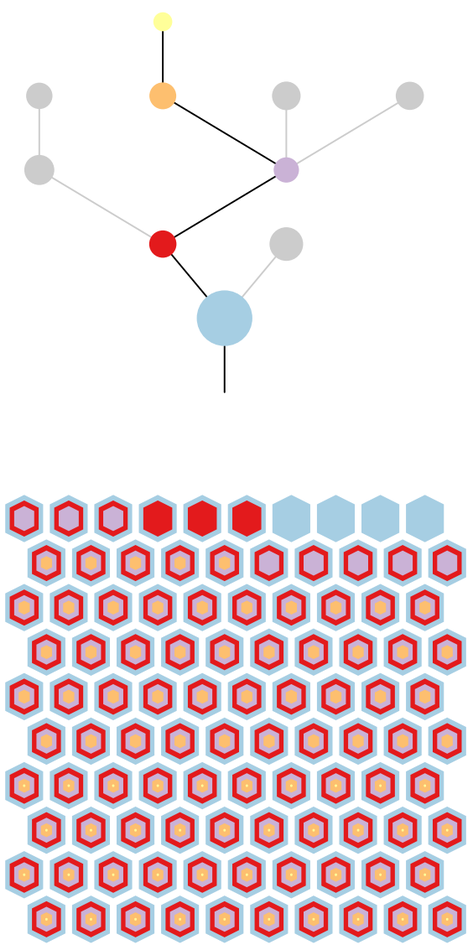
CRUK0071



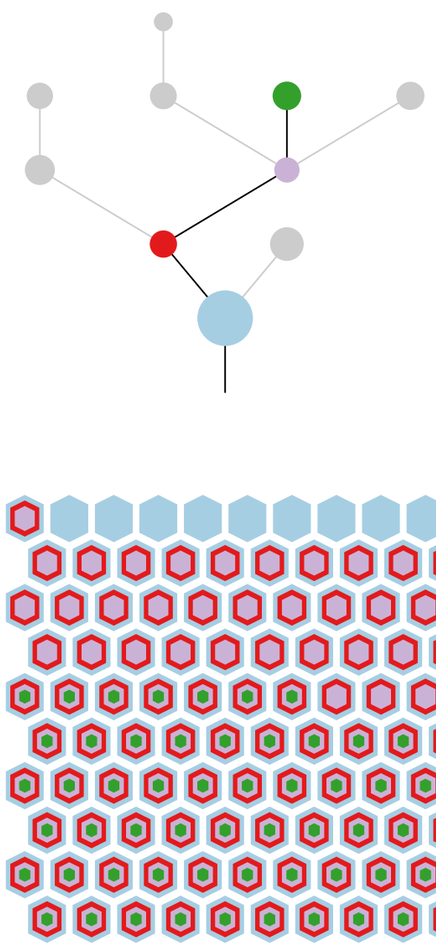
Histology:Squamous, Age:68, PackYears:80, Size:55
Stage:2a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
OMA1	1	1p32.2	Amp
JUN	1	1p32.1	Amp
JAK1	1	1p31.3	Amp
FOXL2	1	3q22.3	Amp
WWTR1	1	3q25.1	Amp
GMPS	1	3q25.31	Amp
MLF1	1	3q25.32	Amp
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
CMTR2	1	16q22.2	SNV
UBR5	4	8q22.3	SNV
CMTR2	?	16q22.2	SNV

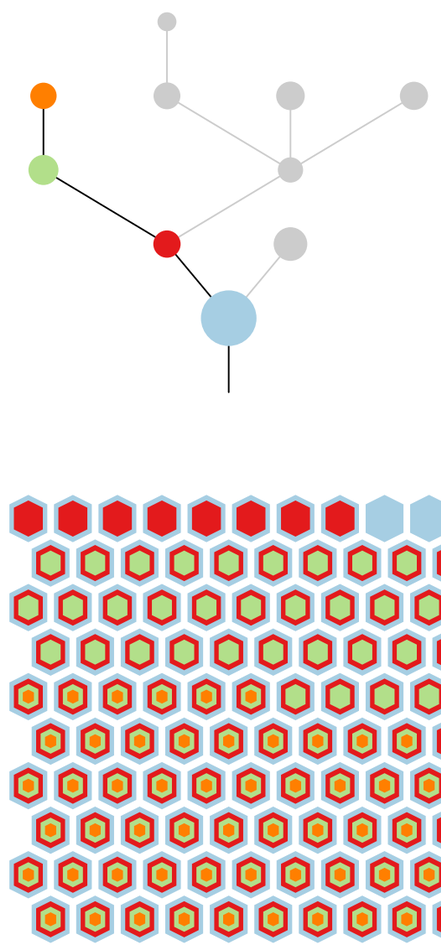
R1



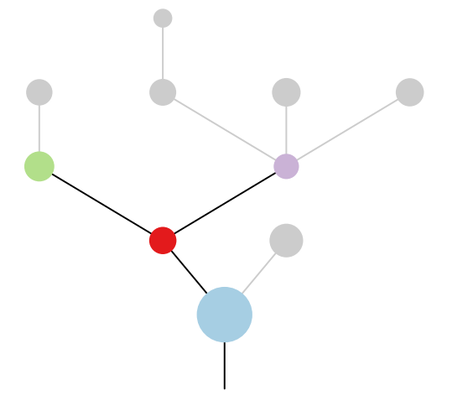
R2



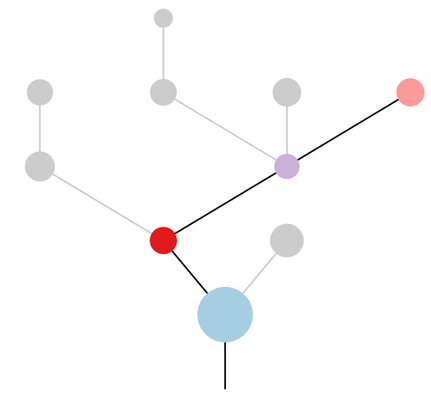
R3



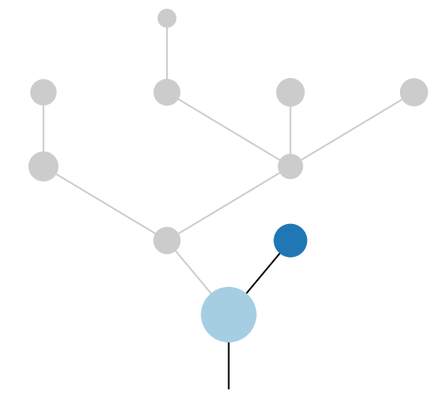
R5

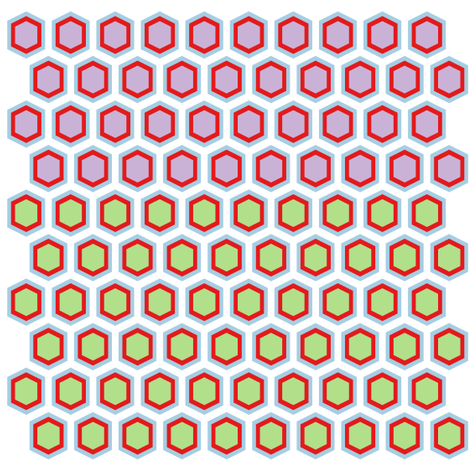


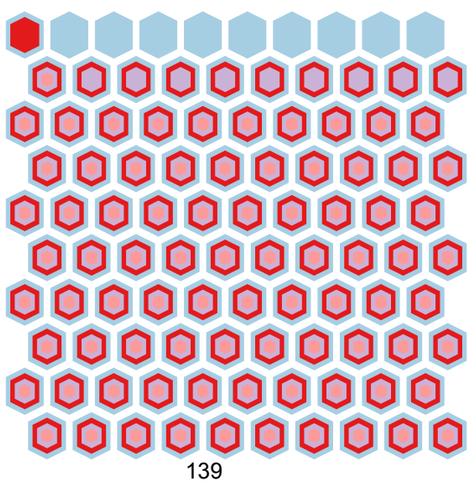
R6



R7







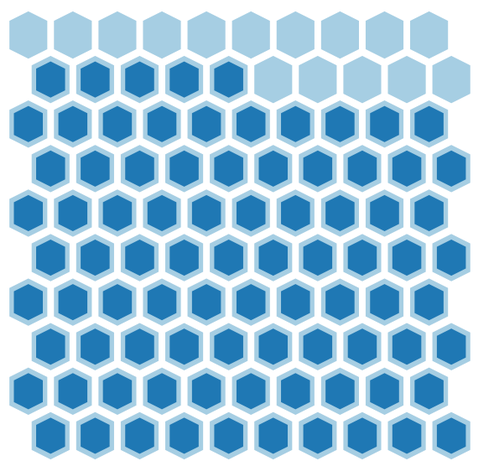
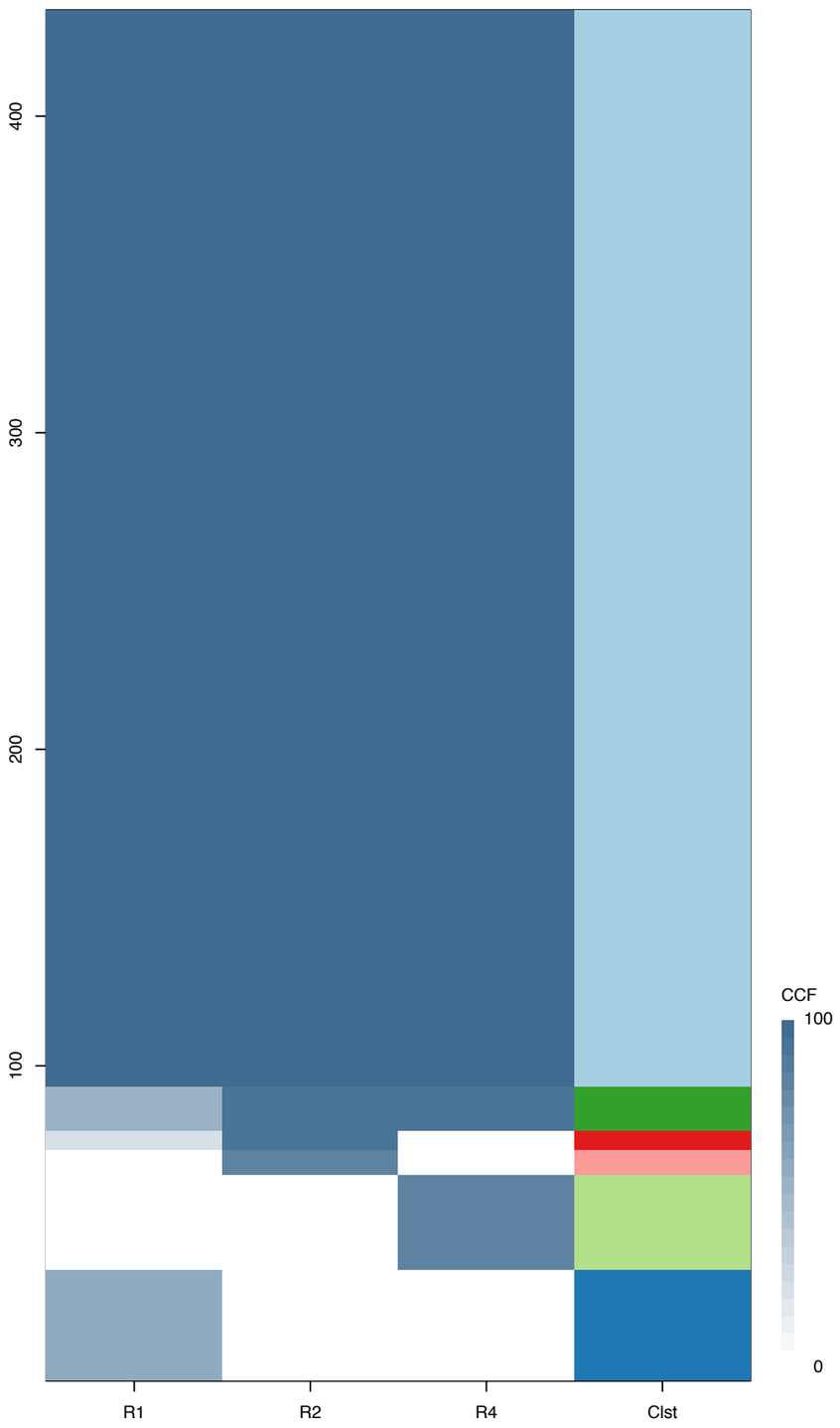
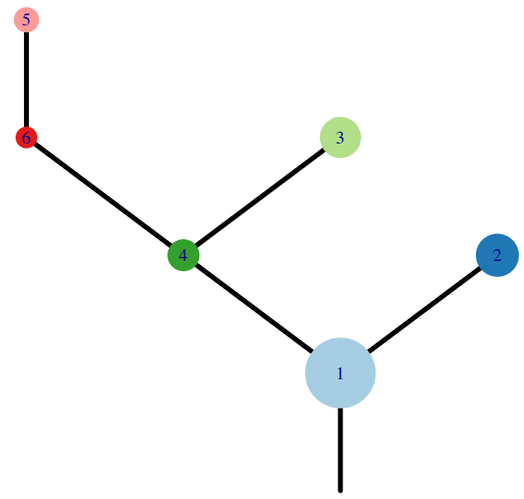


Fig.S12BN



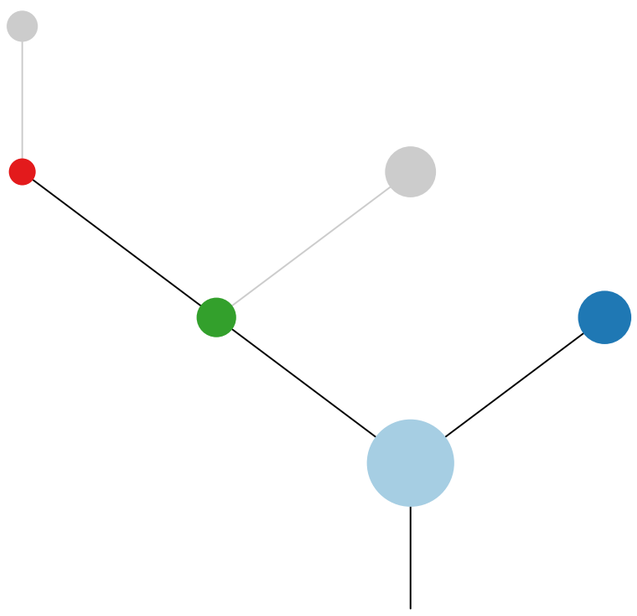
CRUK0072



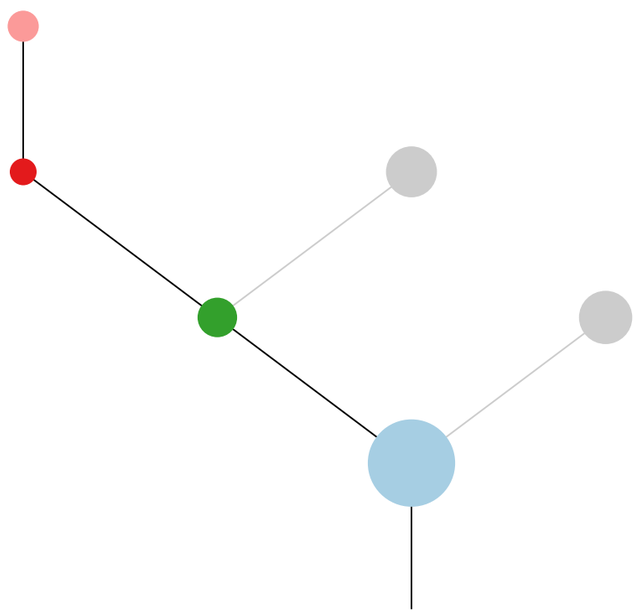
Histology:Squamous, Age:68, PackYears:25.5, Size:35
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
NFE2L2	1	2q31.2	SNV
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
EGFR	1	7p11.2	Amp
MYC	1	8q24.21	Amp
NDRG1	1	8q24.22	Amp
TP53	1	17p13.1	SNV
FCGR2B	2	1q23.3	Amp
C2orf44	2	2p23.3	Amp
NCOA1	2	2p23.3	Amp
BCL11A	2	2p16.1	Amp
REL	2	2p16.1	Amp
XPO1	2	2p15	Amp
BCL11A	6	2p16.1	Amp
REL	6	2p16.1	Amp
XPO1	6	2p15	Amp
WRN	?	8p12	SNV

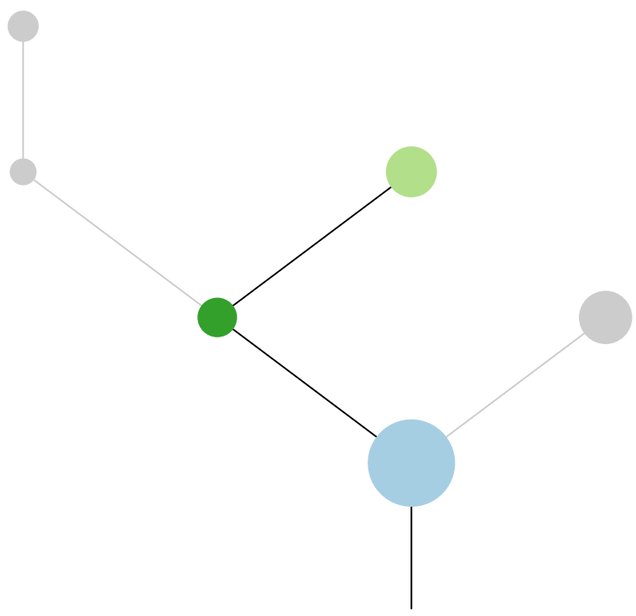
R1

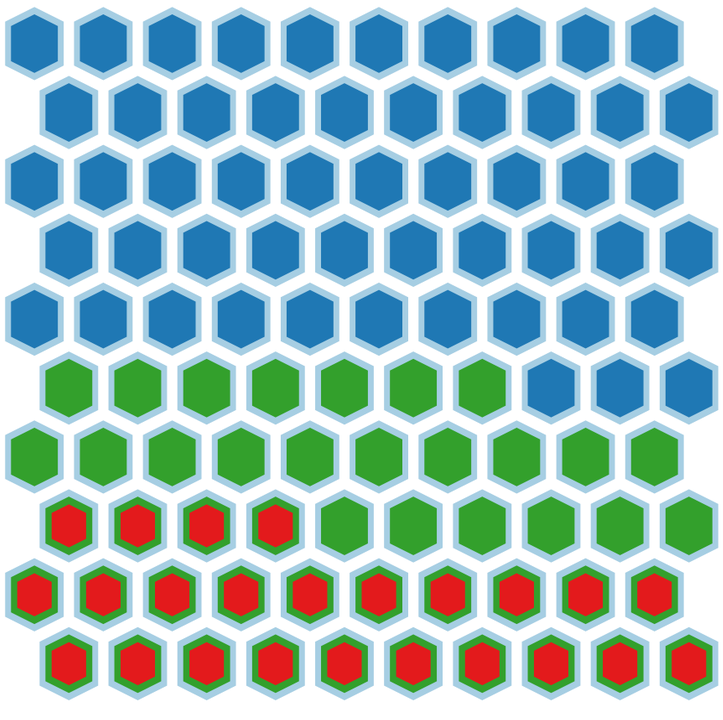


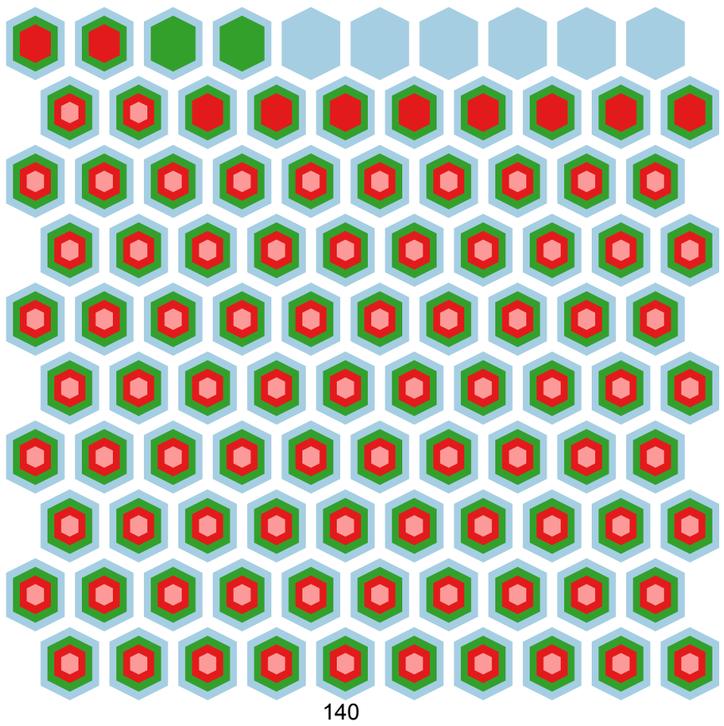
R2



R4







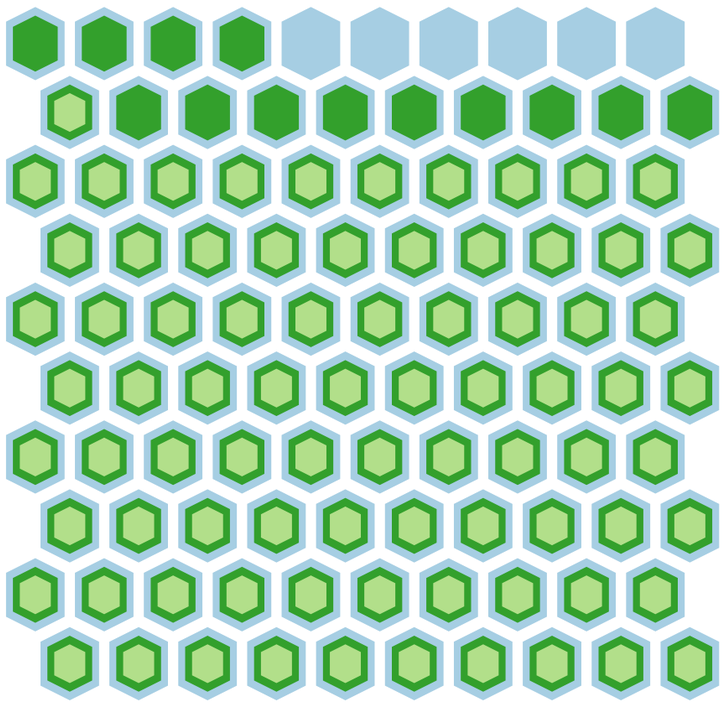
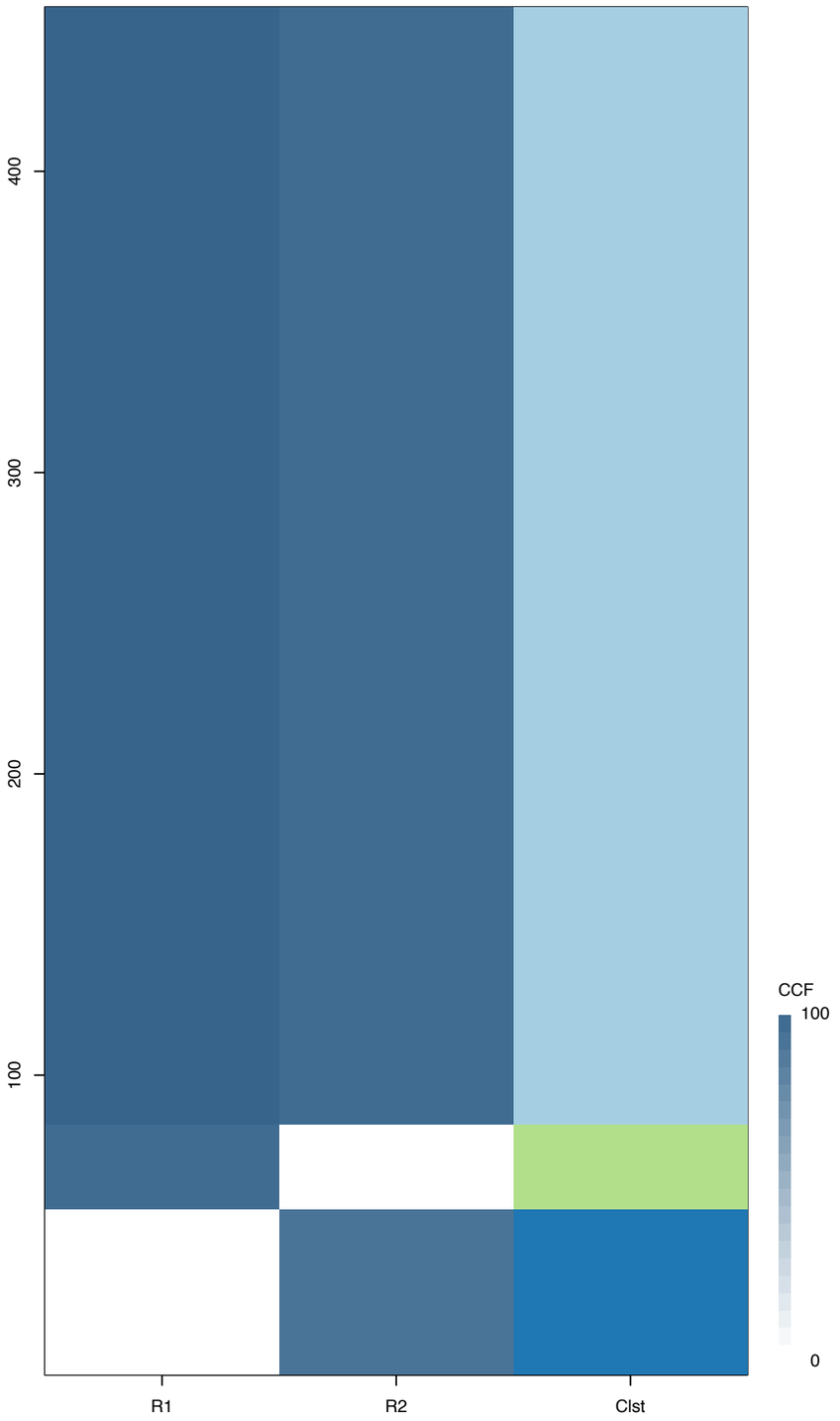
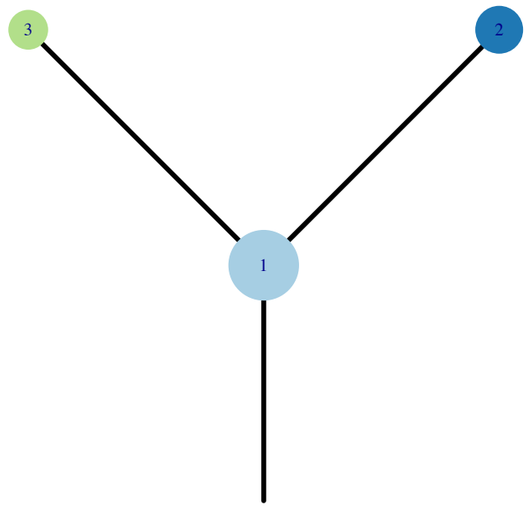


Fig.S12BO



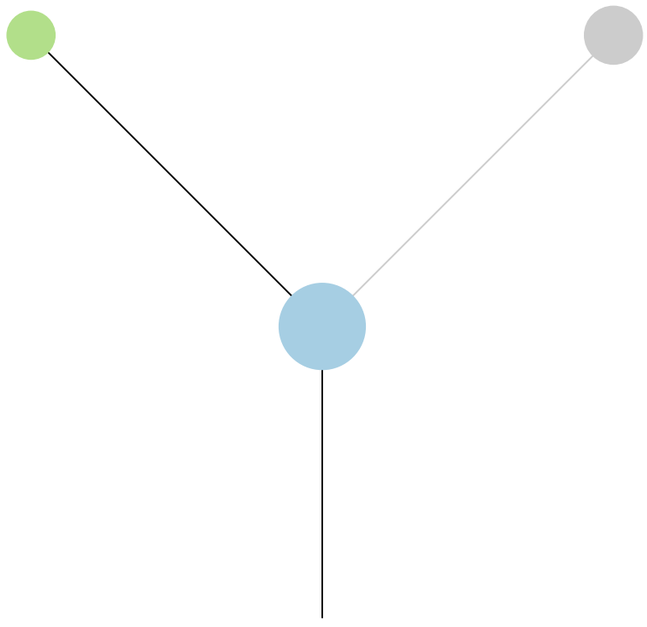
CRUK0073



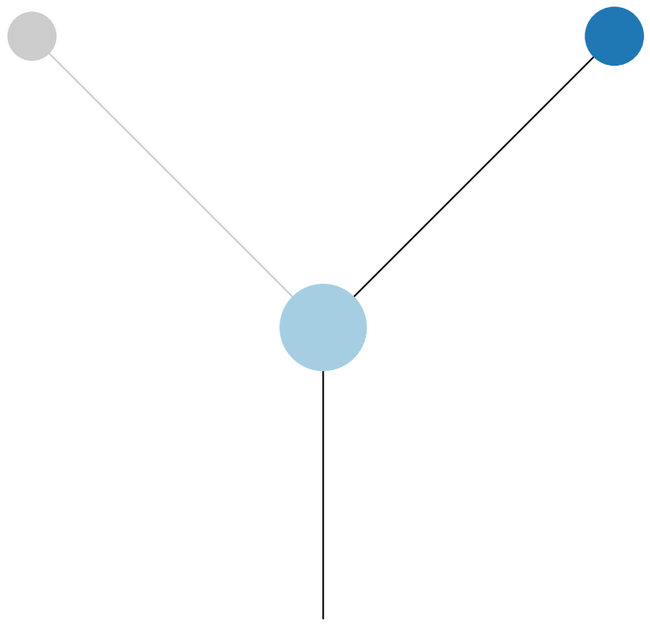
Histology:Squamous, Age:80, PackYears:16, Size:21
Stage:1a, Gender:Female, GD:Clonal GD, Recur:no

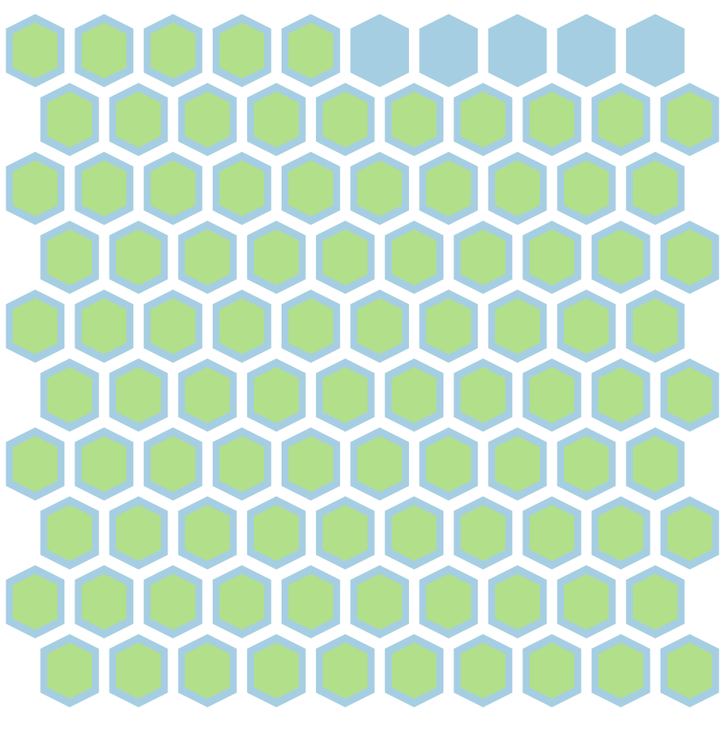
Gene	Cluster	Cytoband	Type
NOTCH2	1	1p12	SNV
NFE2L2	1	2q31.2	SNV
FAT1	1	4q35.2	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
KAT6A	1	8p11.21	Amp
IKBKB	1	8p11.21	Amp
HOOK3	1	8p11.21	Amp
TCEA1	1	8q11.23	Amp
PLAG1	1	8q12.1	Amp
CHCHD7	1	8q12.1	Amp
NCOA2	1	8q13.3	Amp
HEY1	1	8q21.13	Amp
CNBD1	1	8q21.3	Amp
RUNX1T1	1	8q21.3	Amp
COX6C	1	8q22.2	Amp
RSPO2	1	8q23.1	Amp
EIF3E	1	8q23.1	Amp
MYC	1	8q24.21	Amp
NDRG1	1	8q24.22	Amp
CDKN2A	1	9p21.3	SNV
KDM5A	1	12p13.33	Amp
ERC1	1	12p13.33	Amp
KMT2D	1	12q13.12	SNV
DICER1	1	14q32.13	SNV
PLXNB2	2	22q13.33	SNV

R1



R2





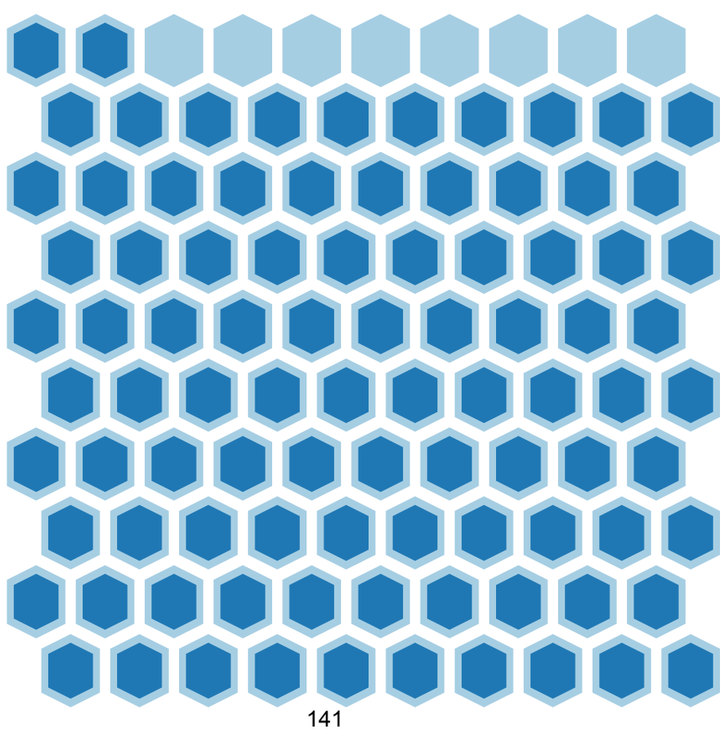
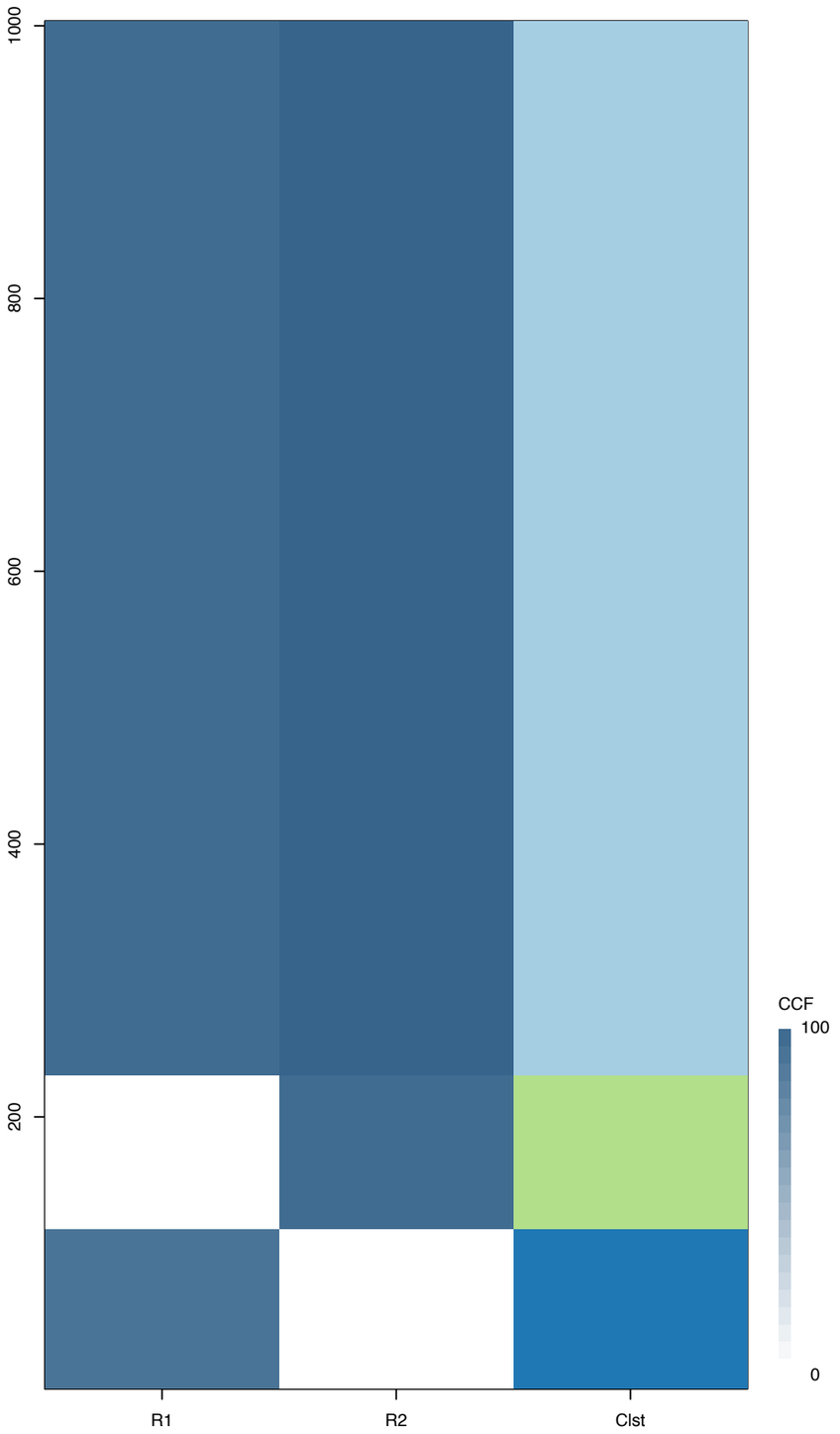
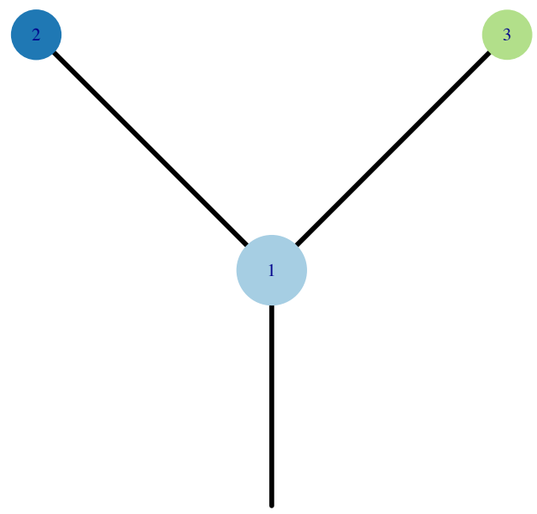


Fig.S12BP



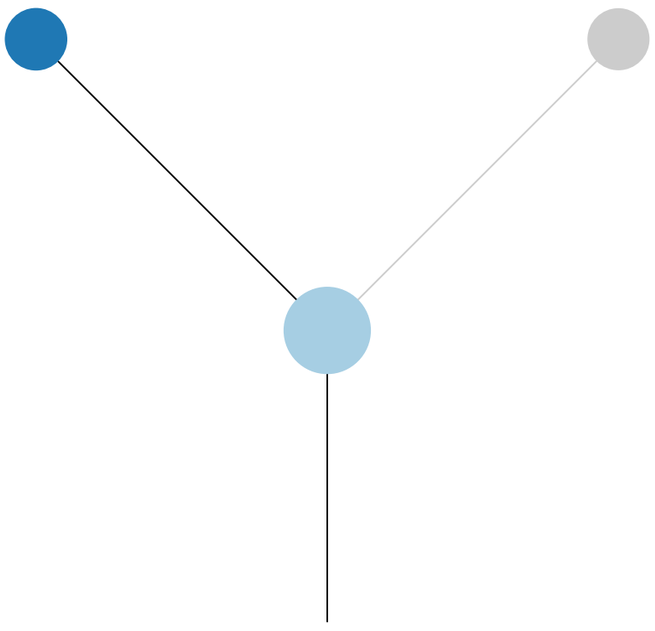
CRUK0074



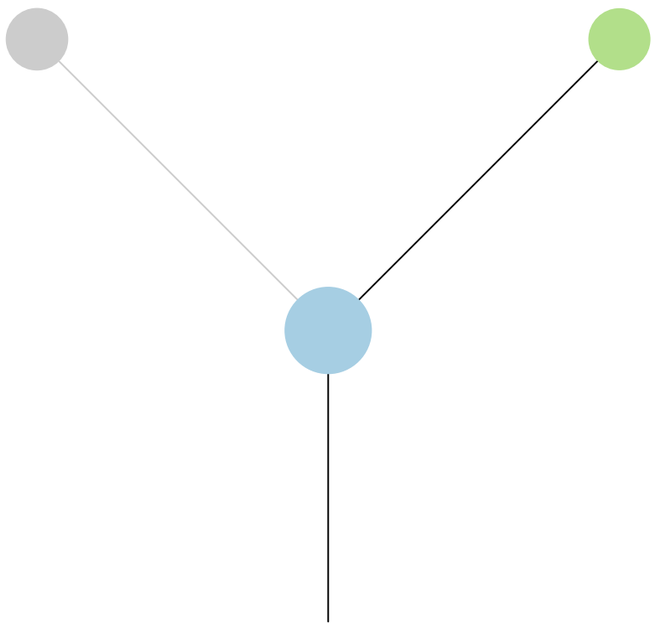
Histology:Squamous, Age:81, PackYears:79.5, Size:60
Stage:2b, Gender:Male, GD:Clonal GD, Recur:no

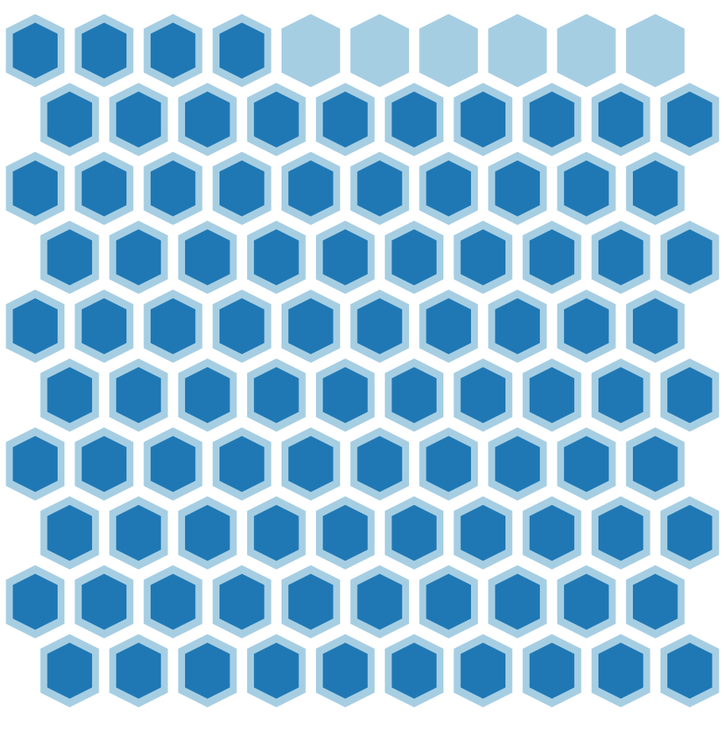
Gene	Cluster	Cytoband	Type
NFE2L2	1	2q31.2	SNV
MLF1	1	3q25.32	Amp
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
CCND1	1	11q13.3	Amp
NKX2−1	1	14q13.3	Amp
FOXA1	1	14q21.1	Amp
TP53	1	17p13.1	SNV
AKT2	1	19q13.2	Amp
CD79A	1	19q13.2	Amp
UBR5	3	8q22.3	SNV
LSM14A	3	19q13.11	Amp

R1



R2





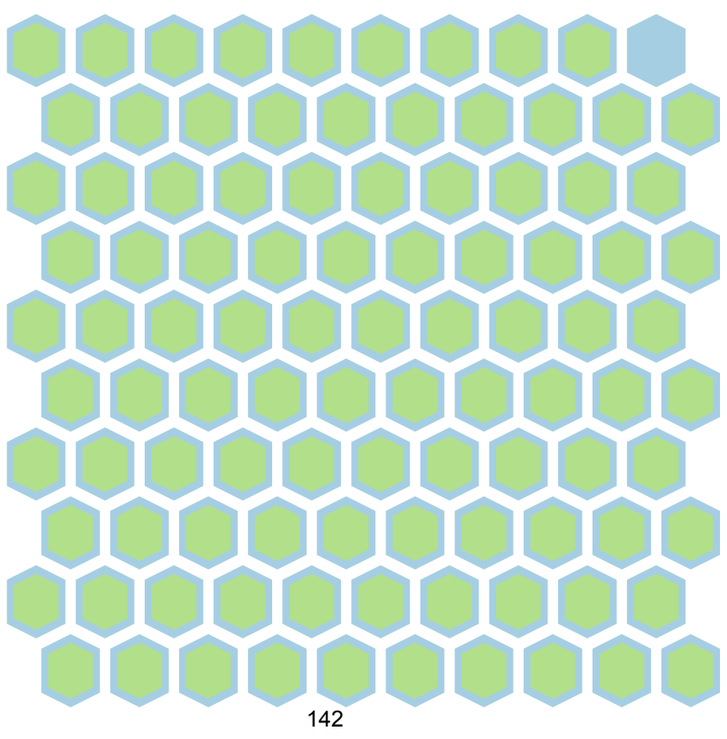
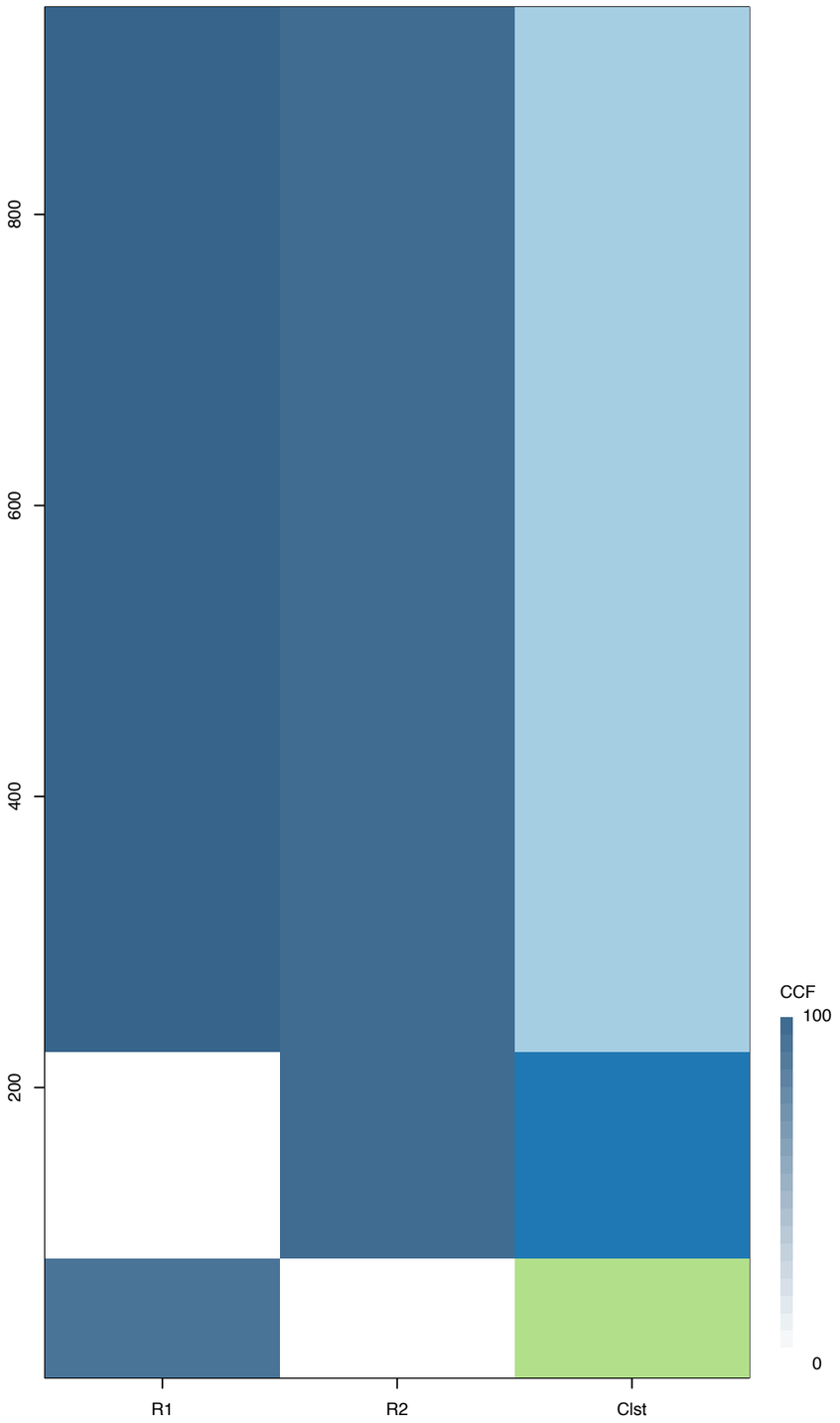
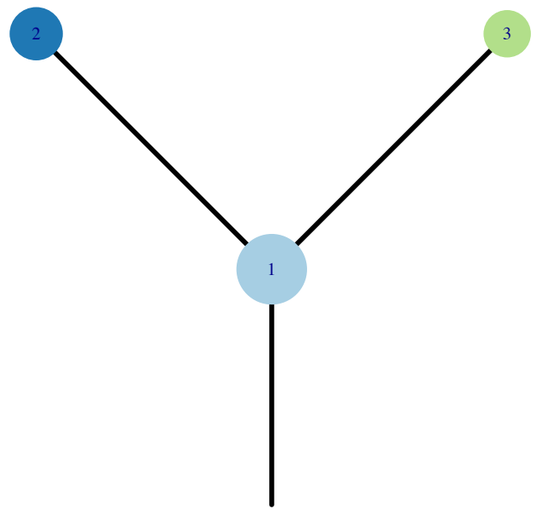


Fig.S12BQ



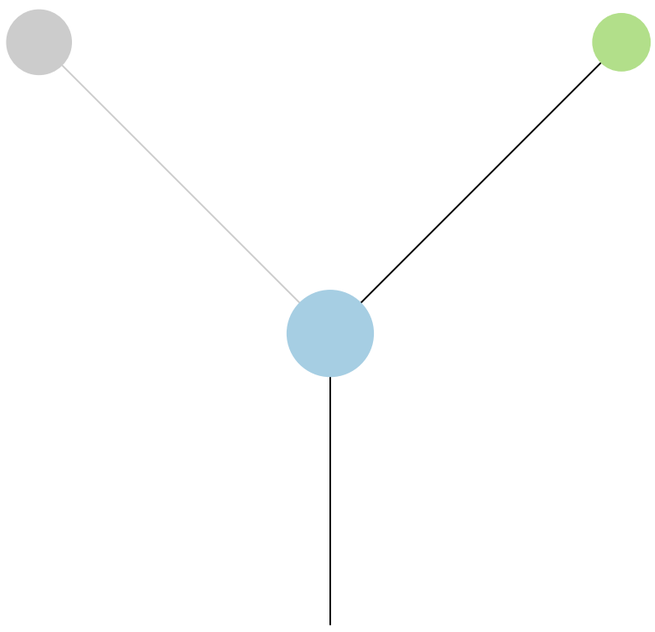
CRUK0075



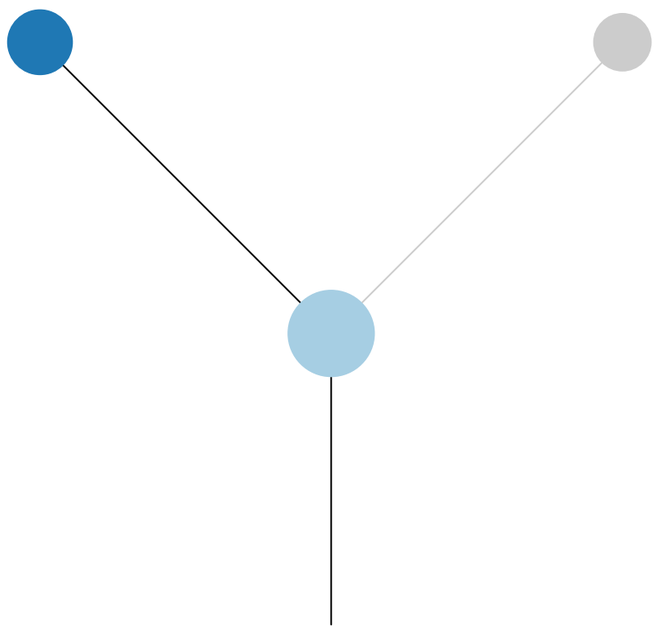
Gene	Cluster	Cytoband	Type
PIK3CA	1	3q26.32	SNV
FAT1	1	4q35.2	SNV
RASA1	1	5q14.3	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
NOTCH1	1	9q34.3	SNV
MGA	1	15q15.1	SNV
TP53	1	17p13.1	SNV
CCNE1	1	19q12	Amp
CEP89	1	19q13.11	Amp
CEBPA	1	19q13.11	Amp
LSM14A	1	19q13.11	Amp
AKT2	1	19q13.2	Amp
CD79A	1	19q13.2	Amp
ZNF180	1	19q13.31	Amp
BCL3	1	19q13.32	Amp
NFE2L2	2	2q31.2	SNV
ERCC5	?	13q33.1	SNV

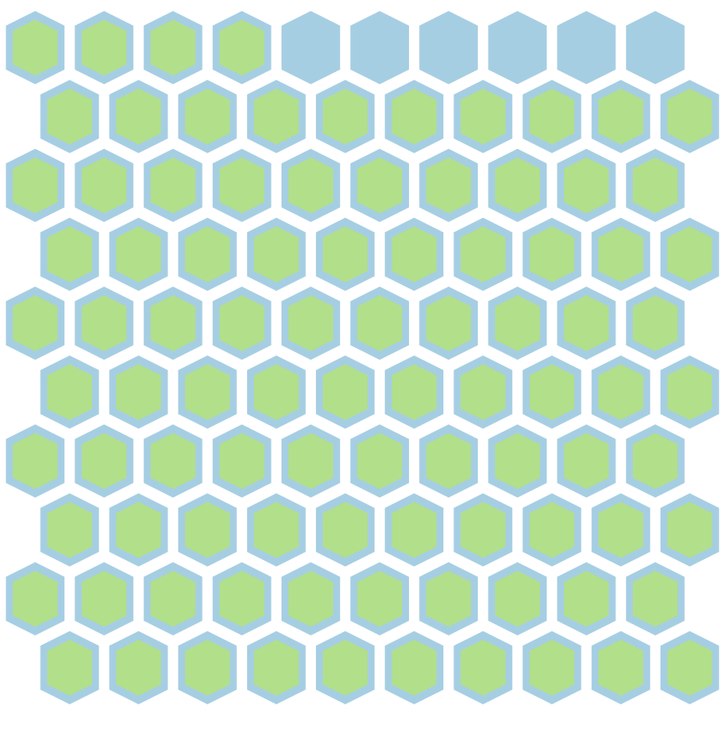
Histology:Squamous, Age:76, PackYears:41.3, Size:30
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

R1



R2





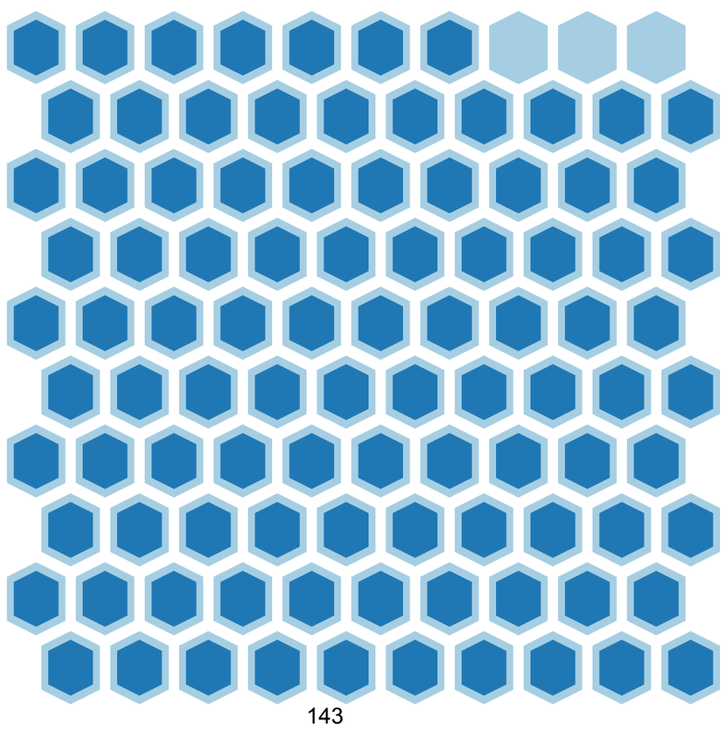
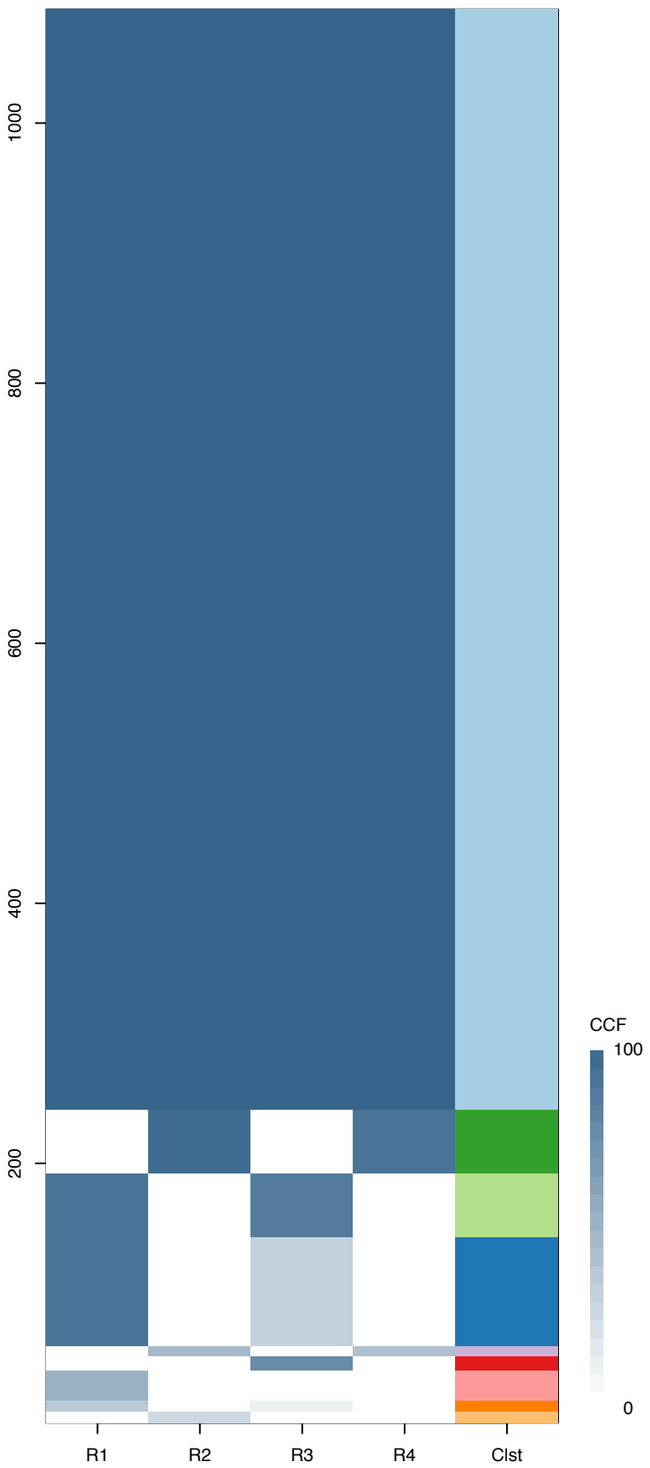
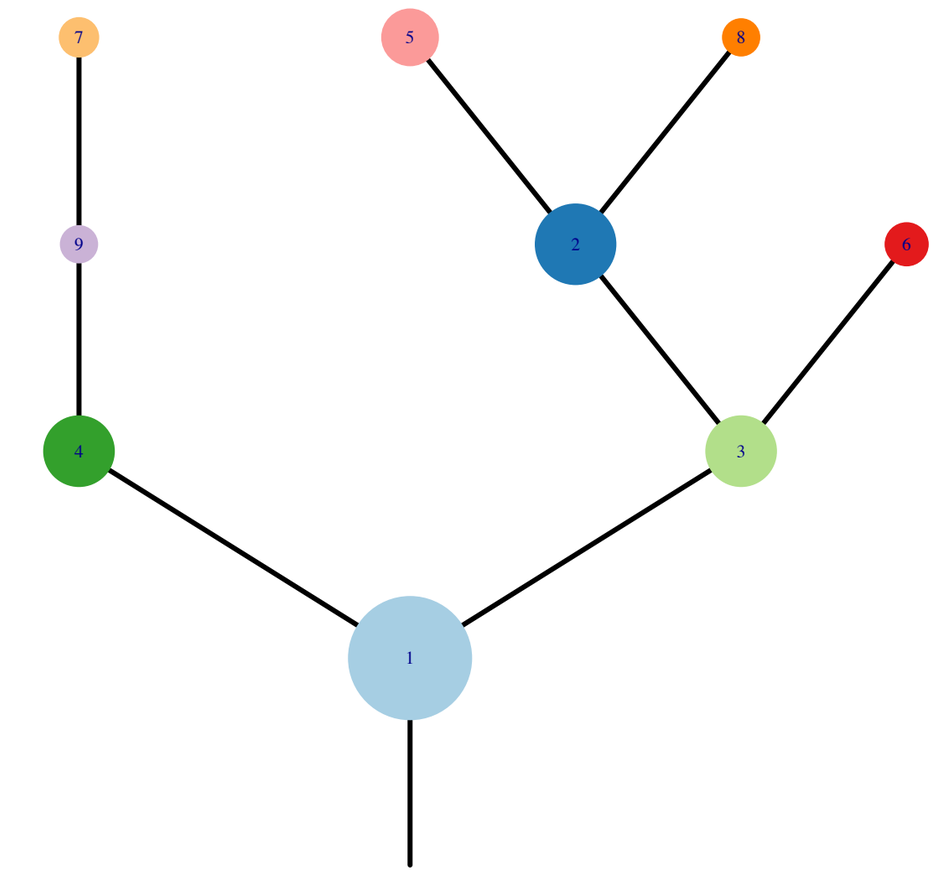


Fig.S12BR



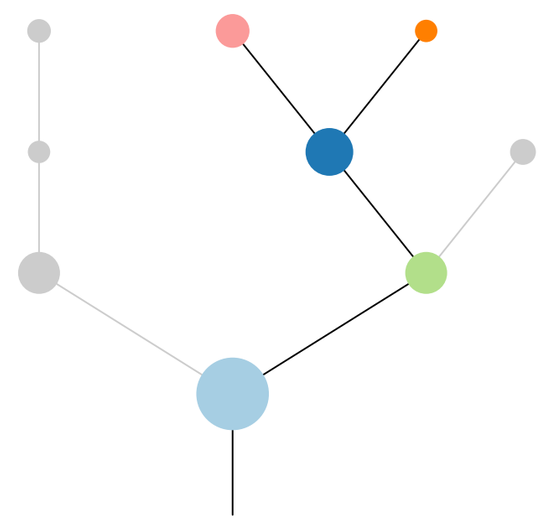
CRUK0076_A



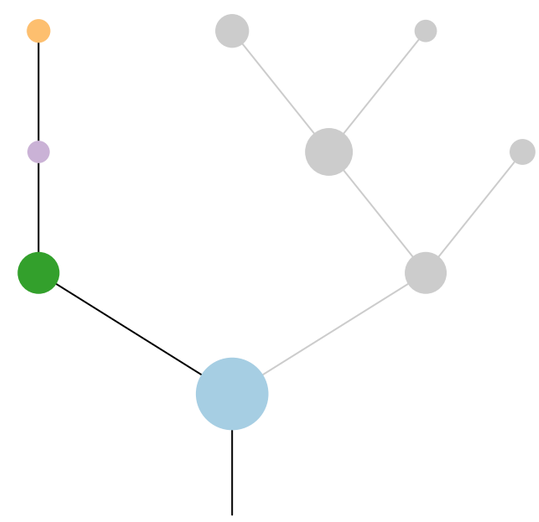
Histology:Squamous, Age:65, PackYears:102, Size:32
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
REL	1	2p16.1	Amp
XPO1	1	2p15	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
ARID1B	1	6q25.3	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
TP53	1	17p13.1	SNV
SERPINB13	1	18q21.33	SNV
NCOR1	2	17p11.2	SNV
COL5A2	3	2q32.2	SNV
BCL11A	?	2p16.1	Amp
ZRSR2	?		SNV

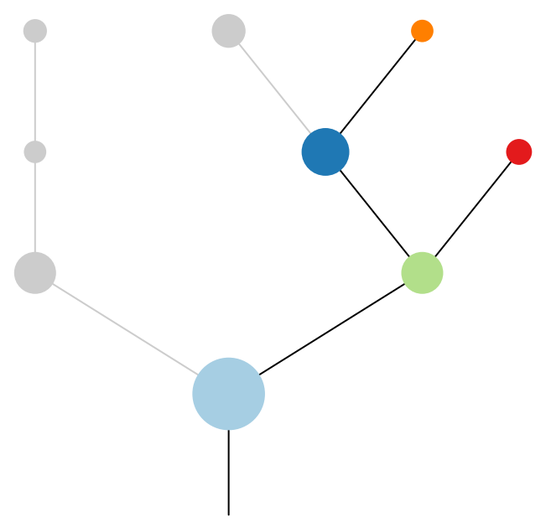
R1



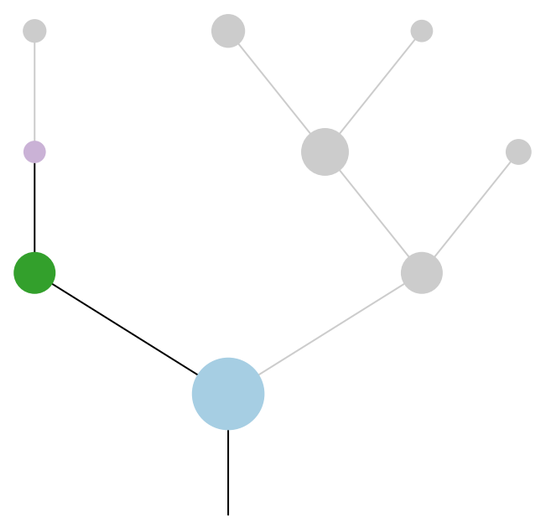
R2

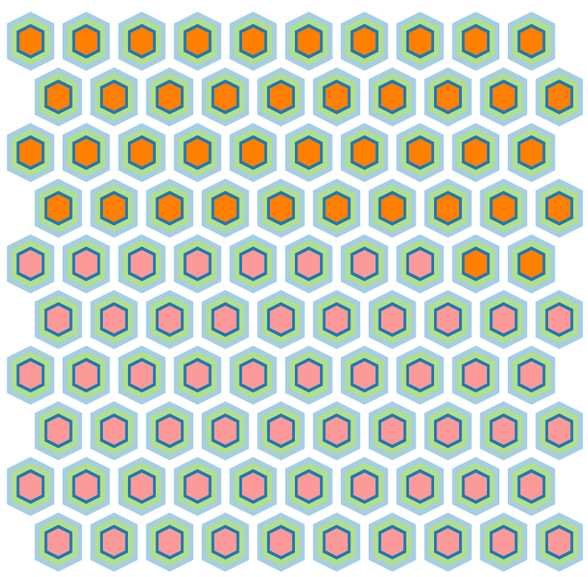


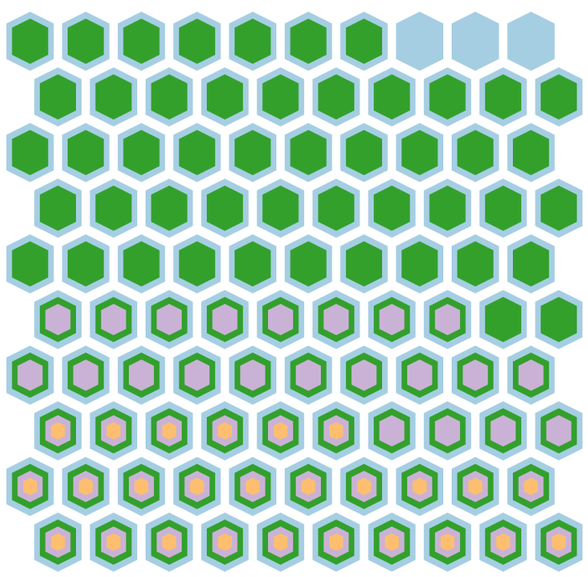
R3

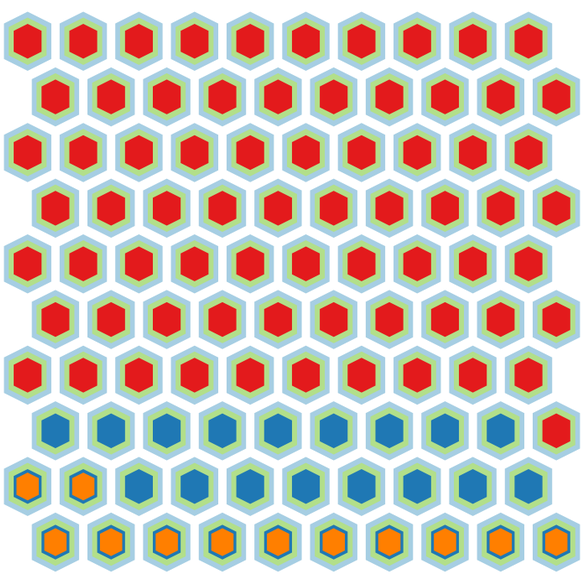


R4









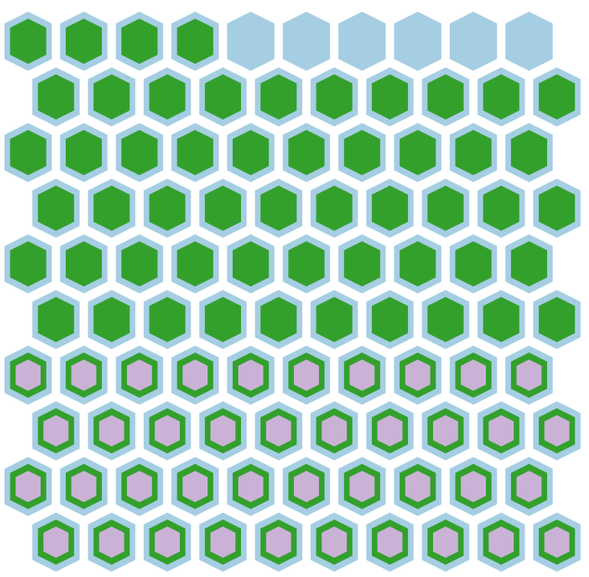
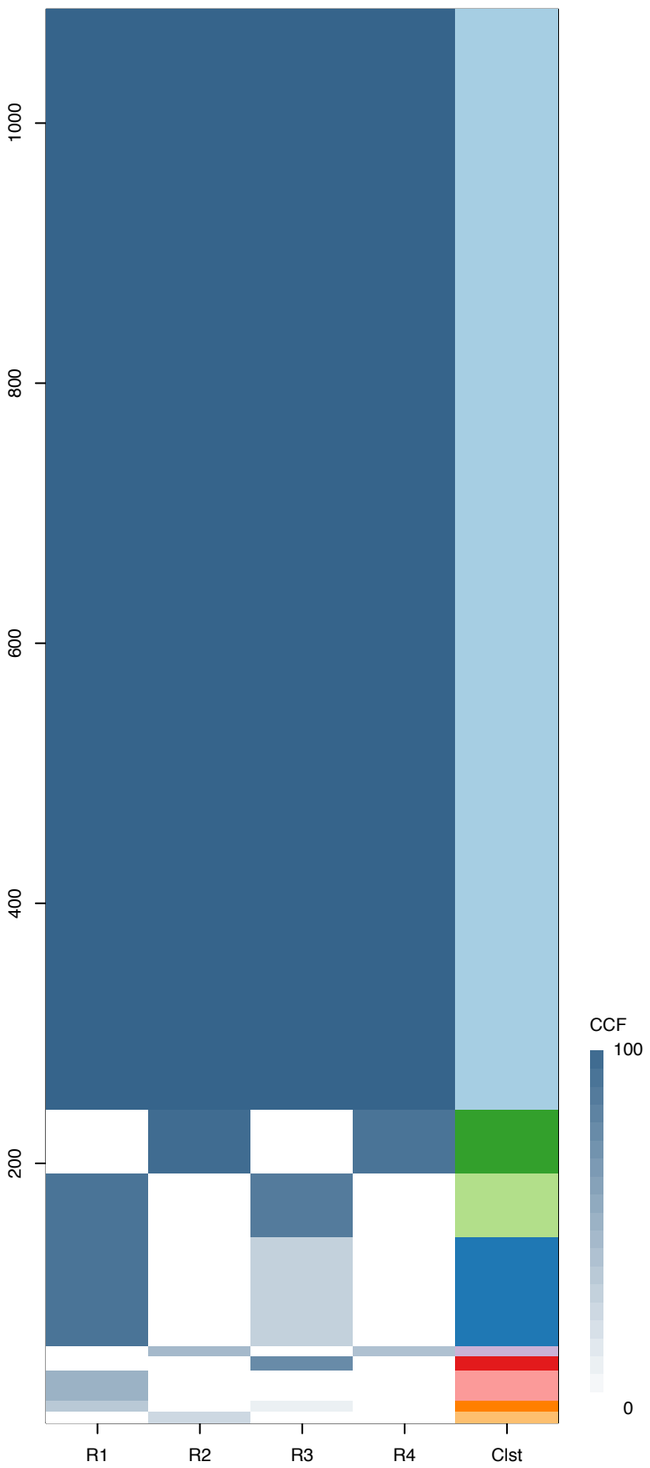
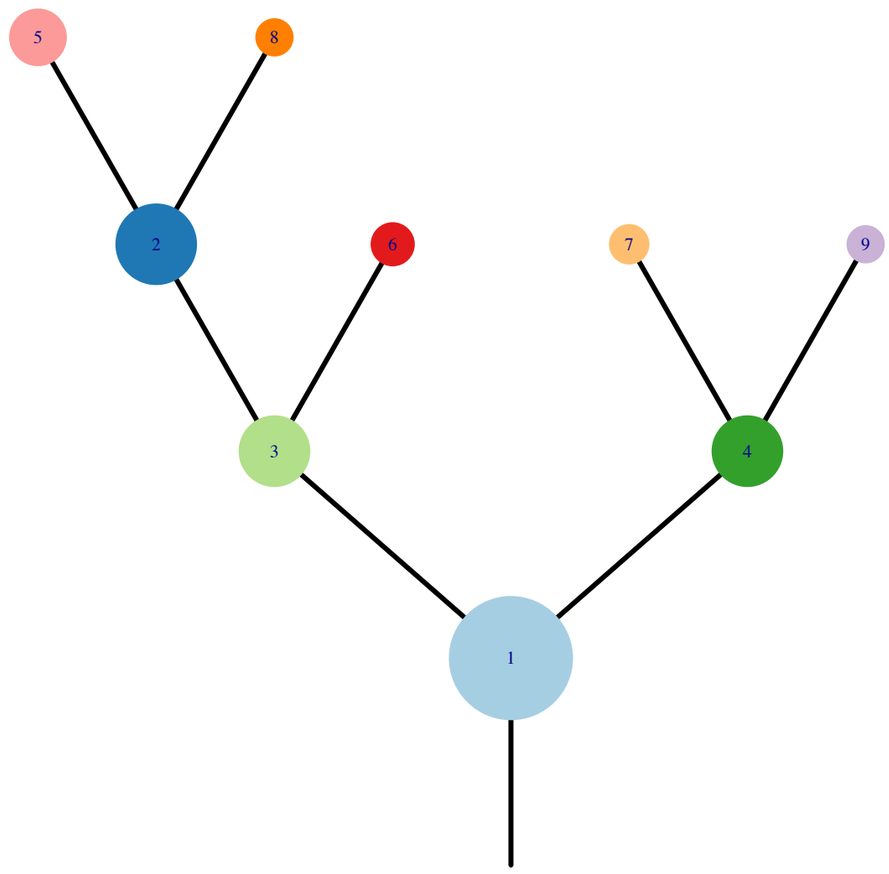


Fig.S12BR



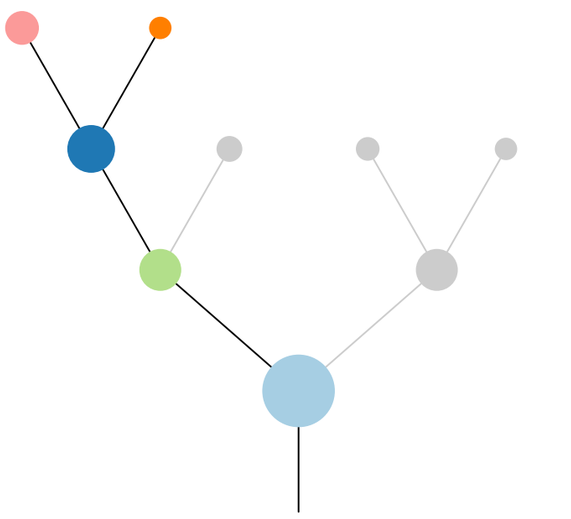
CRUK0076_B



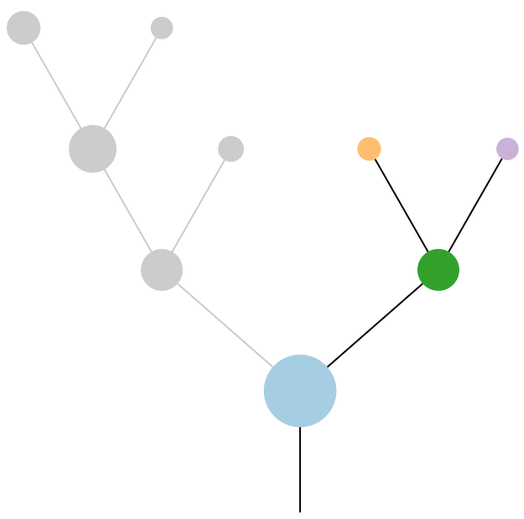
Histology:Squamous, Age:65, PackYears:102, Size:32
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
REL	1	2p16.1	Amp
XPO1	1	2p15	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
ARID1B	1	6q25.3	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
TP53	1	17p13.1	SNV
SERPINB13	1	18q21.33	SNV
NCOR1	2	17p11.2	SNV
COL5A2	3	2q32.2	SNV
BCL11A	?	2p16.1	Amp
ZRSR2	?		SNV

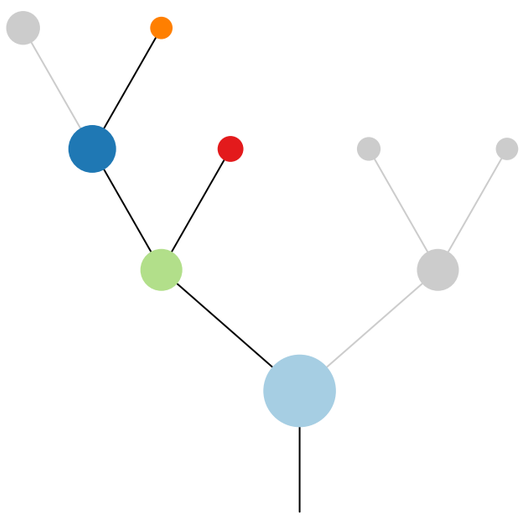
R1



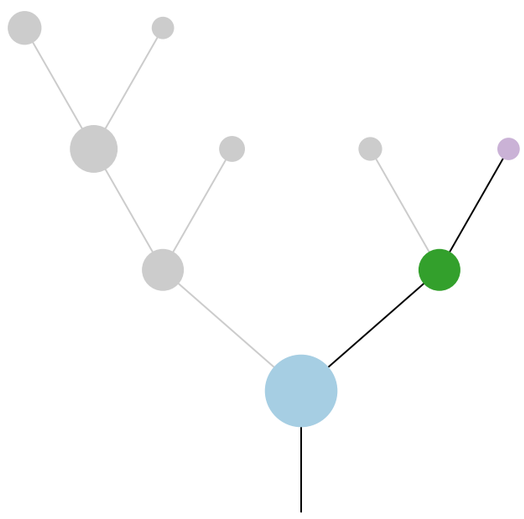
R2



R3



R4



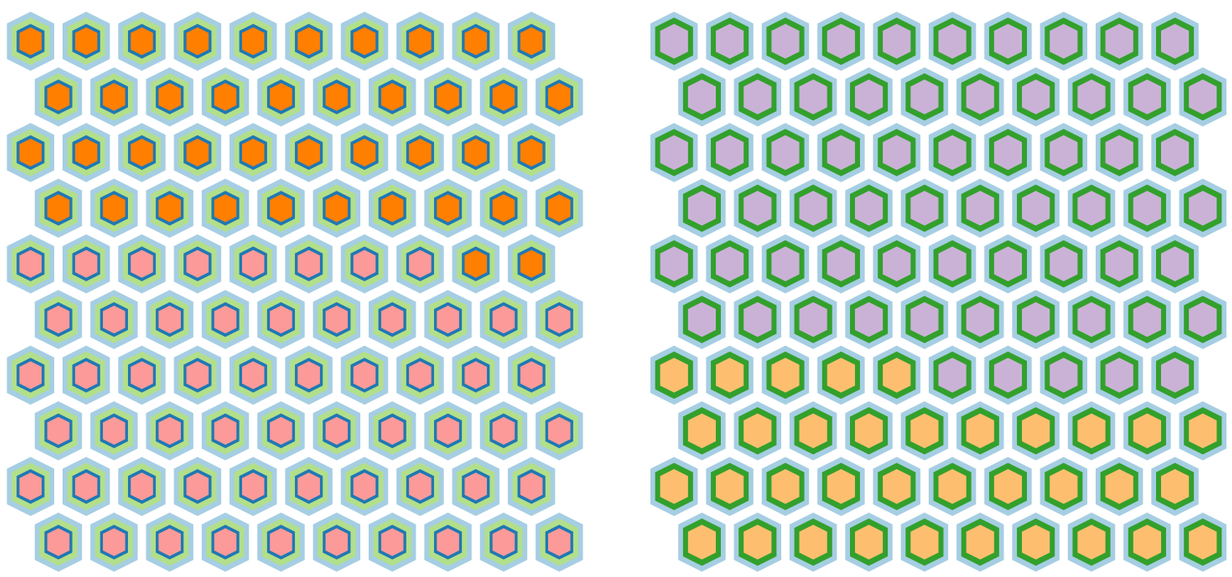
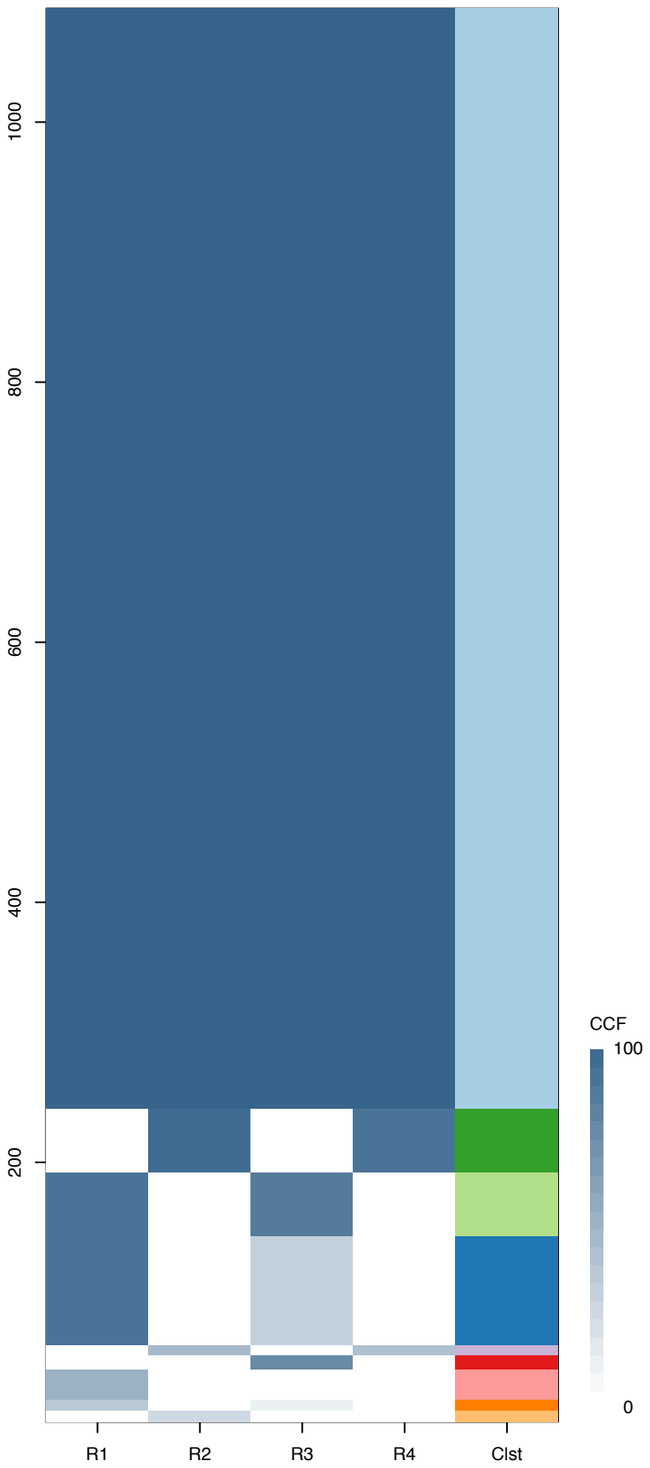
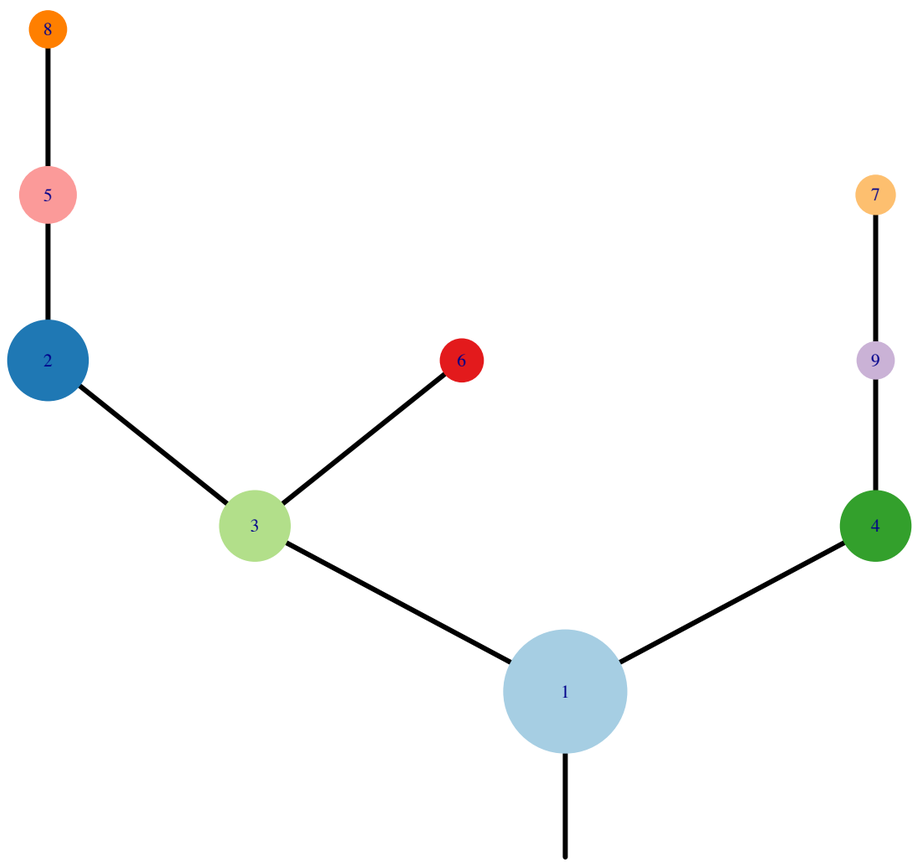


Fig.S12BR



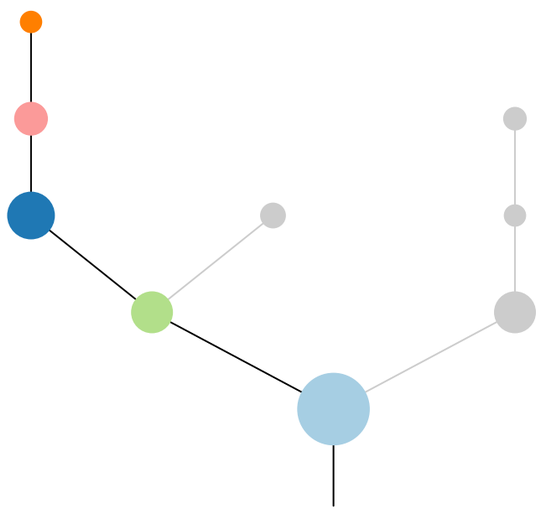
CRUK0076_C



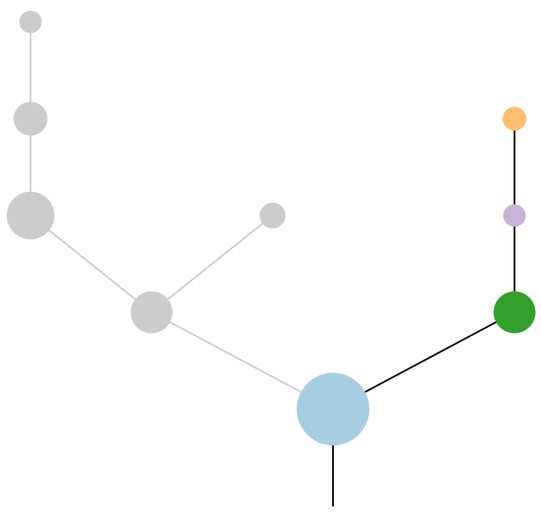
Histology:Squamous, Age:65, PackYears:102, Size:32
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
REL	1	2p16.1	Amp
XPO1	1	2p15	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
ARID1B	1	6q25.3	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
TP53	1	17p13.1	SNV
SERPINB13	1	18q21.33	SNV
NCOR1	2	17p11.2	SNV
COL5A2	3	2q32.2	SNV
BCL11A	?	2p16.1	Amp
ZRSR2	?		SNV

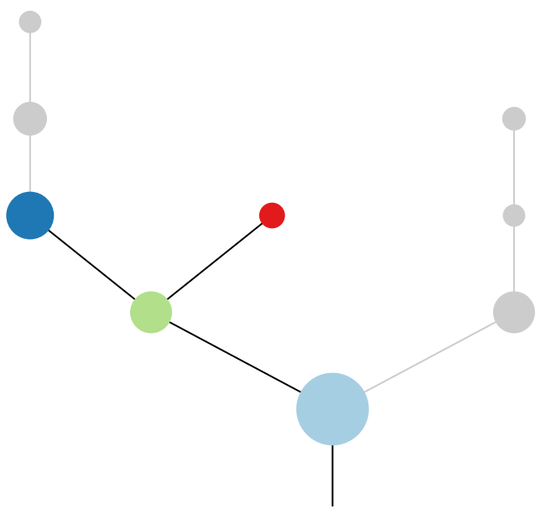
R1



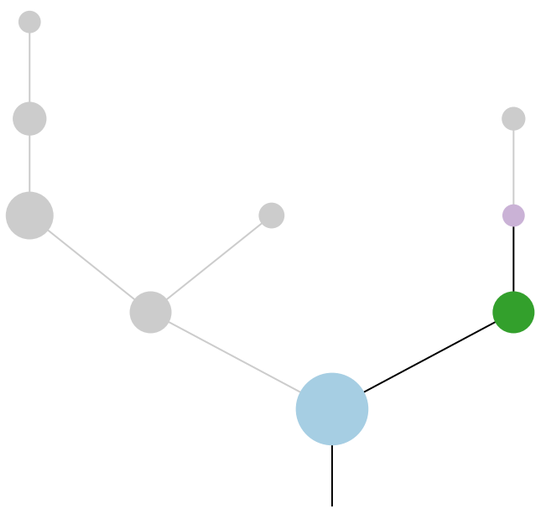
R2



R3



R4



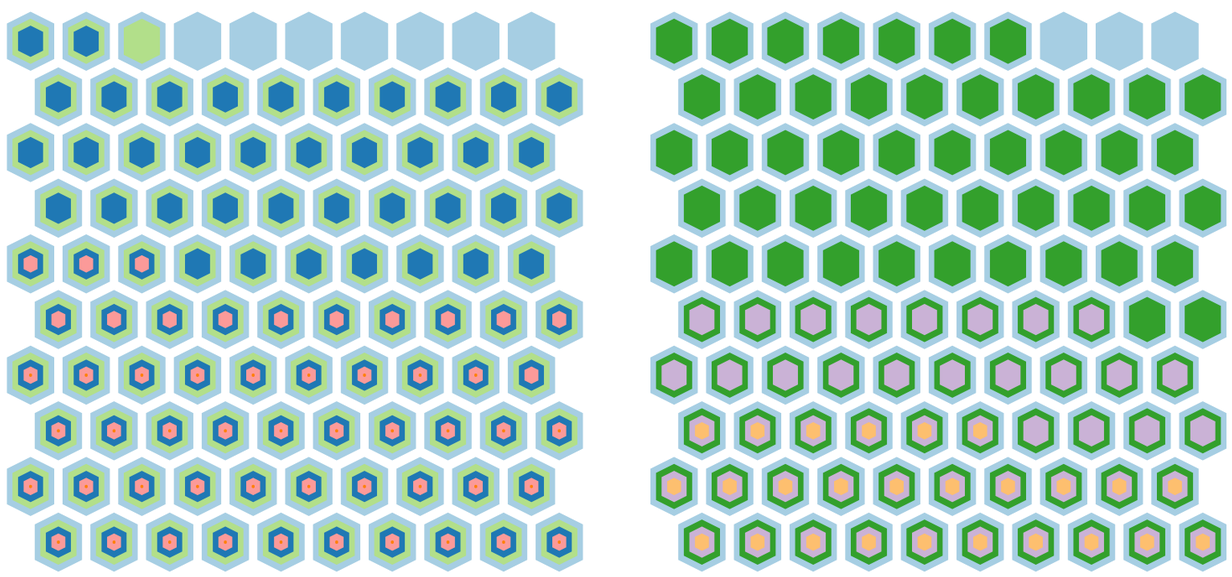
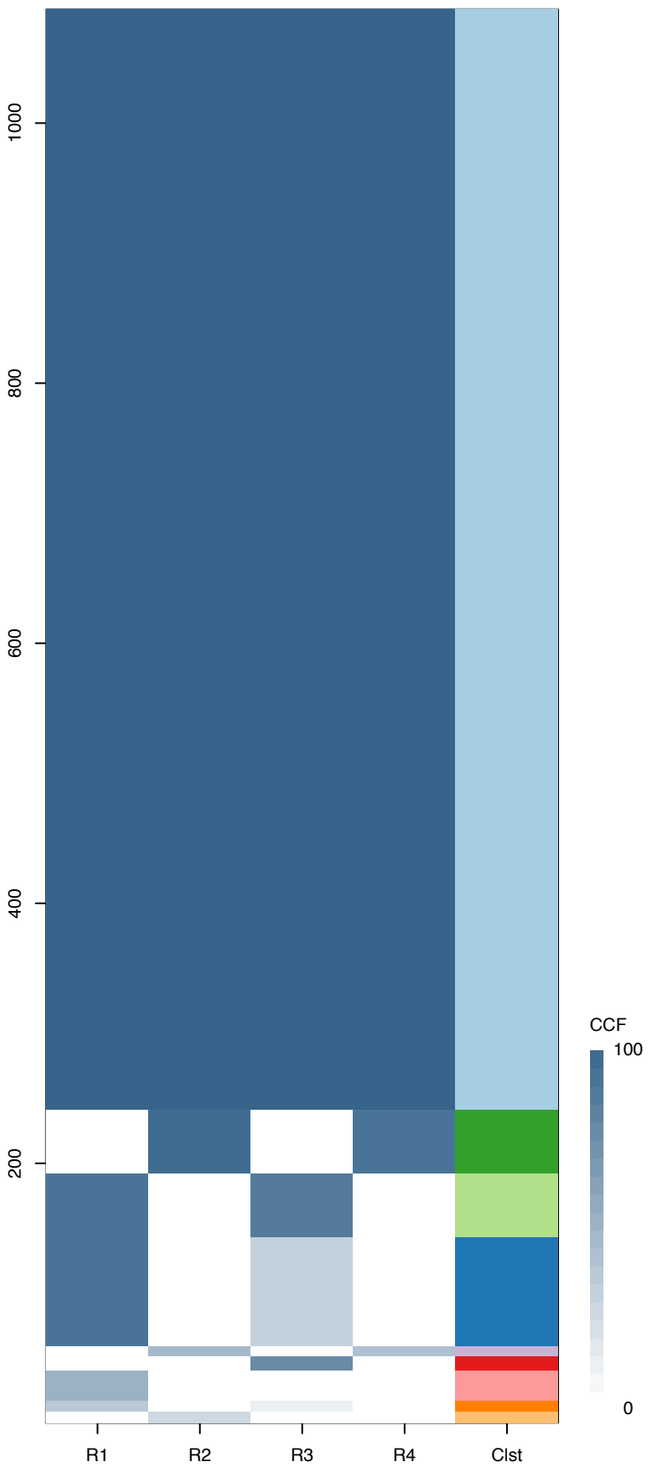
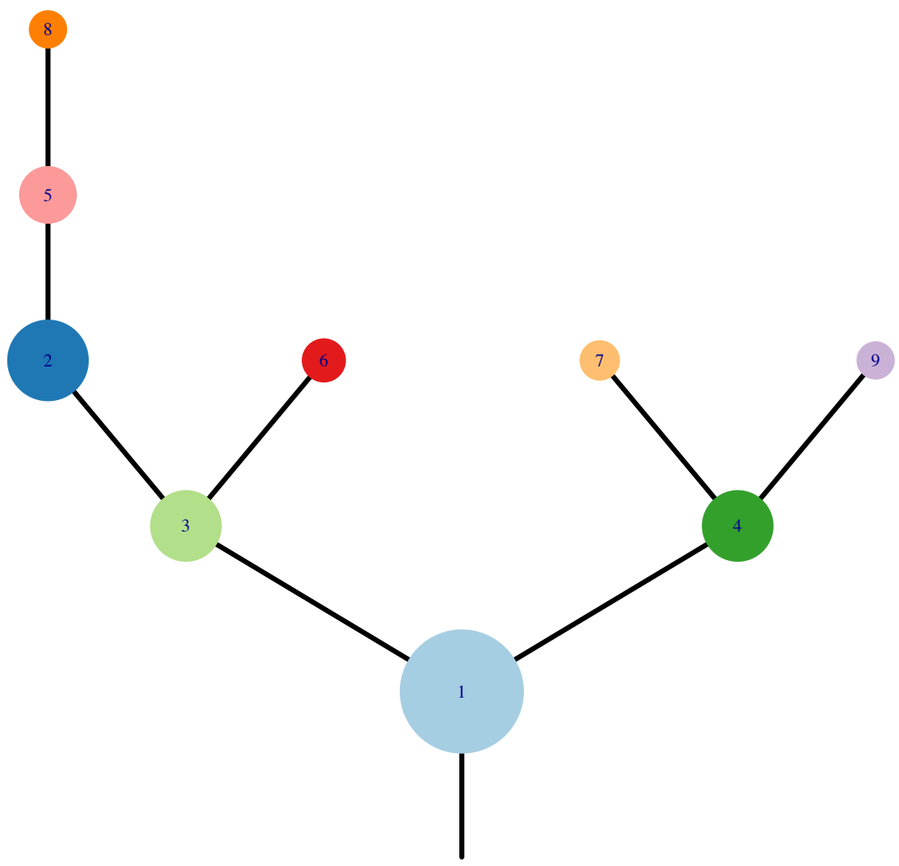


Fig.S12BR



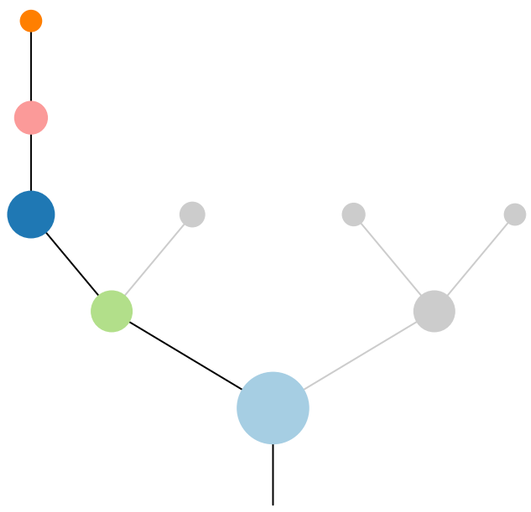
CRUK0076_D



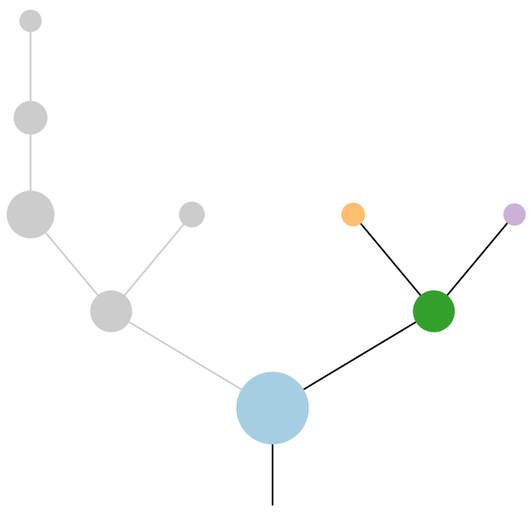
Histology:Squamous, Age:65, PackYears:102, Size:32
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
REL	1	2p16.1	Amp
XPO1	1	2p15	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
IL7R	1	5p13.2	Amp
LIFR	1	5p13.1	Amp
ARID1B	1	6q25.3	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
TP53	1	17p13.1	SNV
SERPINB13	1	18q21.33	SNV
NCOR1	2	17p11.2	SNV
COL5A2	3	2q32.2	SNV
BCL11A	?	2p16.1	Amp
ZRSR2	?		SNV

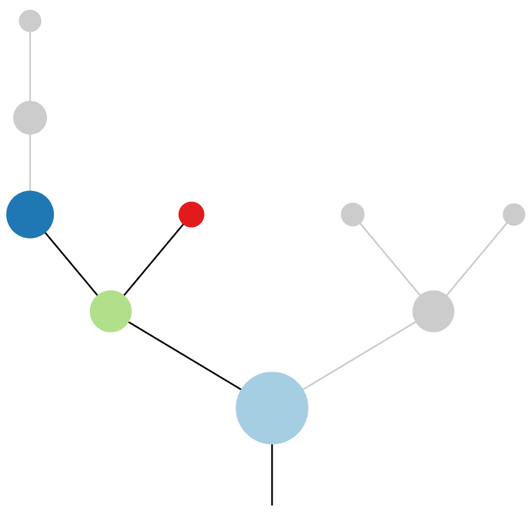
R1



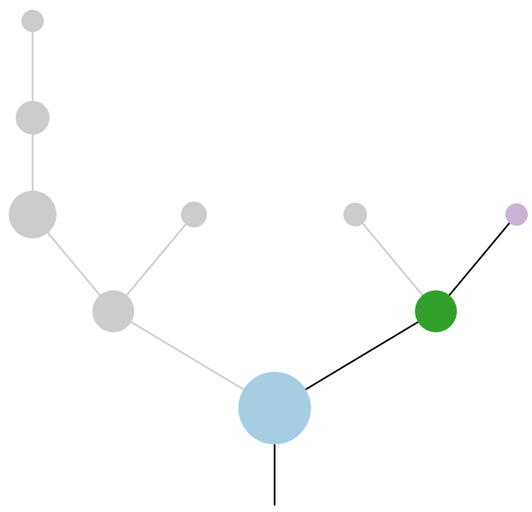
R2



R3



R4



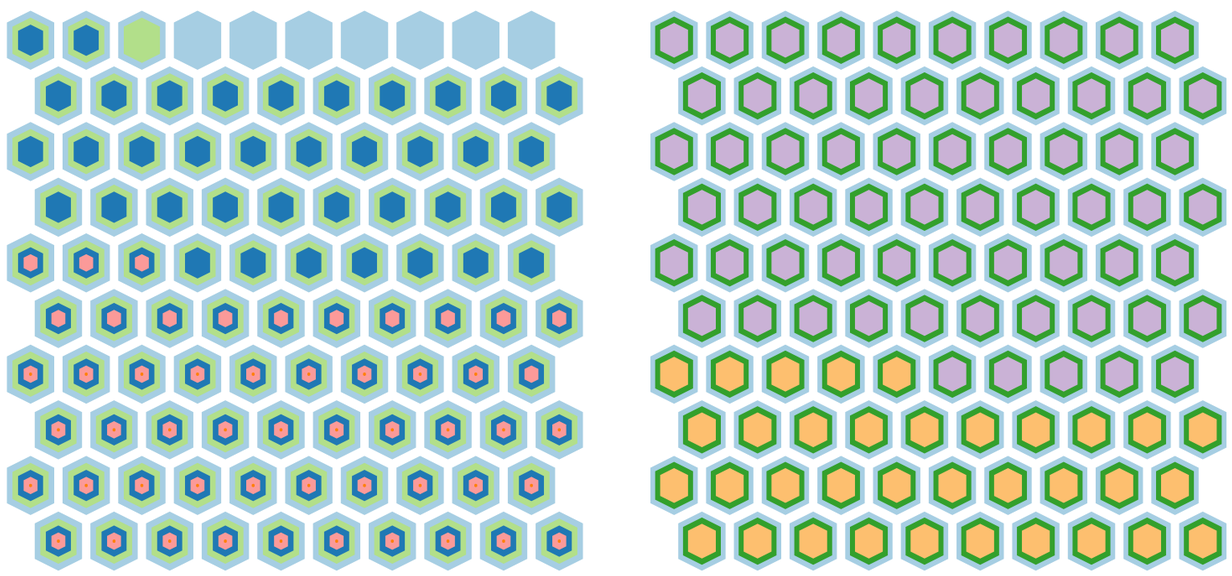
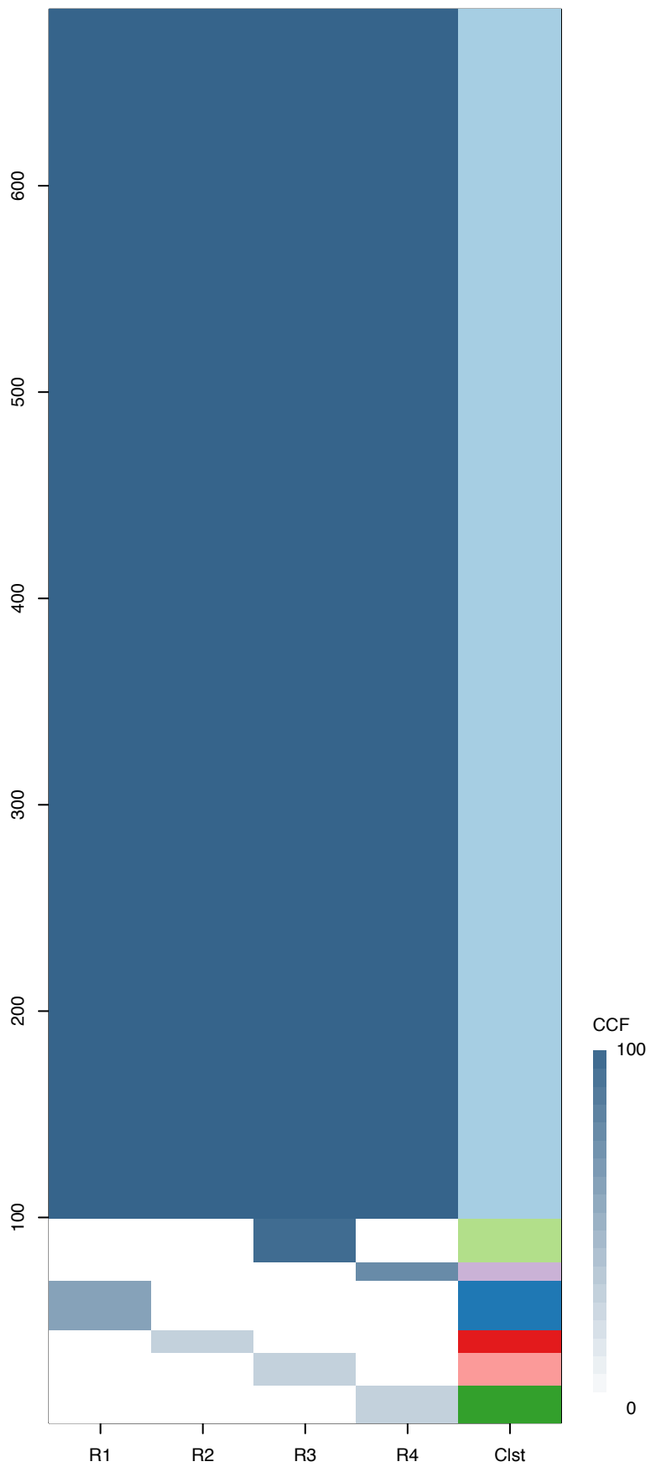
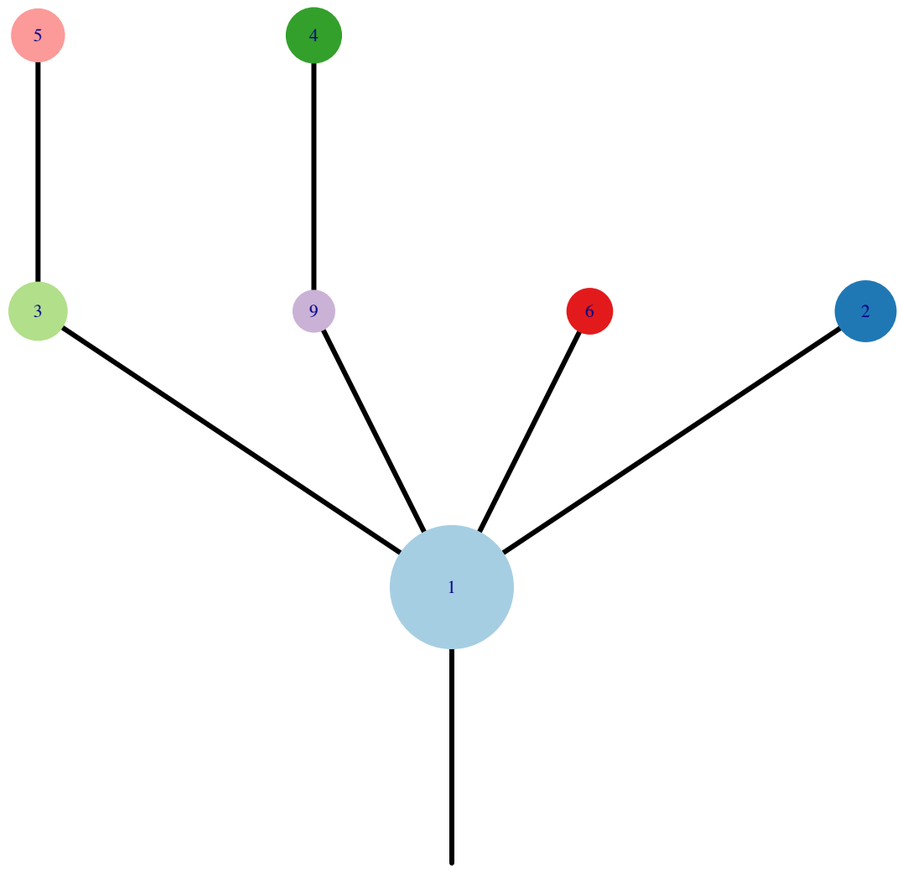


Fig.S12BS



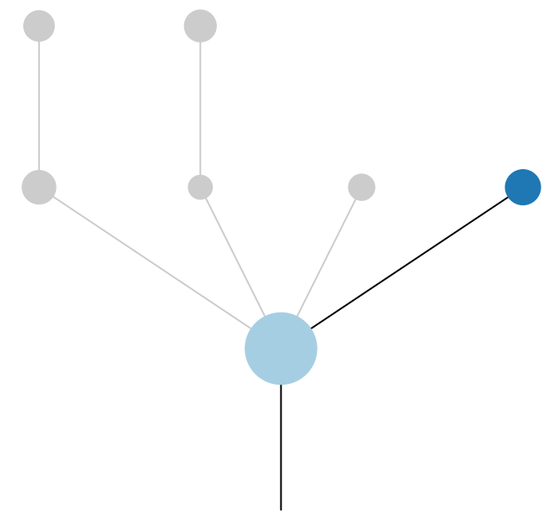
CRUK0077_A



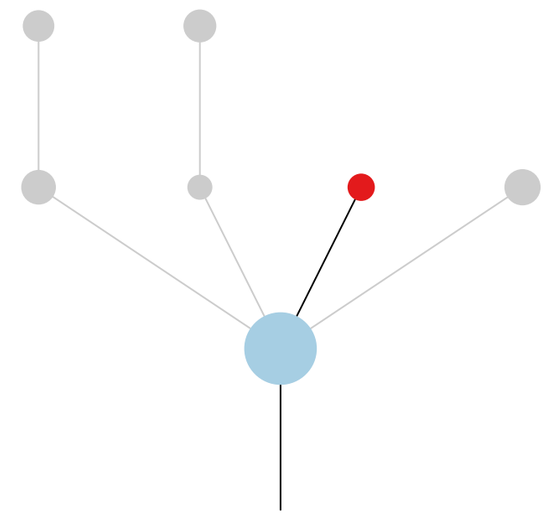
Histology:Squamous, Age:74, PackYears:58, Size:35
Stage:1b, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
TFEB	1	6p21.1	Amp
CCND3	1	6p21.1	Amp
HSP90AB1	1	6p21.1	Amp
LATS1	1	6q25.1	SNV
TP53	1	17p13.1	SNV
KEAP1	1	19p13.2	SNV

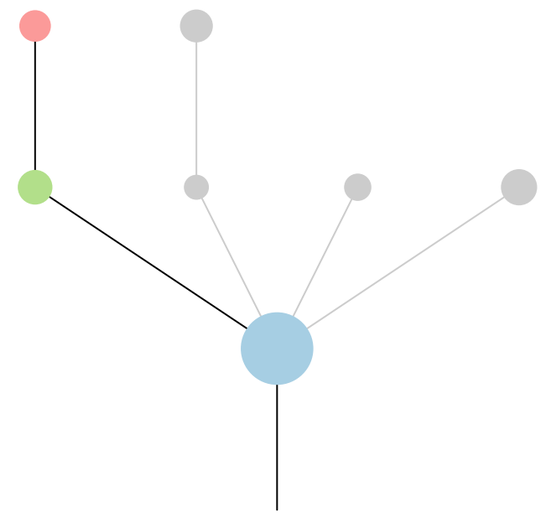
R1



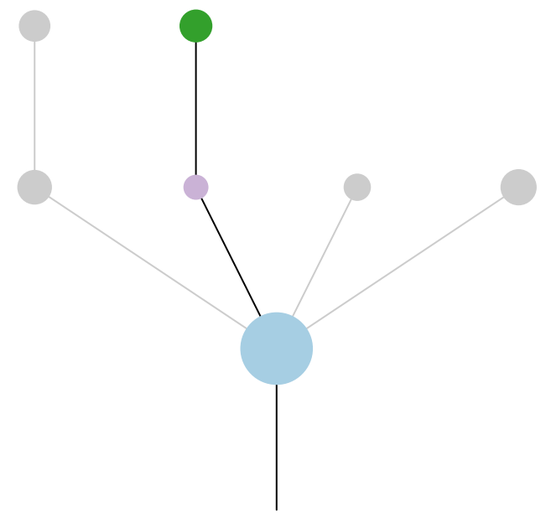
R2

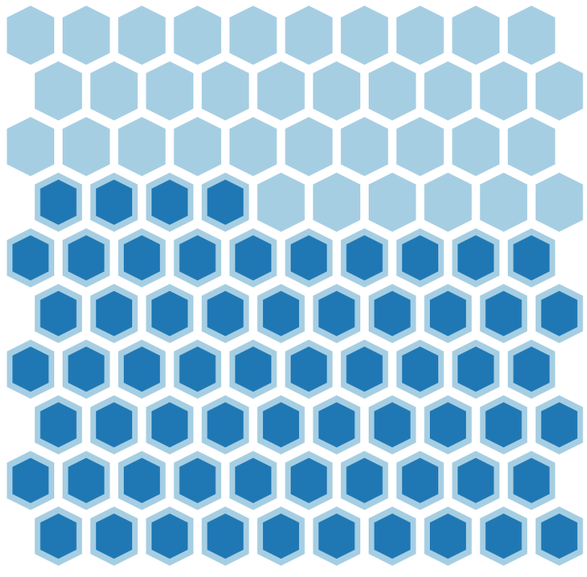


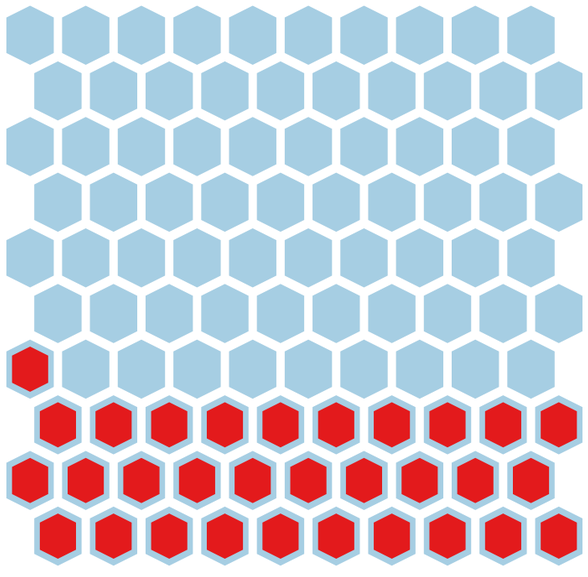
R3

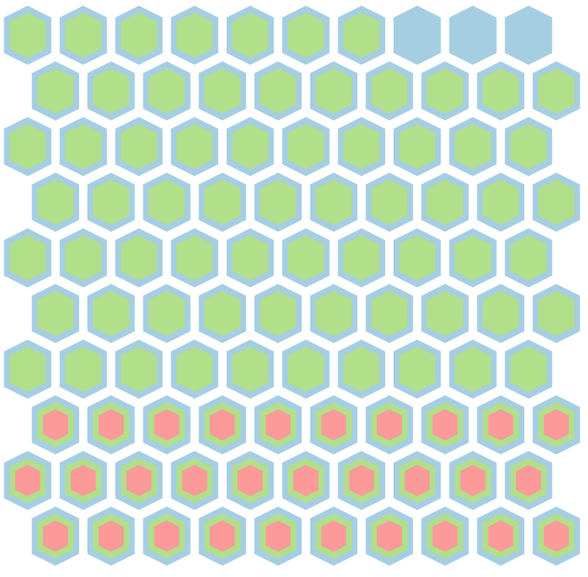


R4









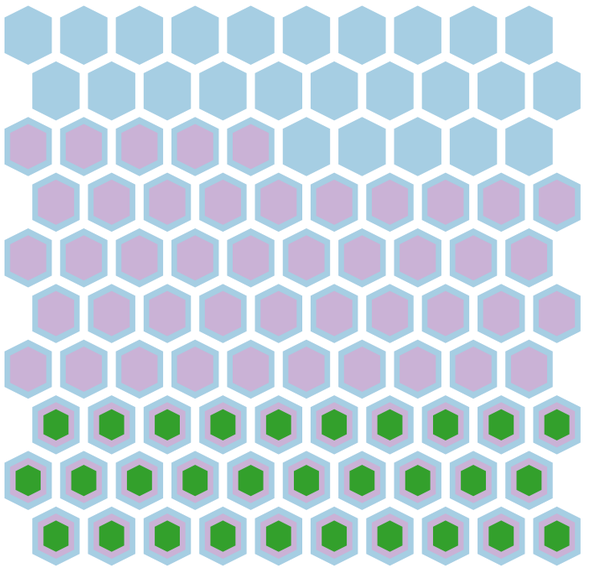
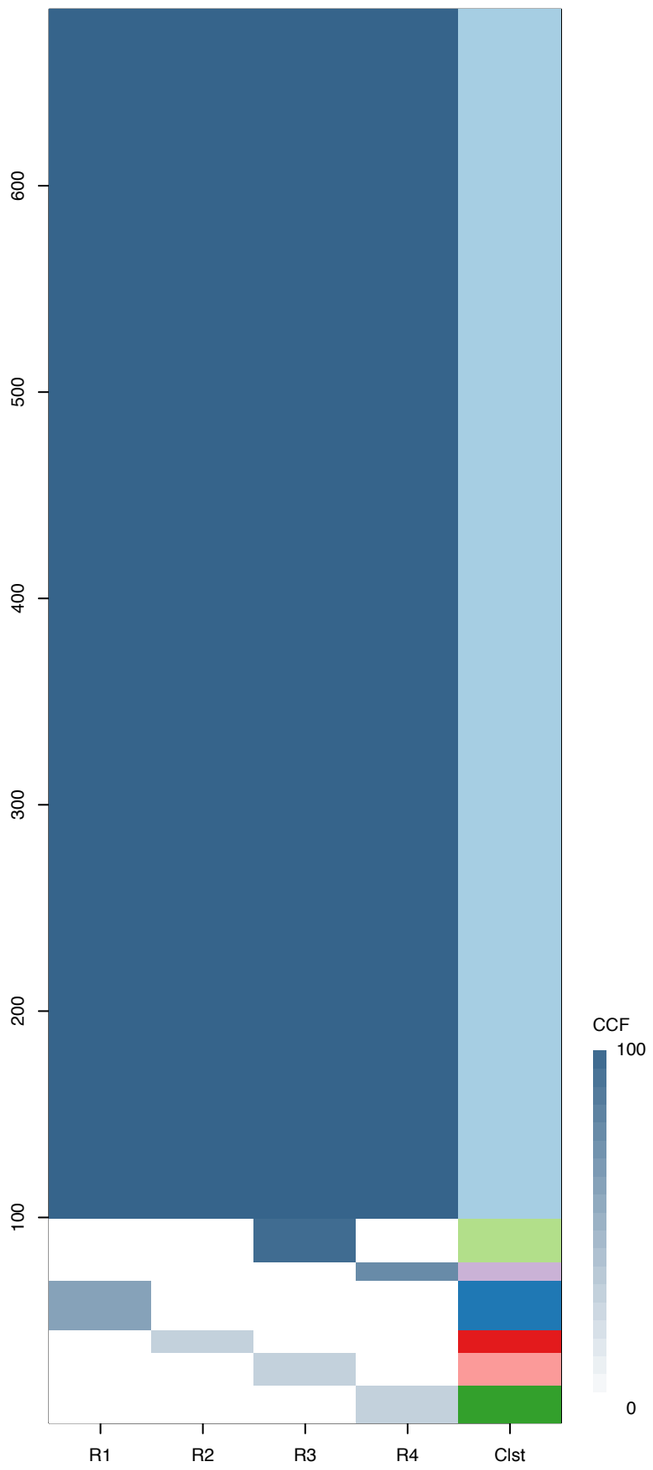
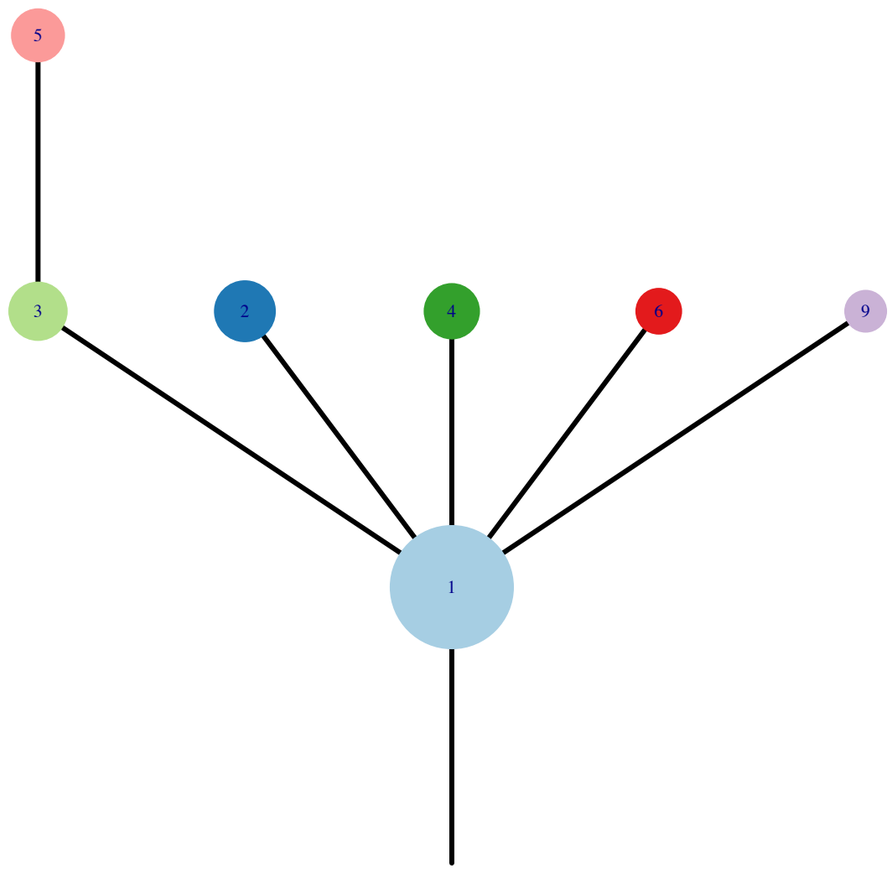


Fig.S12BS



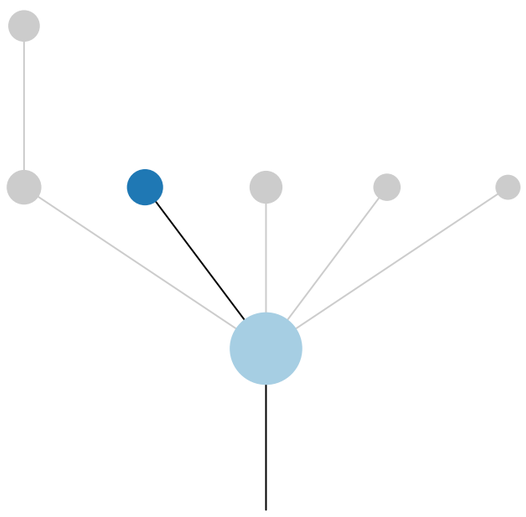
CRUK0077_B



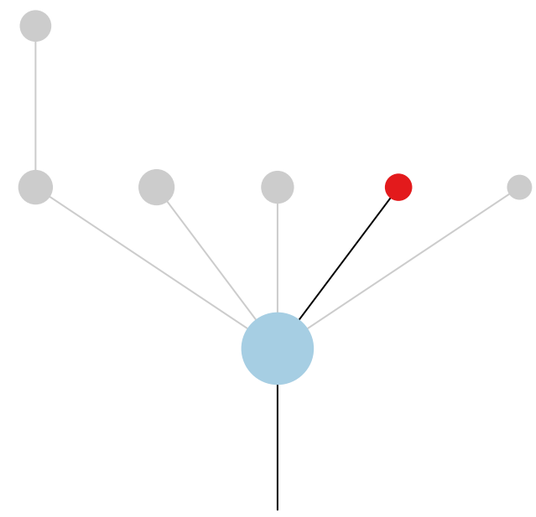
Histology:Squamous, Age:74, PackYears:58, Size:35
Stage:1b, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
TFEB	1	6p21.1	Amp
CCND3	1	6p21.1	Amp
HSP90AB1	1	6p21.1	Amp
LATS1	1	6q25.1	SNV
TP53	1	17p13.1	SNV
KEAP1	1	19p13.2	SNV

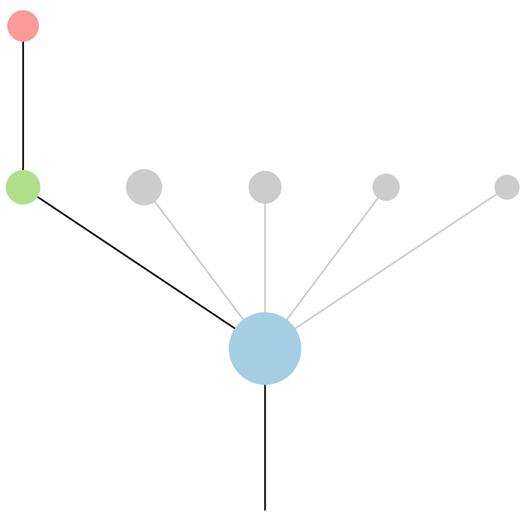
R1



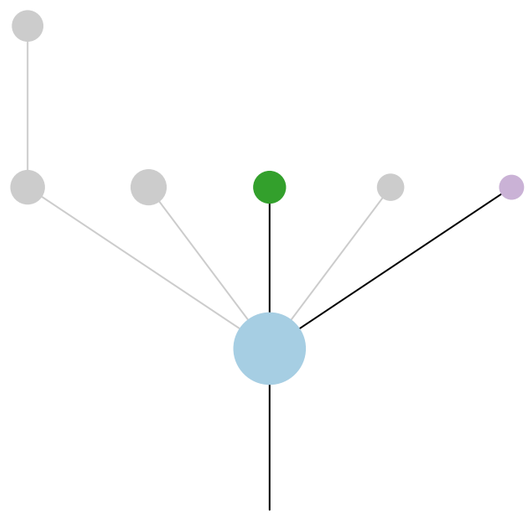
R2



R3



R4



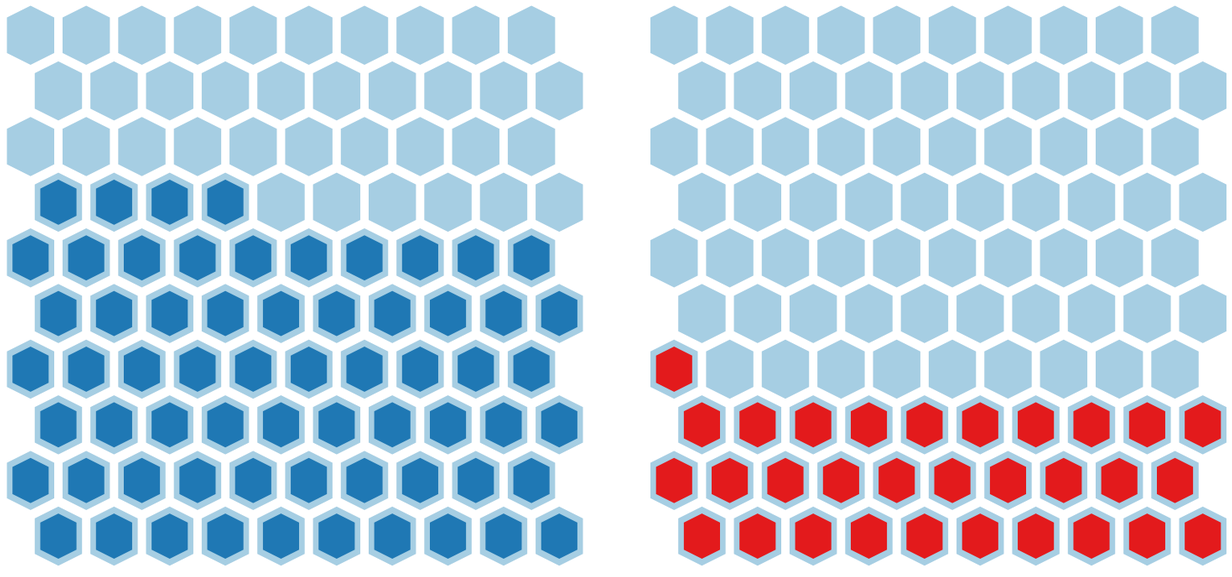
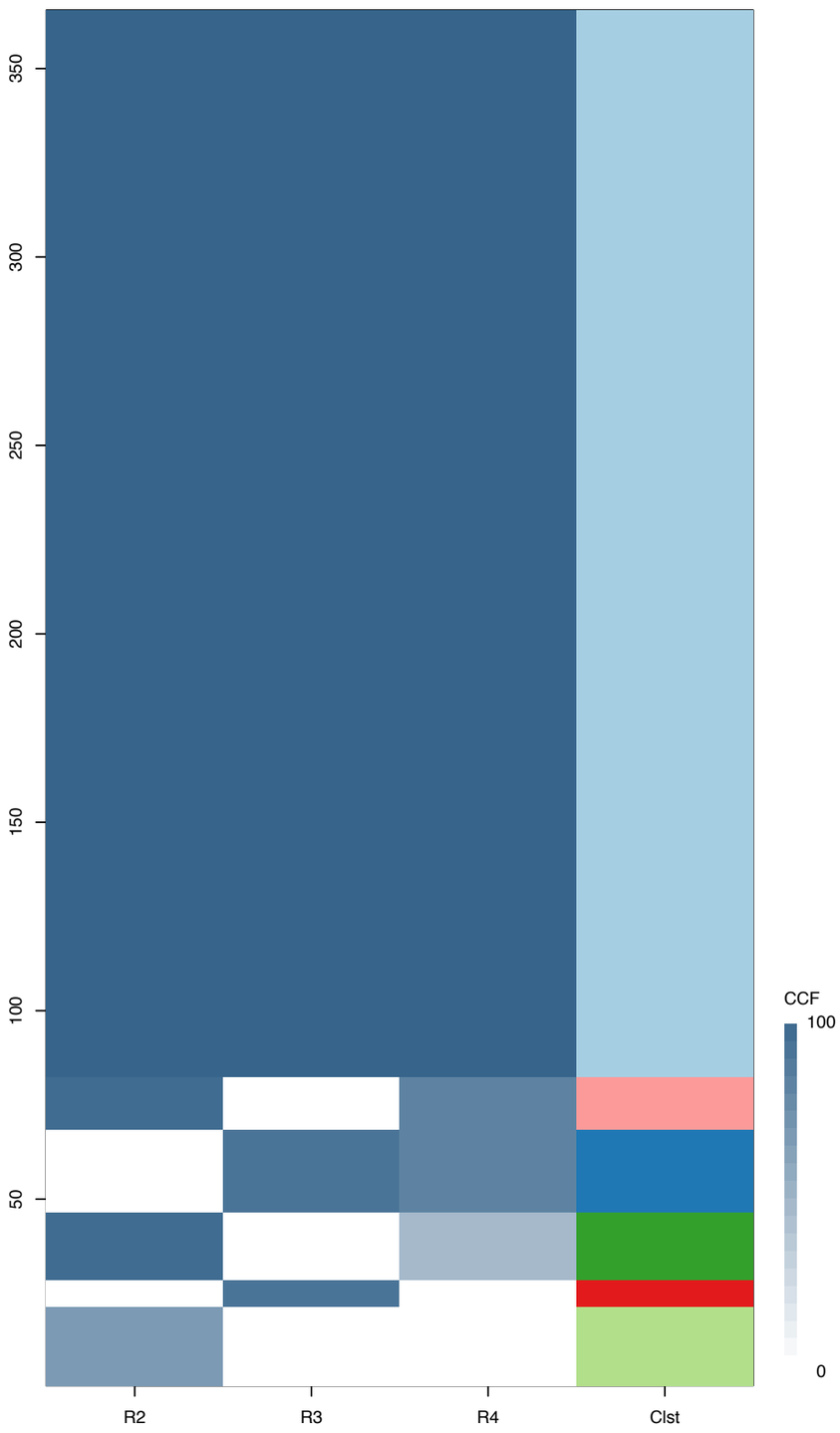
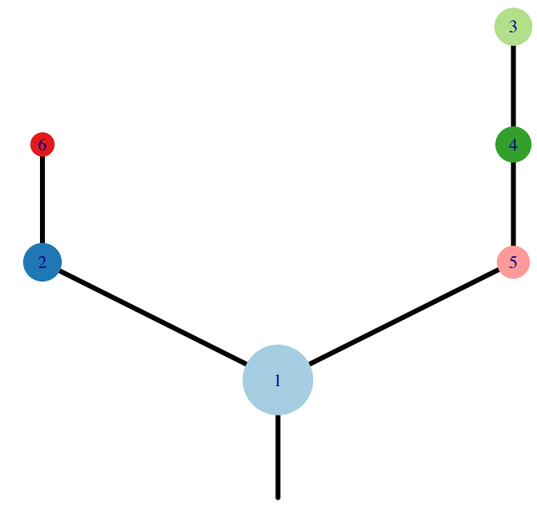


Fig.S12BT



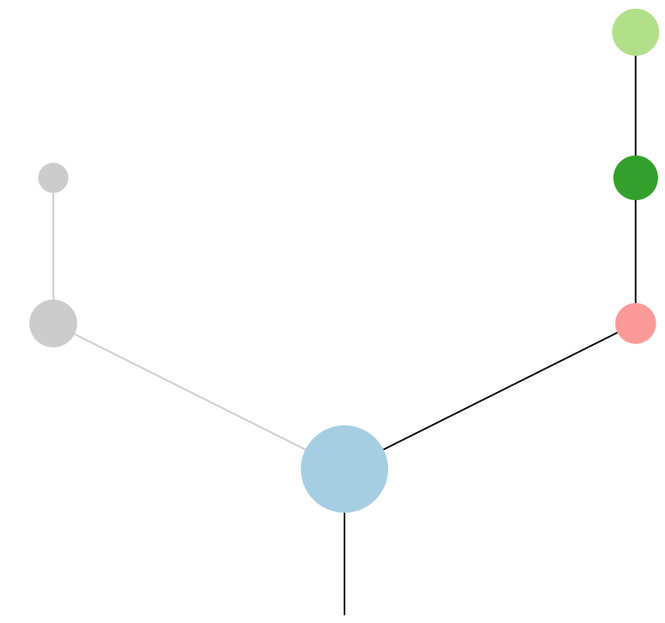
CRUK0078



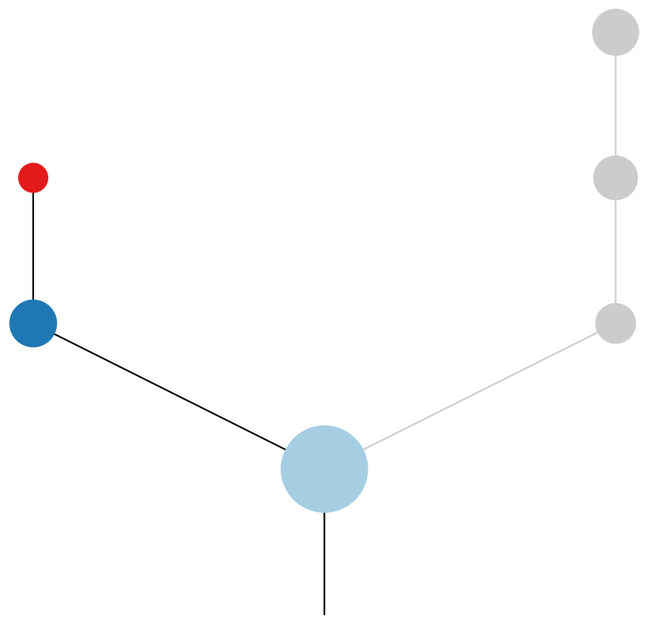
Histology:Squamous, Age:59, PackYears:63, Size:38
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
KAT6A	1	8p11.21	Amp
PTEN	1	10q23.31	SNV
KDM5A	1	12p13.33	Amp
ERC1	1	12p13.33	Amp
SS18L1	1	20q13.33	Amp
PLXNB2	1	22q13.33	SNV
ERCC5	4	13q33.1	SNV

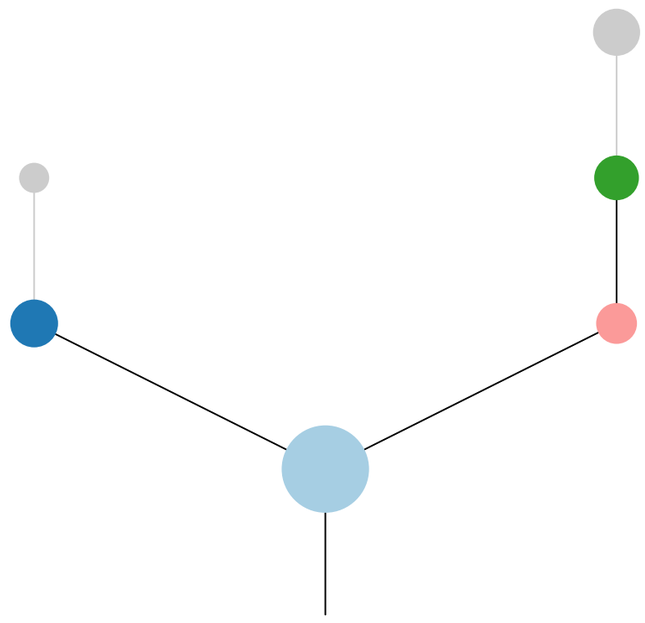
R2

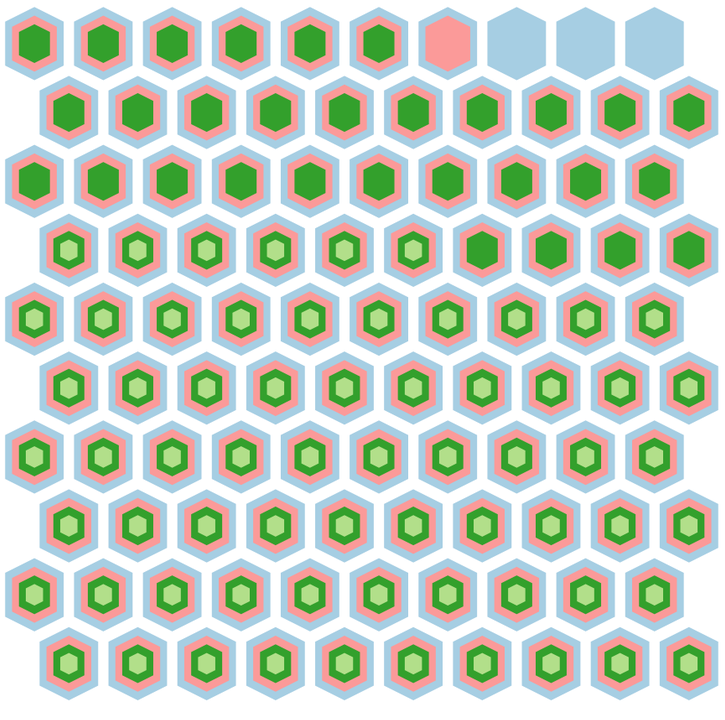


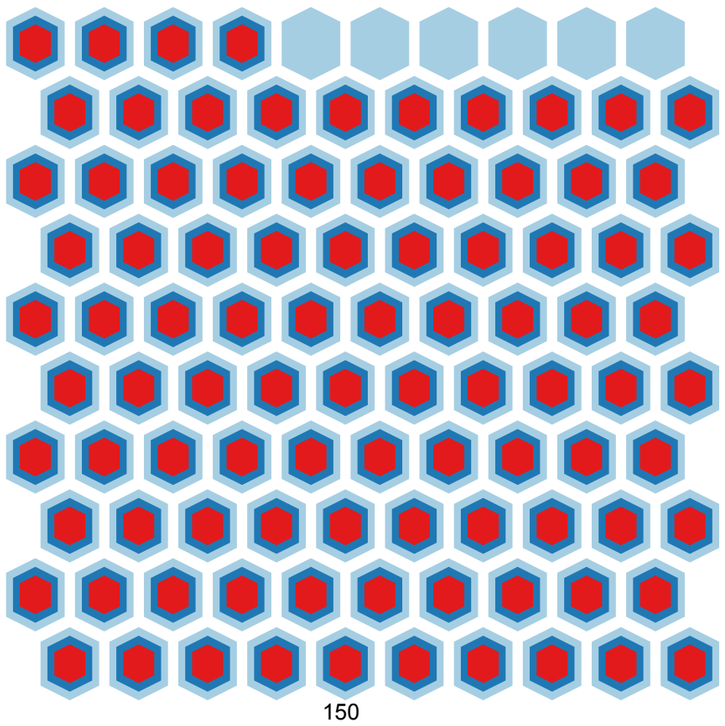
R3



R4







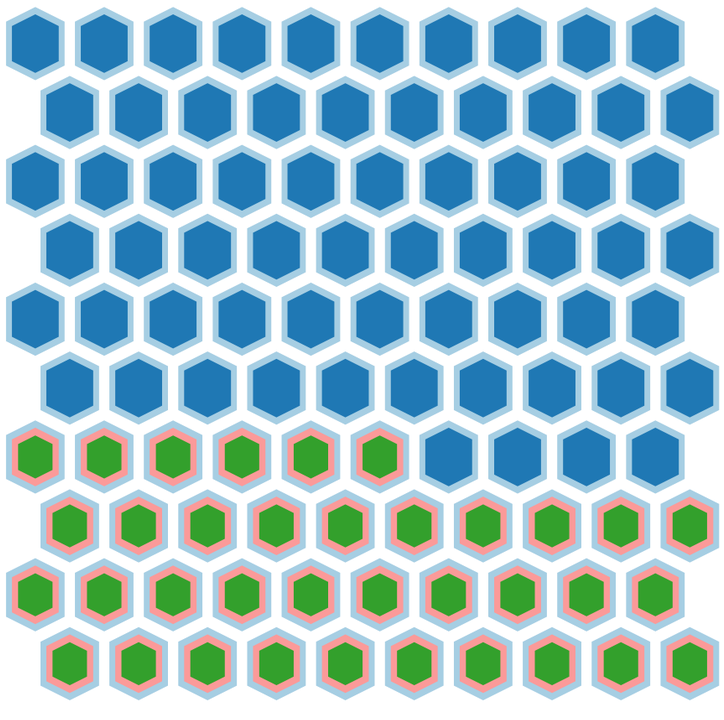
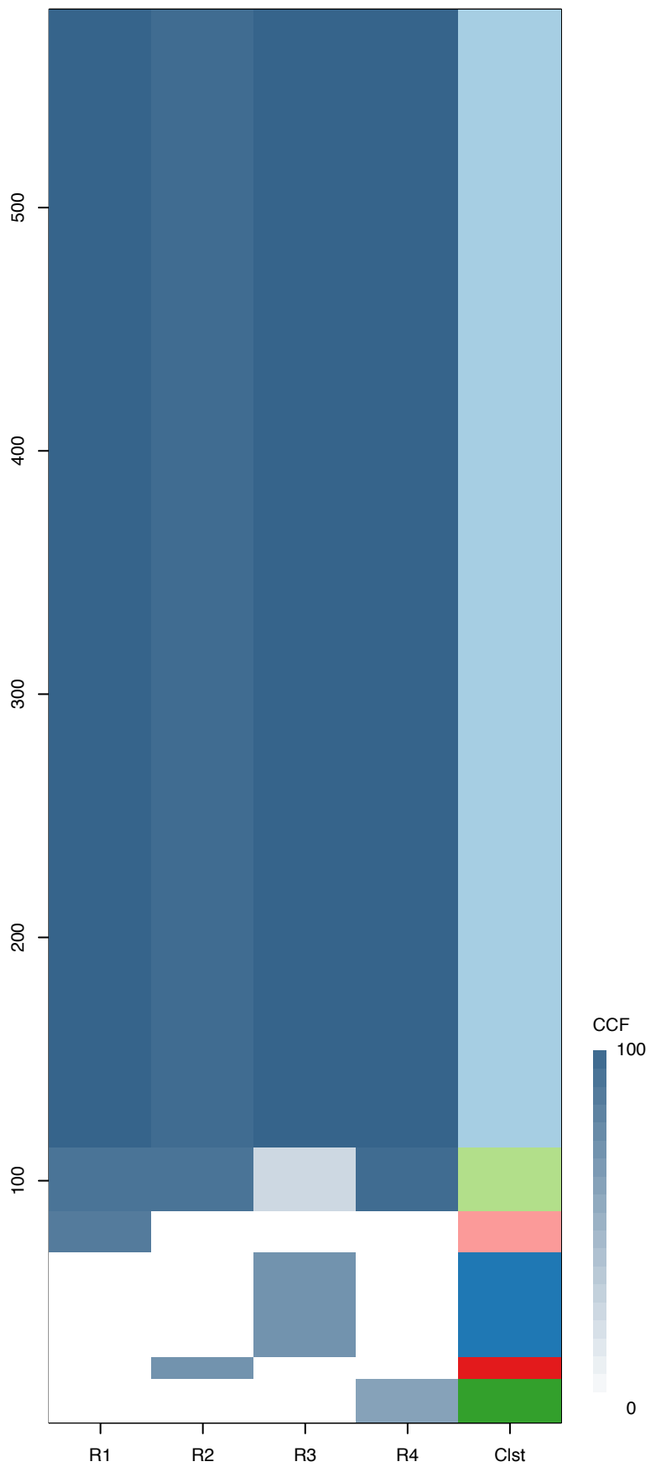
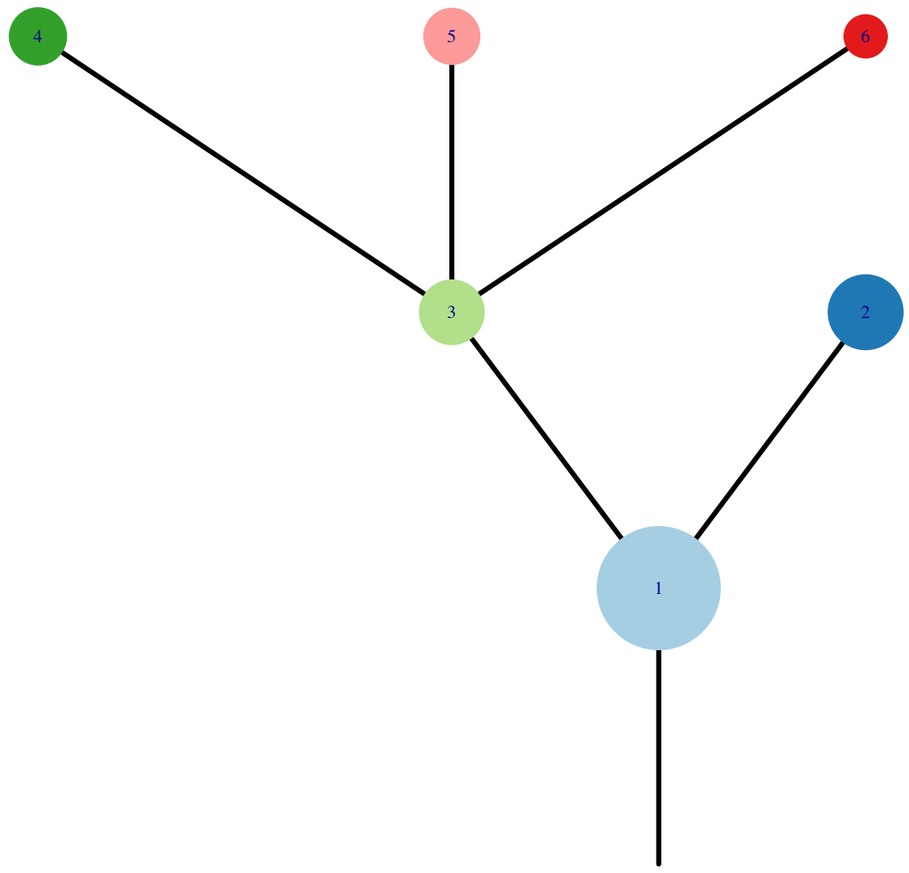


Fig.S12BU



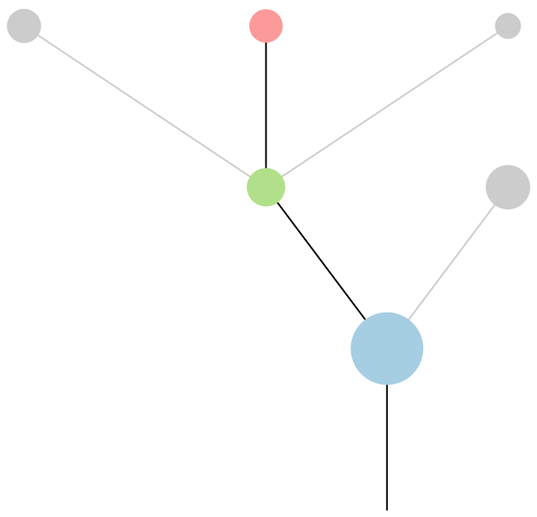
CRUK0079

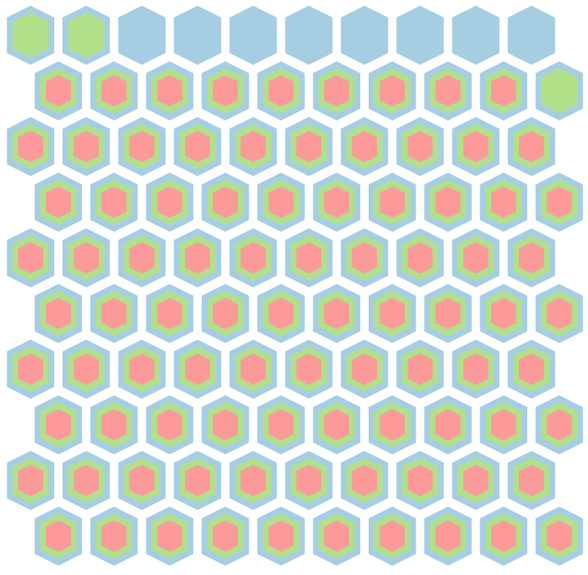


Histology:Squamous, Age:64, PackYears:15, Size:60
Stage:2a, Gender:Female, GD:Clonal GD, Recur:no

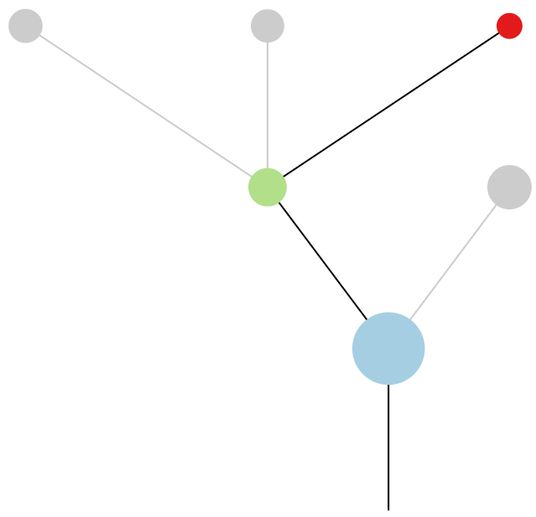
Gene	Cluster	Cytoband	Type
WWTR1	1	3q25.1	Amp
GMPS	1	3q25.31	Amp
MLF1	1	3q25.32	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
FAT1	1	4q35.2	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
KAT6A	1	8p11.21	Amp
IKBKB	1	8p11.21	Amp
HOOK3	1	8p11.21	Amp
POLE	1	12q24.33	SNV
TP53	1	17p13.1	SNV
CCNE1	1	19q12	Amp
CEP89	1	19q13.11	Amp
CEBPA	1	19q13.11	Amp
LSM14A	1	19q13.11	Amp
AKT2	1	19q13.2	Amp
CD79A	1	19q13.2	Amp
ZNF180	1	19q13.31	Amp
BCL3	1	19q13.32	Amp
TNFAIP3	6	6q23.3	SNV

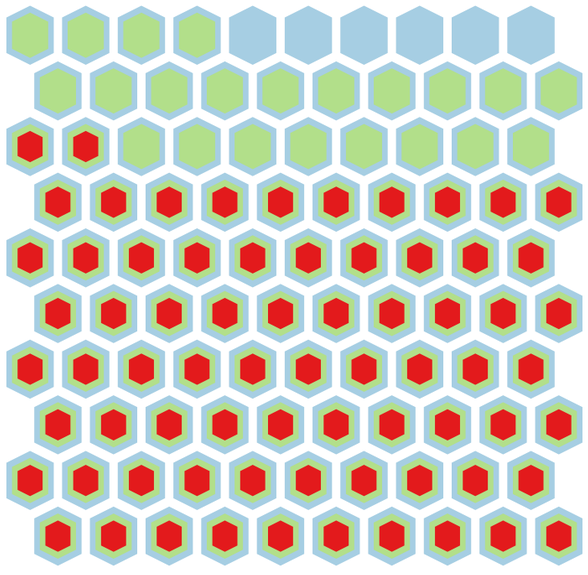
R1



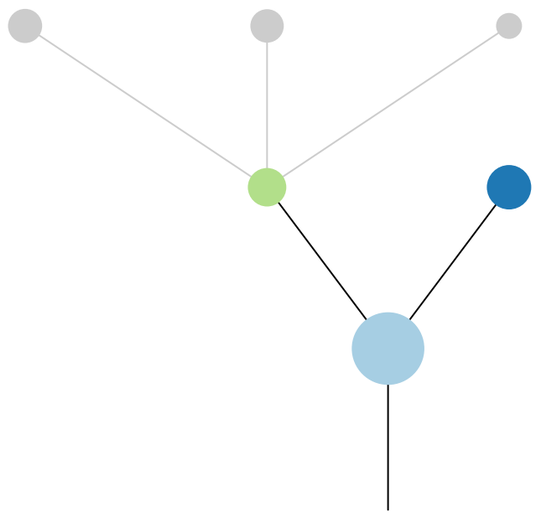


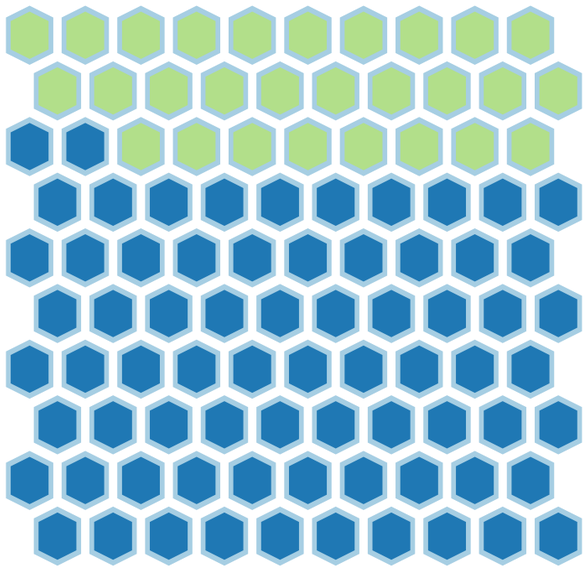
R2



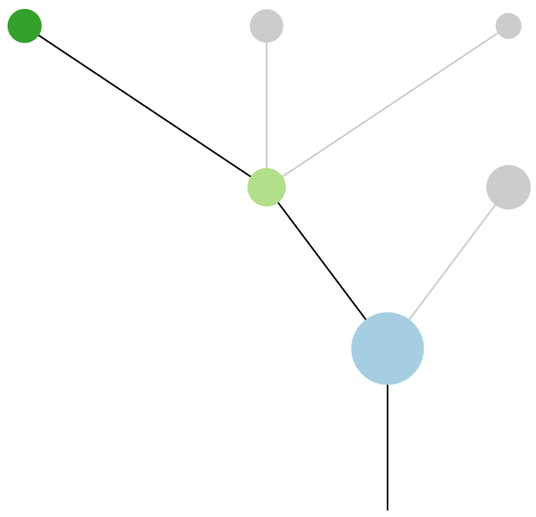


R3





R4



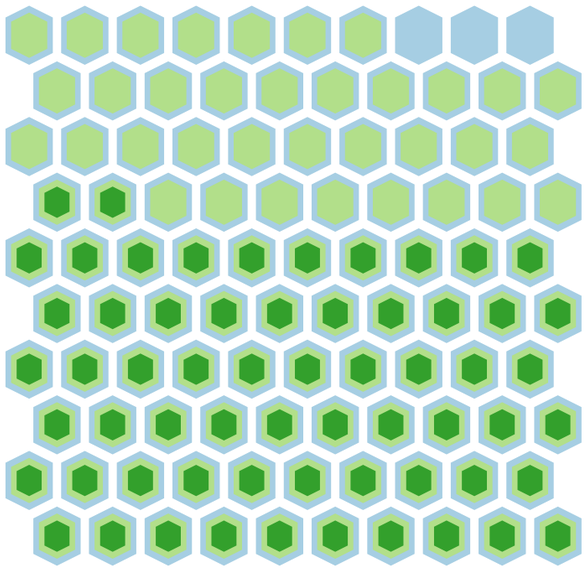
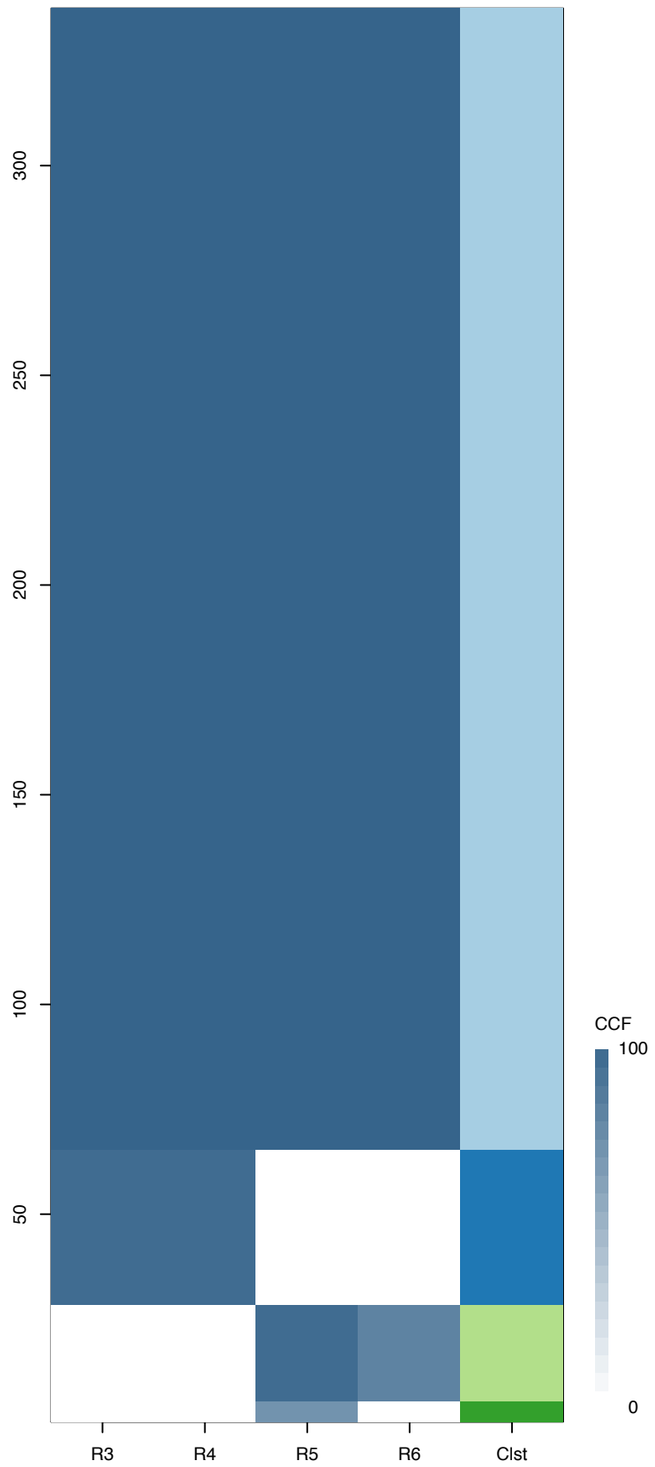
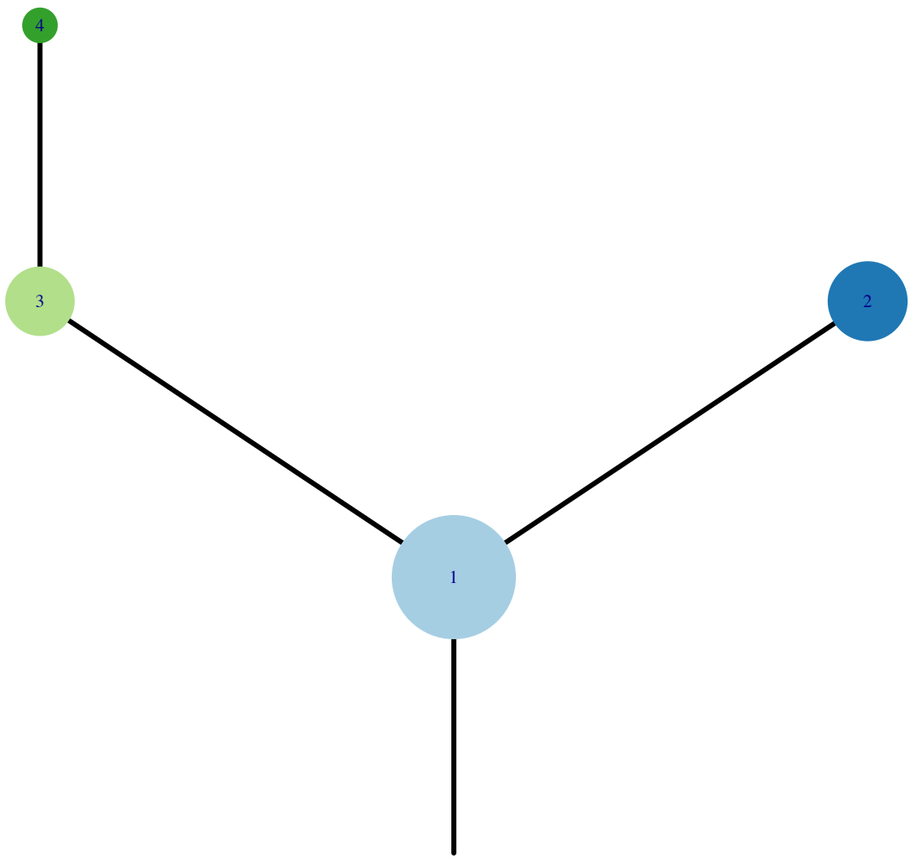


Fig.S12BV



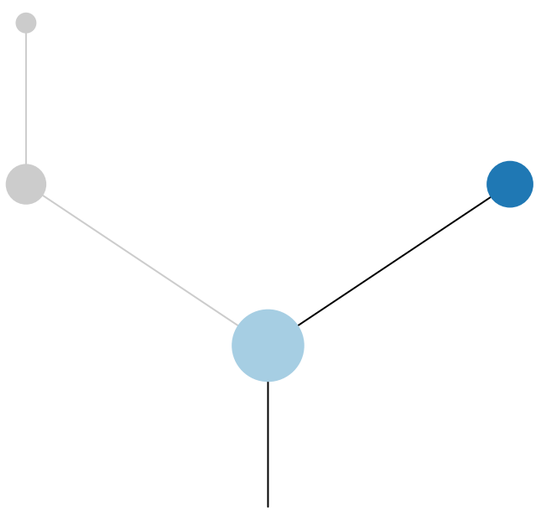
CRUK0080



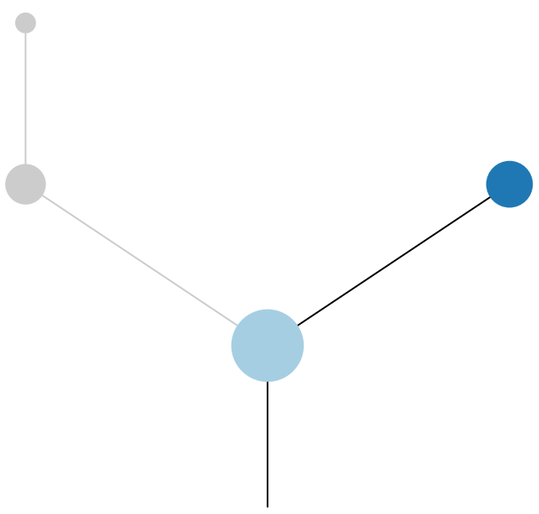
Histology:Squamous, Age:62, PackYears:56.25, Size:37
Stage:2a, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
SRSF3	1	6p21.31	Amp
EGFR	1	7p11.2	Amp
WT1	1	11p13	SNV
CCND1	1	11q13.3	Amp
NUMA1	1	11q13.4	Amp
BIRC3	1	11q22.2	Amp
TP53	1	17p13.1	SNV
IKZF1	3	7p12.2	SNV
CALR	3	19p13.2	Amp
LYL1	3	19p13.2	Amp

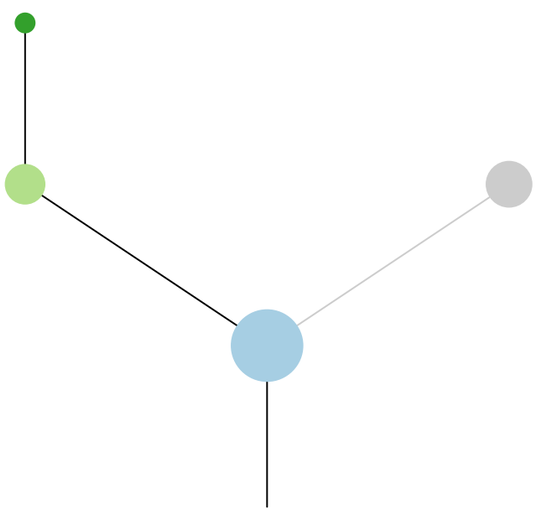
R3



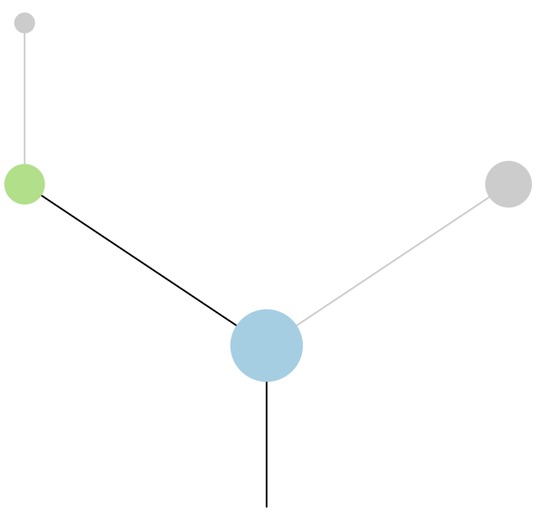
R4



R5



R6



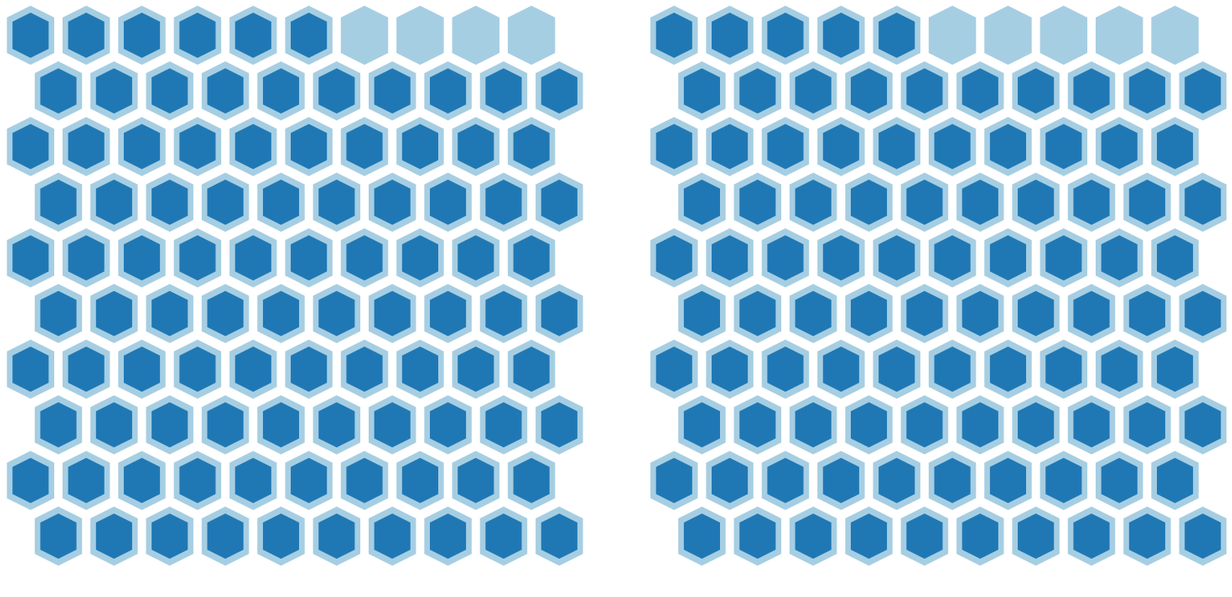
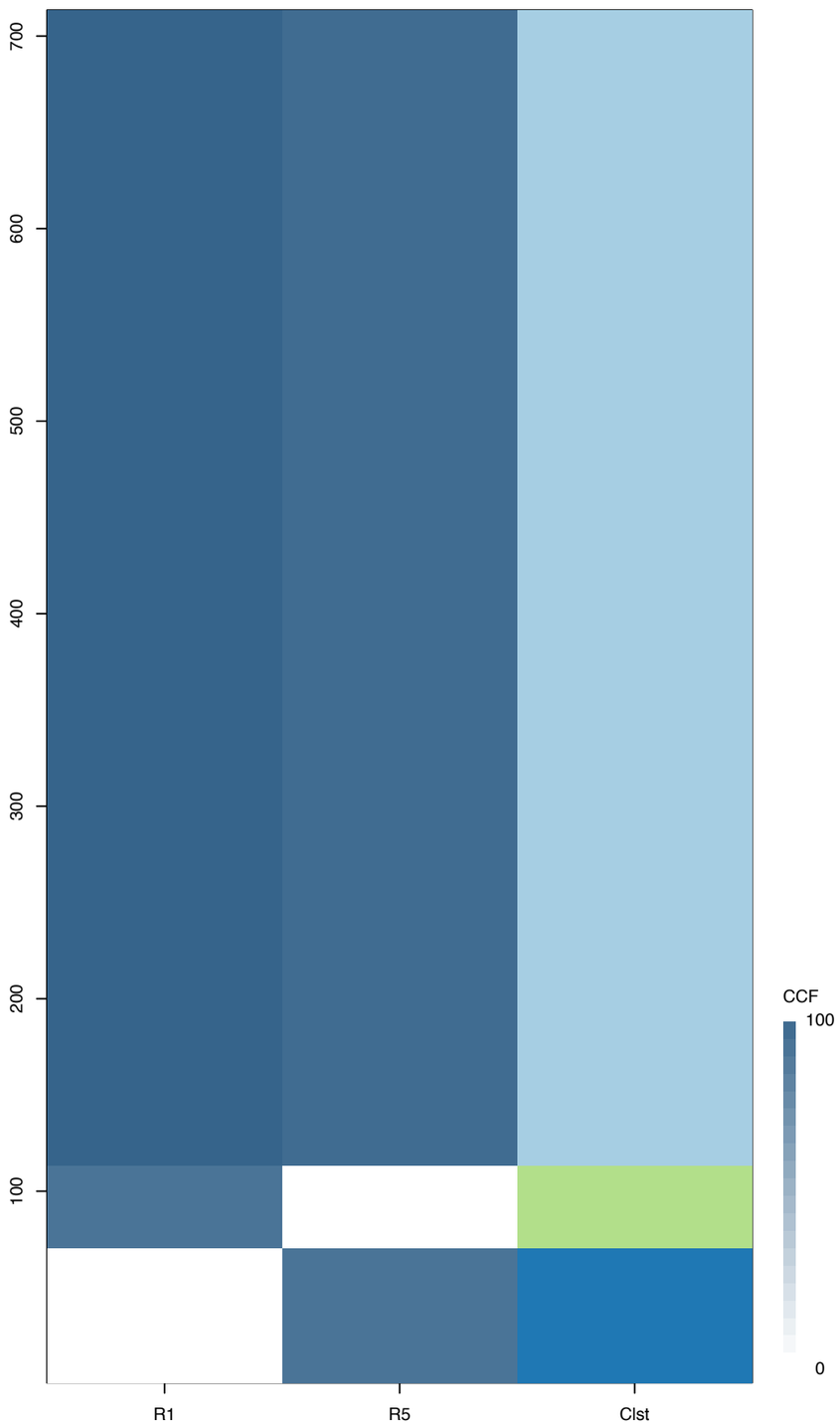
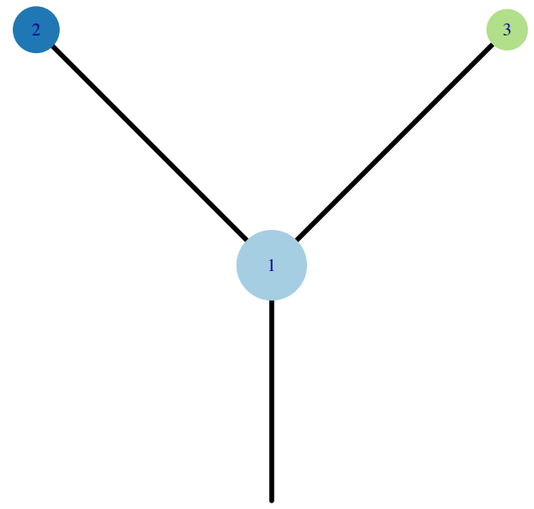


Fig.S12BW



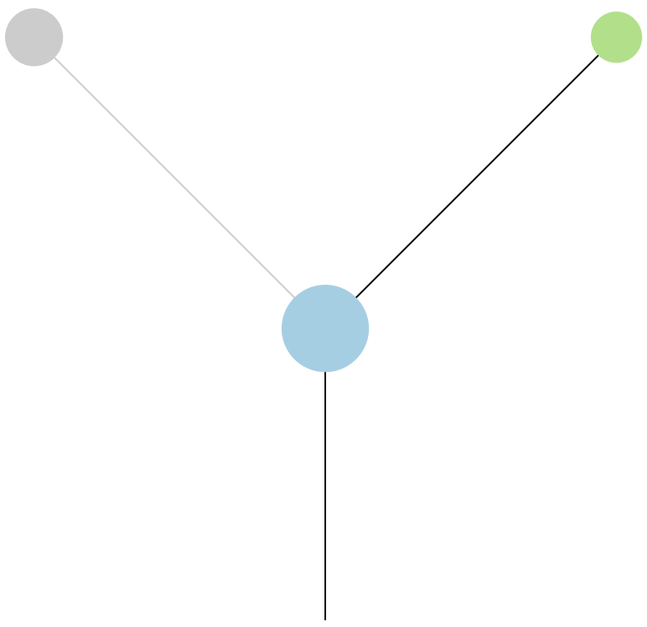
CRUK0081



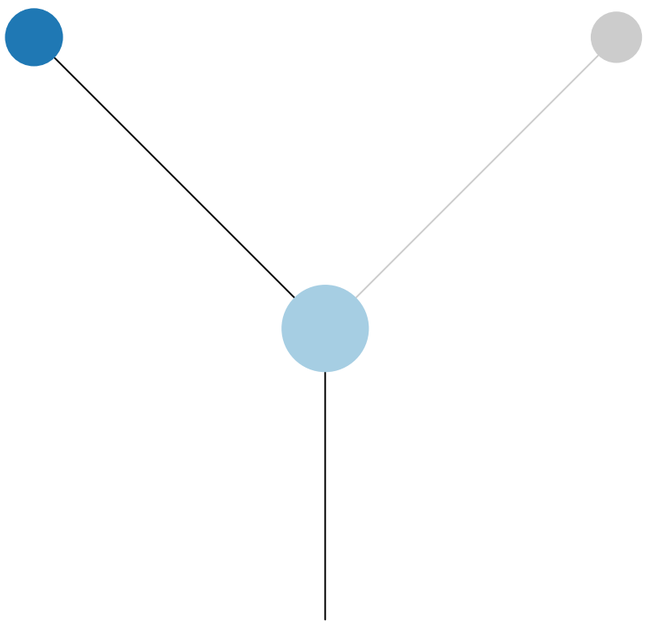
Gene	Cluster	Cytoband	Type
FAT1	1	4q35.2	SNV
CDKN2A	1	9p21.3	SNV
FANCC	1	9q22.32	SNV
NOTCH1	1	9q34.3	SNV
TP53	1	17p13.1	SNV
CYLD	3	16q12.1	SNV

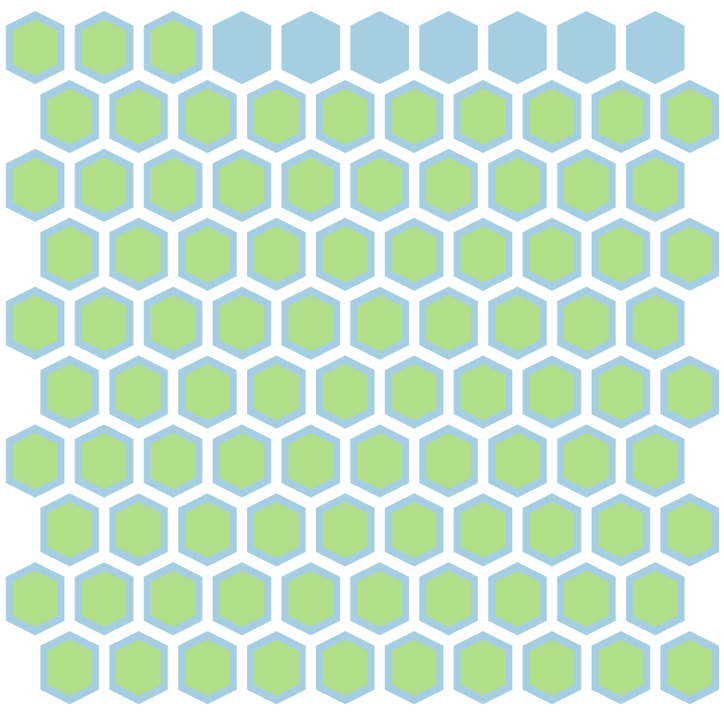
Histology:Squamous, Age:74, PackYears:13.75, Size:60
Stage:2a, Gender:Male, GD:Clonal GD, Recur:yes

R1



R5





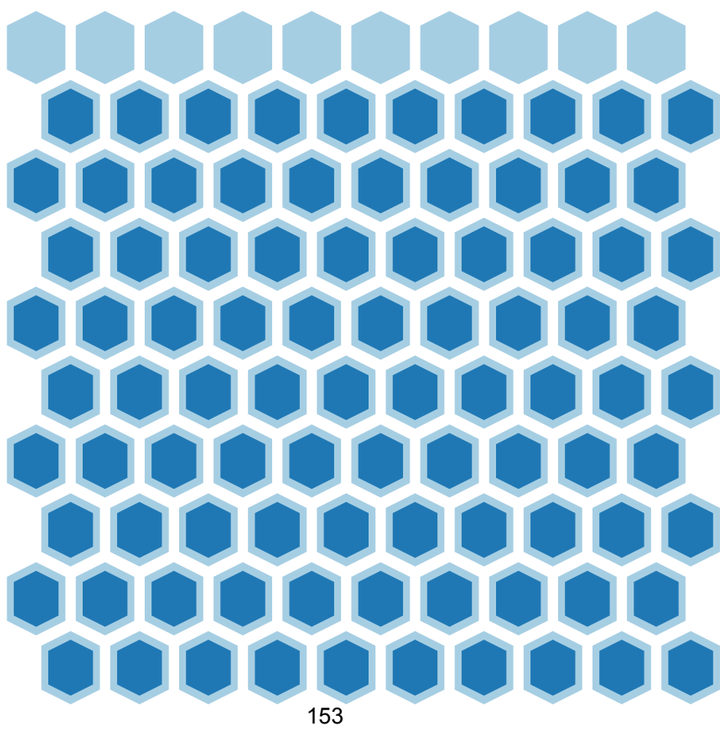
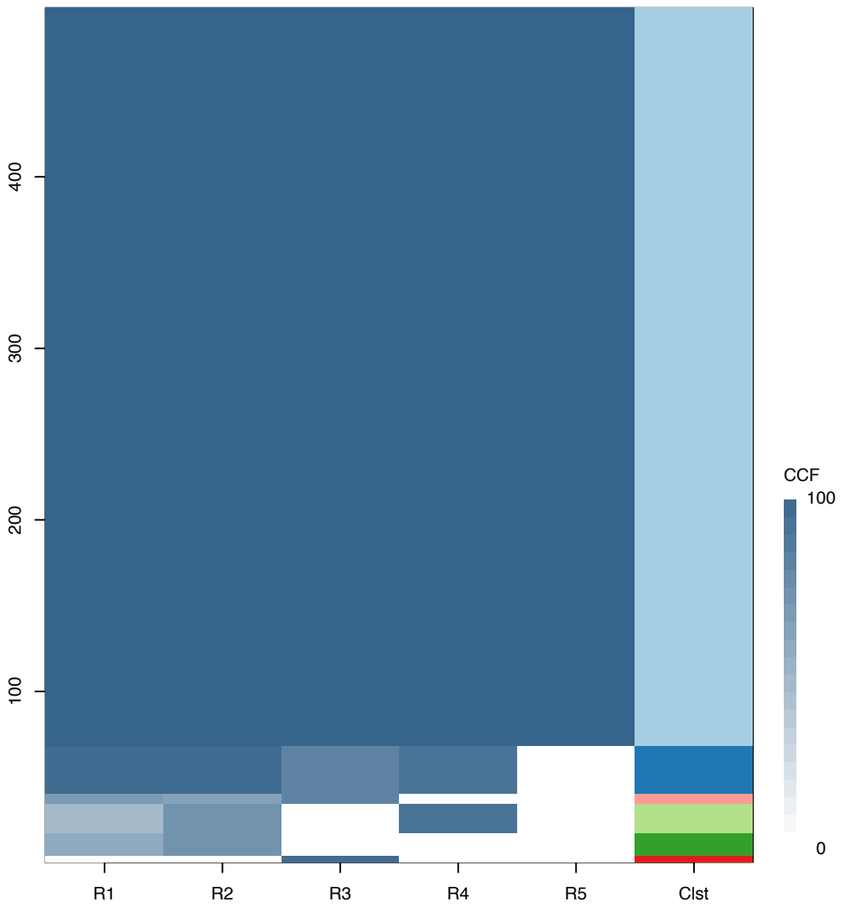
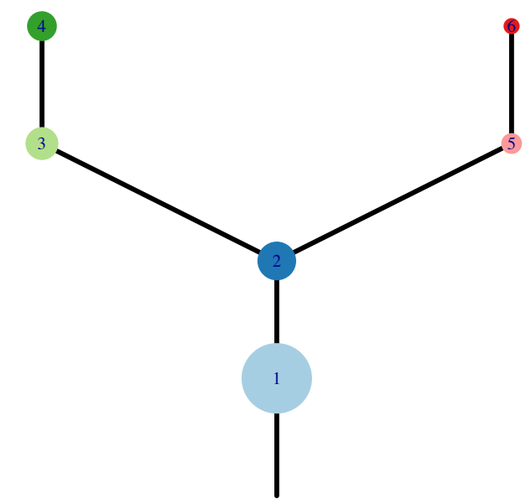


Fig.S12BX



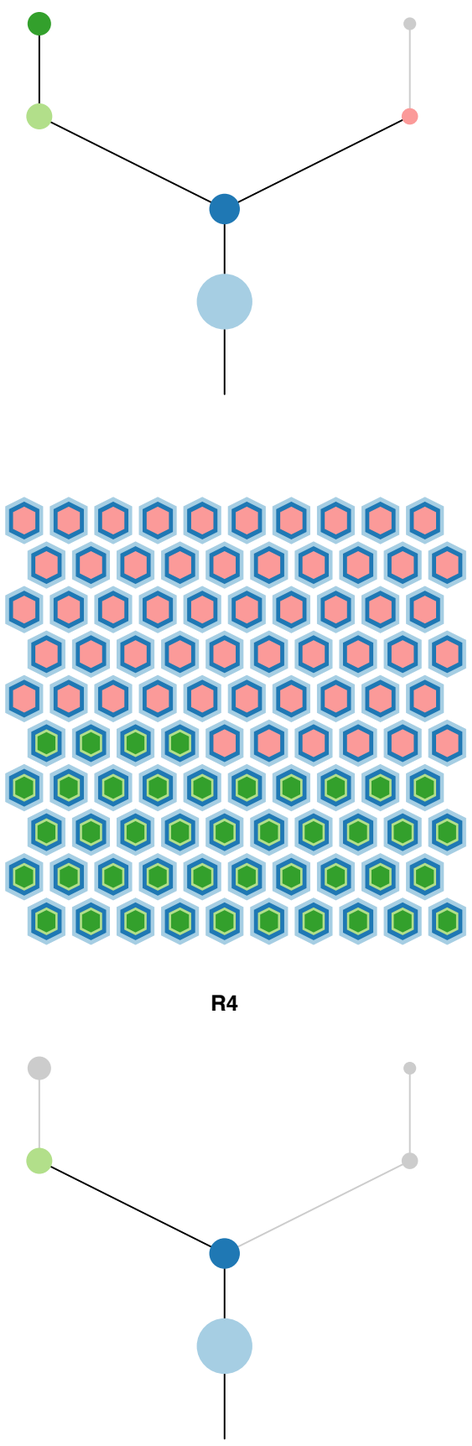
CRUK0082



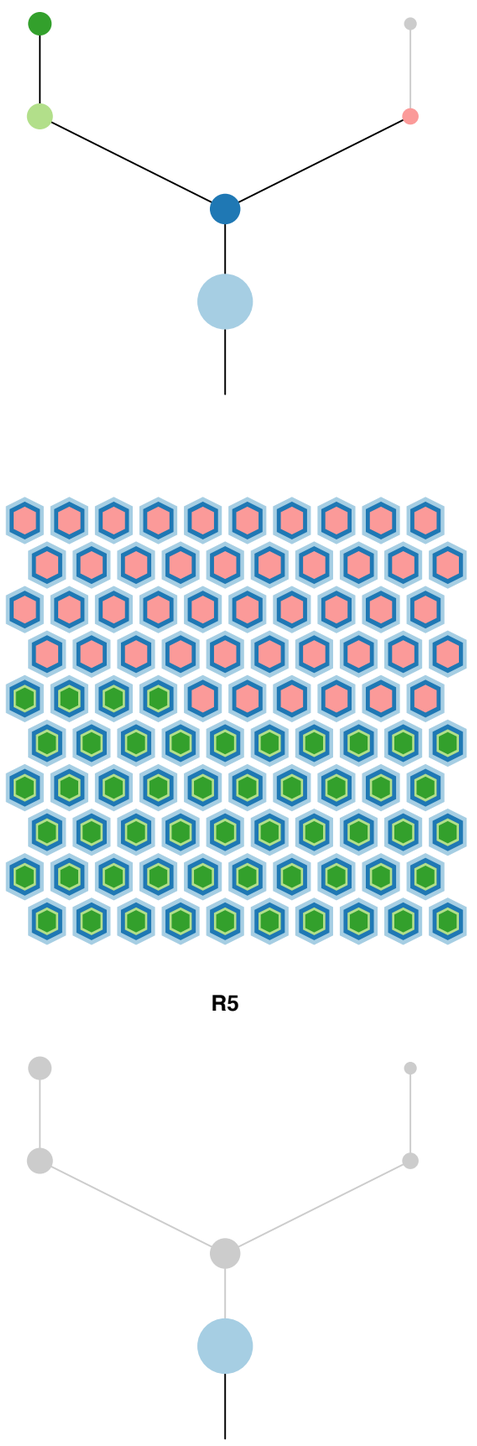
Histology:Squamous, Age:65, PackYears:NA, Size:115
Stage:3a, Gender:Female, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
COL5A2	1	2q32.2	SNV
TFG	1	3q12.2	Amp
GATA2	1	3q21.3	Amp
RPN1	1	3q21.3	Amp
CNBP	1	3q21.3	Amp
FOXL2	1	3q22.3	Amp
WWTR1	1	3q25.1	Amp
GMPS	1	3q25.31	Amp
MLF1	1	3q25.32	Amp
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
PTEN	1	10q23.31	SNV
WT1	1	11p13	SNV
KMT2D	1	12q13.12	SNV
TP53	1	17p13.1	SNV
FSTL3	1	19p13.3	Amp
DNM2	6	19p13.2	SNV
PTPRB	?	12q15	SNV

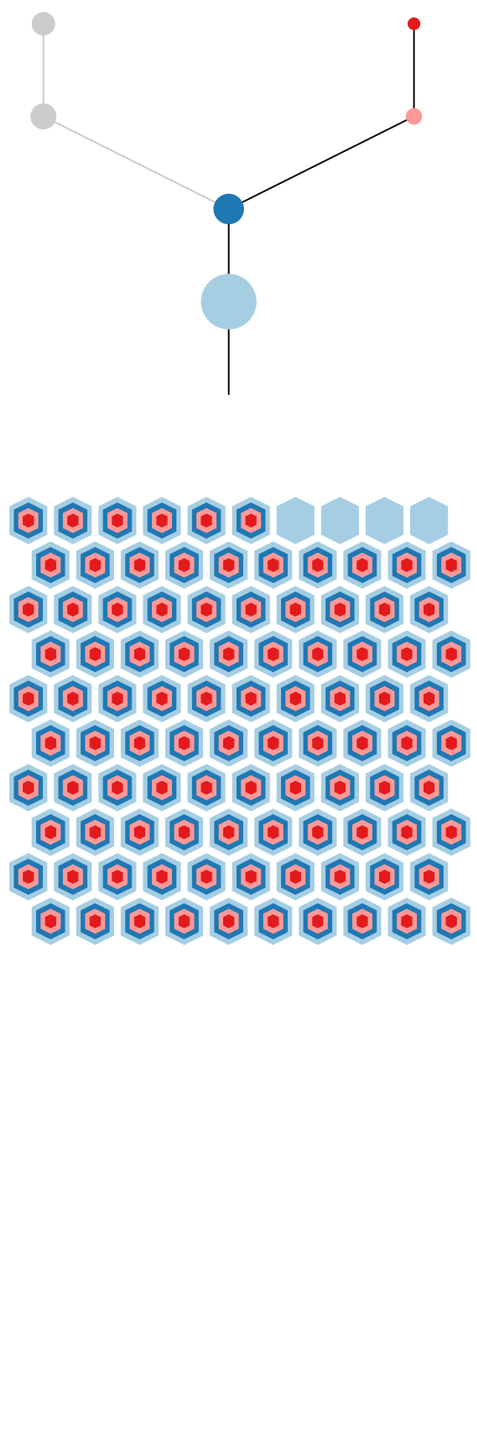
R1



R2



R3

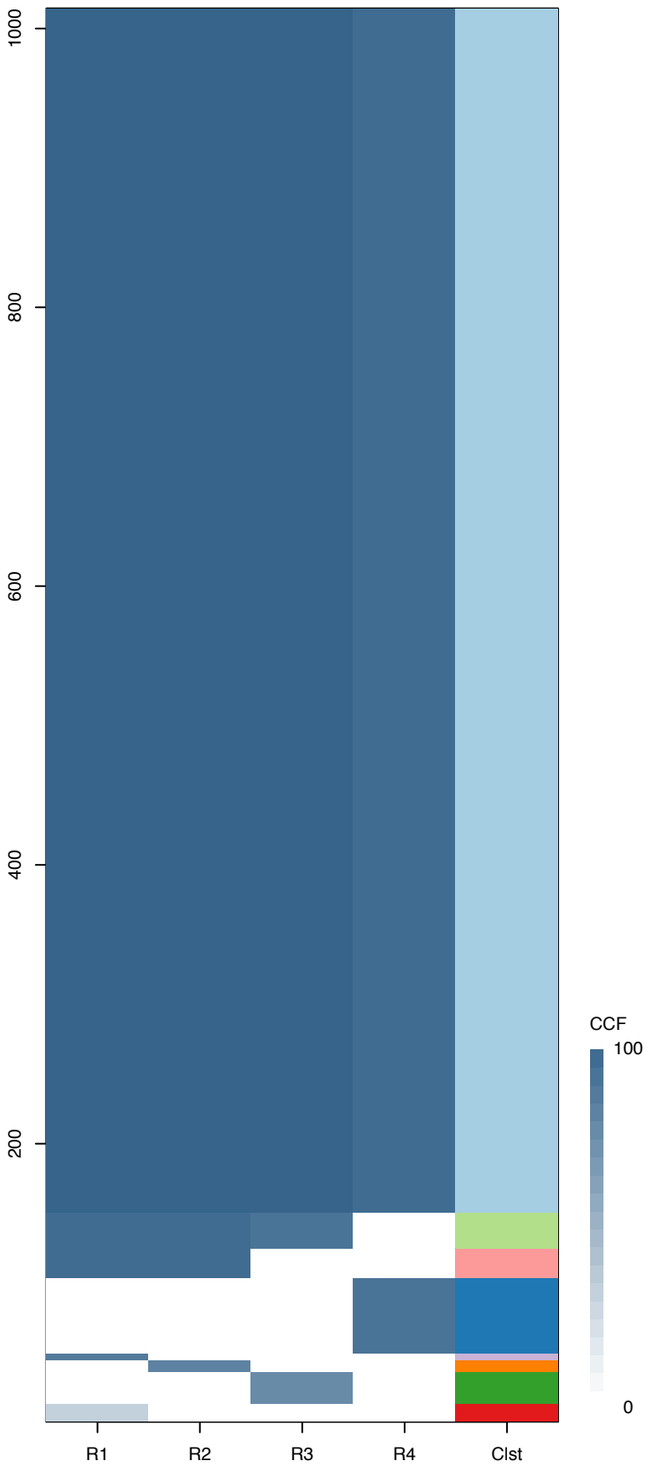


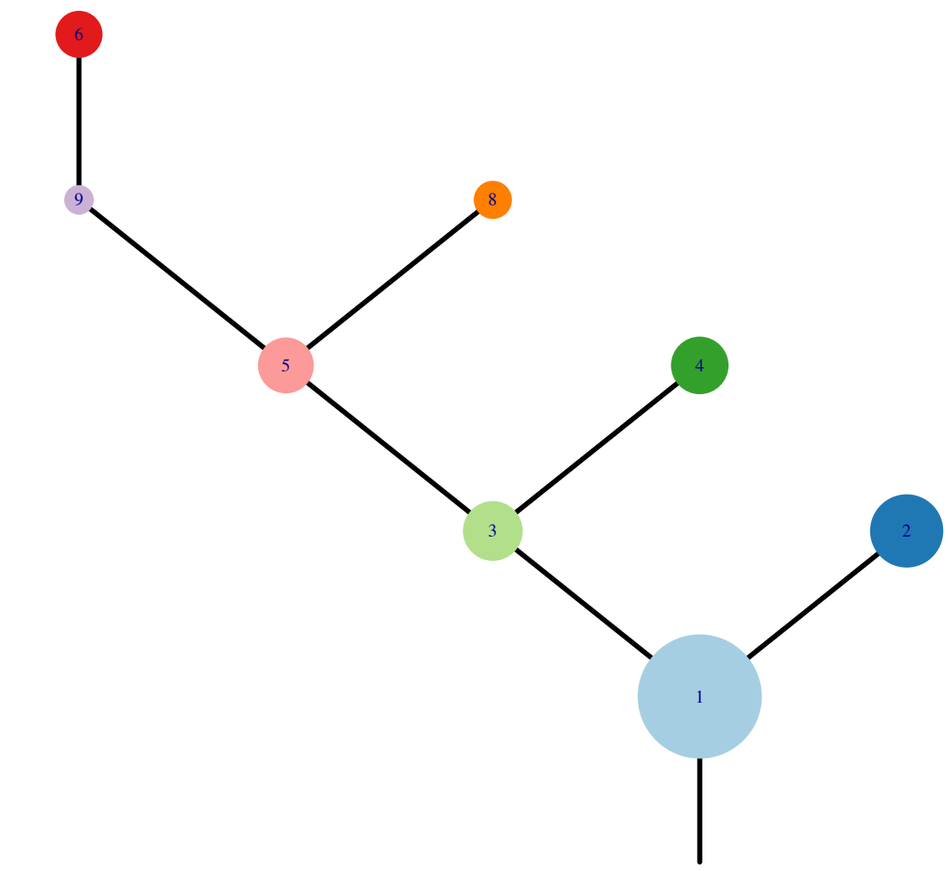
R4

R5

Fig.S12BY

CRUK0083

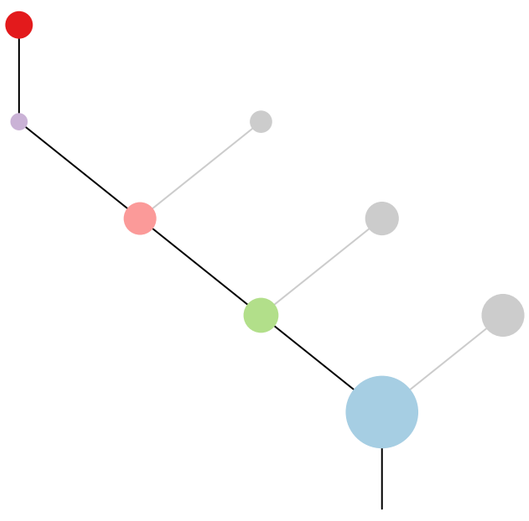




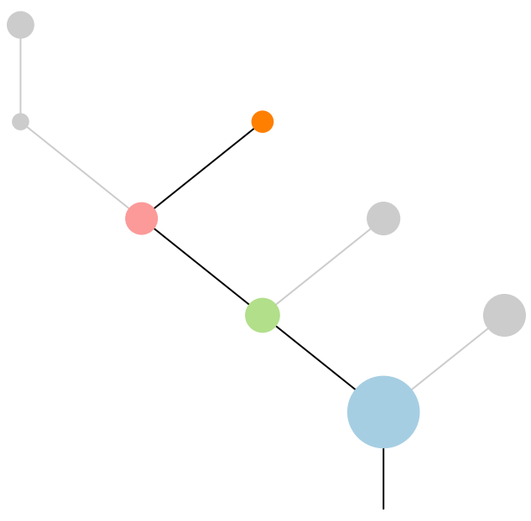
Histology:Squamous, Age:73, PackYears:50, Size:39
Stage:2b, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
WWTR1	1	3q25.1	Amp
GMPS	1	3q25.31	Amp
MLF1	1	3q25.32	Amp
MECOM	1	3q26.2	Amp
PIK3CA	1	3q26.32	Amp
SOX2	1	3q26.33	Amp
ETV5	1	3q27.2	Amp
EIF4A2	1	3q27.3	Amp
BCL6	1	3q27.3	Amp
LPP	1	3q27.3	Amp
TFRC	1	3q29	Amp
FBXW7	1	4q31.3	SNV
RASA1	1	5q14.3	SNV
WHSC1L1	1	8p11.23	Amp
FGFR1	1	8p11.23	Amp
KAT6A	1	8p11.21	Amp
IKBKB	1	8p11.21	Amp
HOOK3	1	8p11.21	Amp
MYC	1	8q24.21	Amp
NDRG1	1	8q24.22	Amp
TP53	1	17p13.1	SNV
CCNB1IP1	2	14q11.2	Amp
NKX2−1	2	14q13.3	Amp
FOXA1	2	14q21.1	Amp
CCNB1IP1	3	14q11.2	Amp
NKX2−1	3	14q13.3	Amp
FOXA1	3	14q21.1	Amp
PIK3CA	?	3q26.32	SNV
FAT1	?	4q35.2	SNV

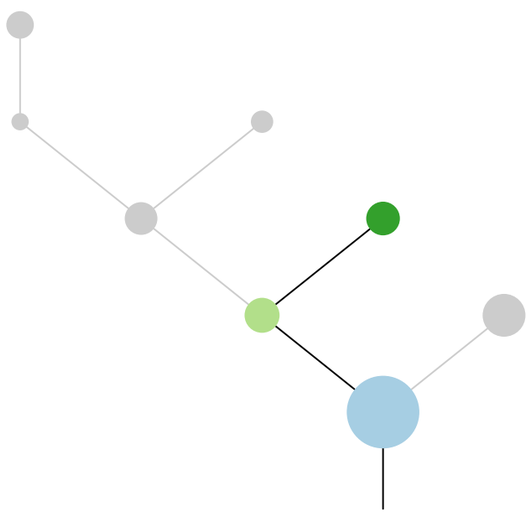
R1



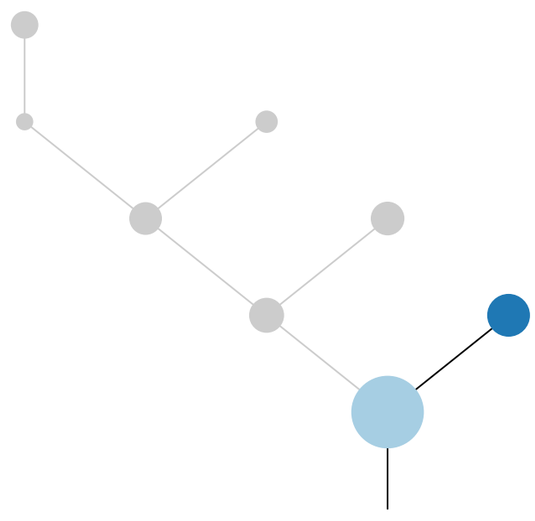
R2



R3



R4



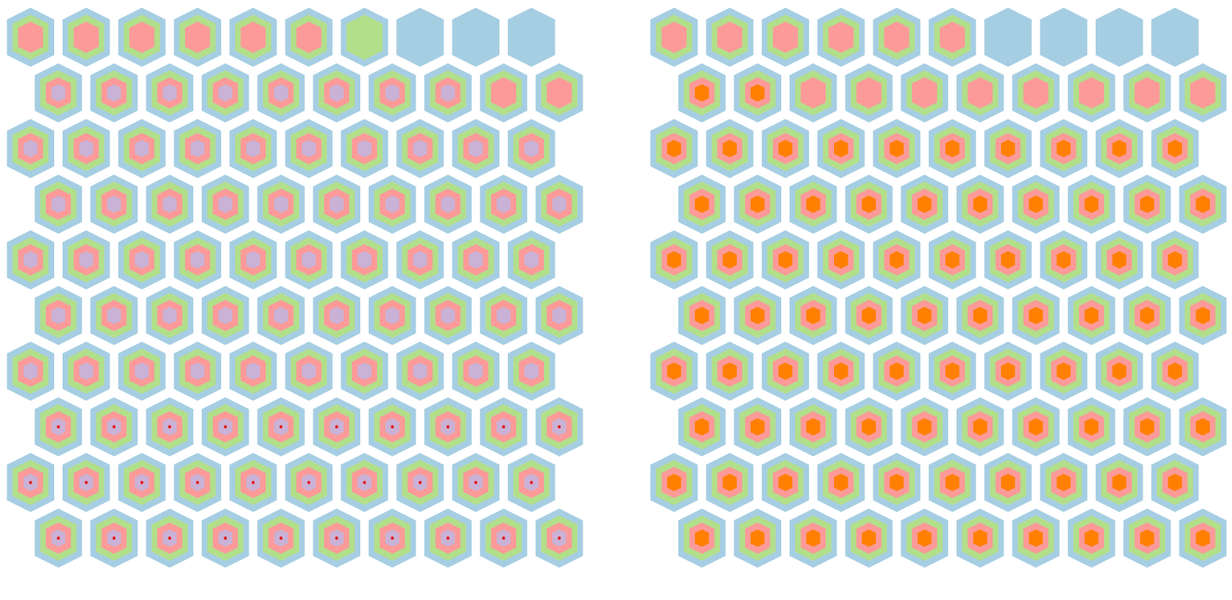
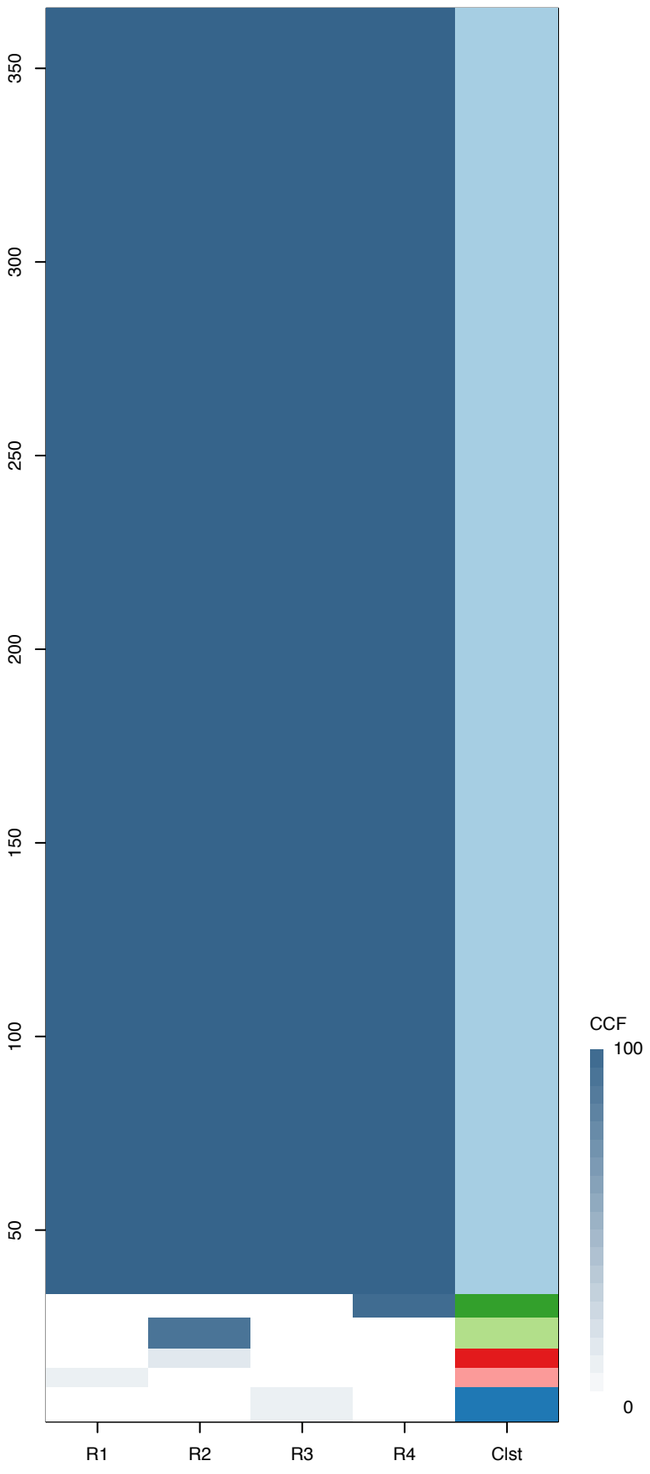
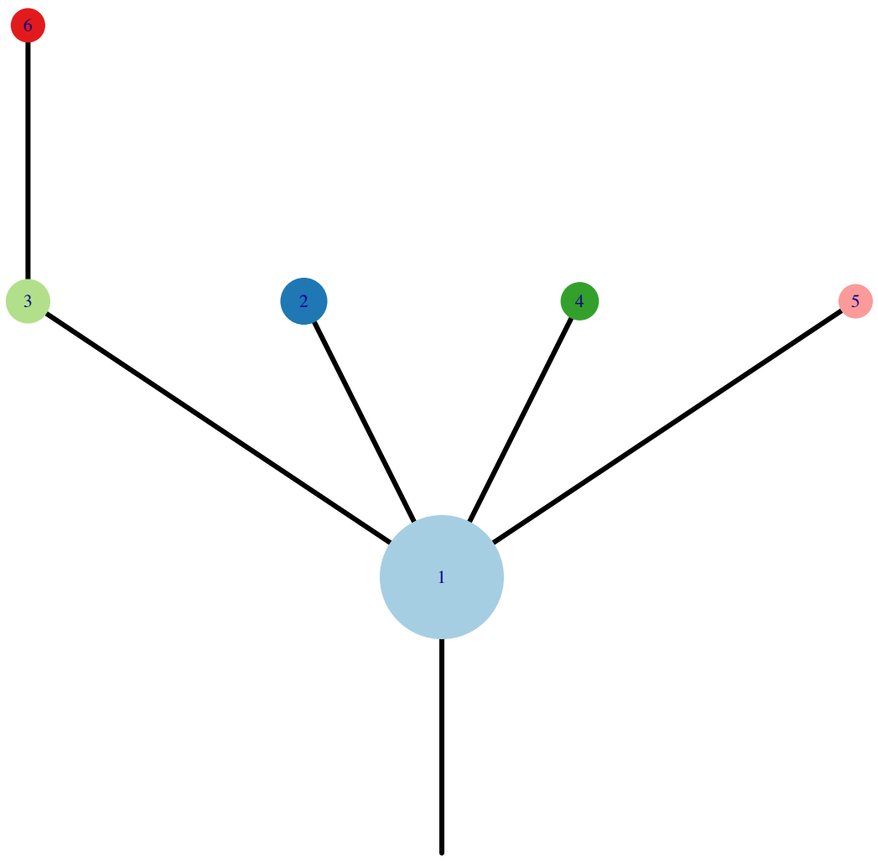


Fig.S12BZ

CRUK0084_A

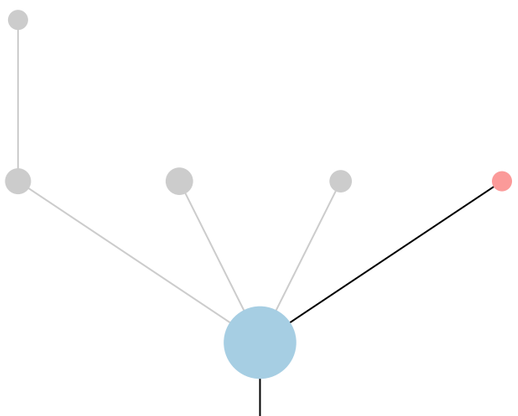




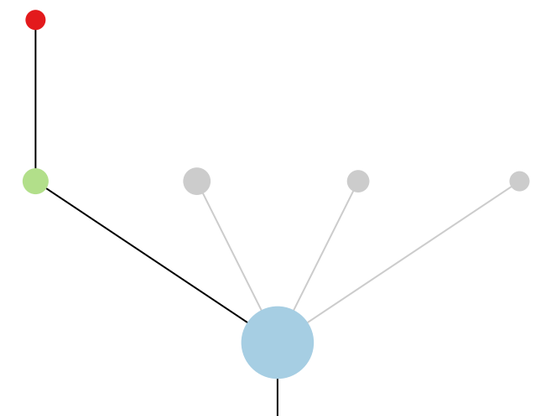
Histology:Squamous, Age:66, PackYears:32.25, Size:32
Stage:1b, Gender:Female, GD:Not GD, Recur:no

Gene	Cluster	Cytoband	Type
CREBBP	1	16p13.3	SNV
KDM5C	?		SNV

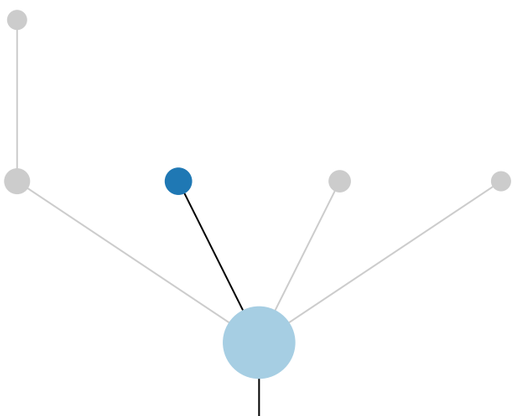
R1



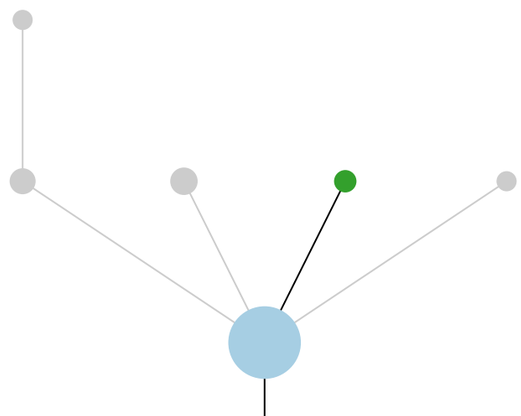
R2



R3



R4



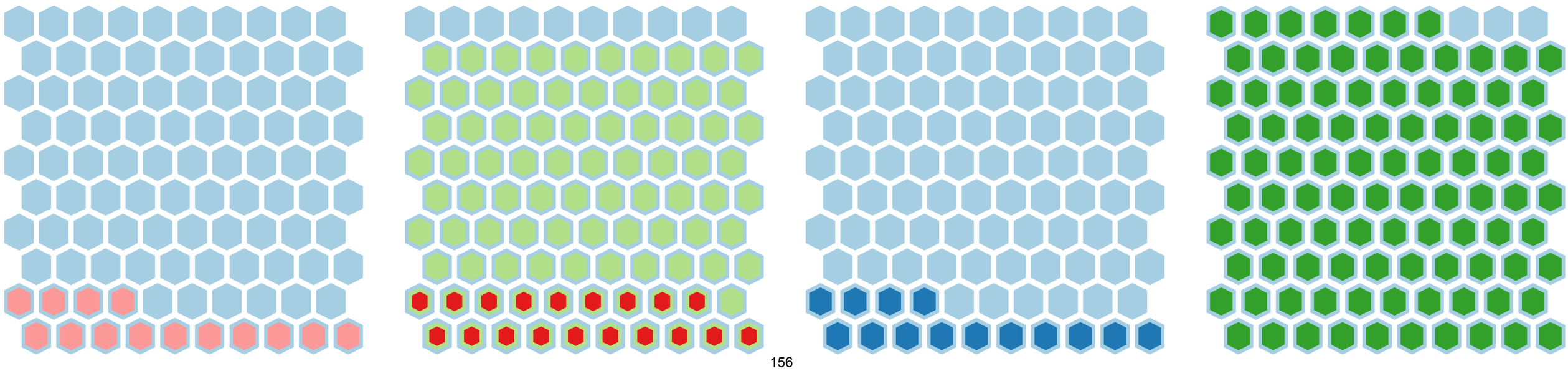
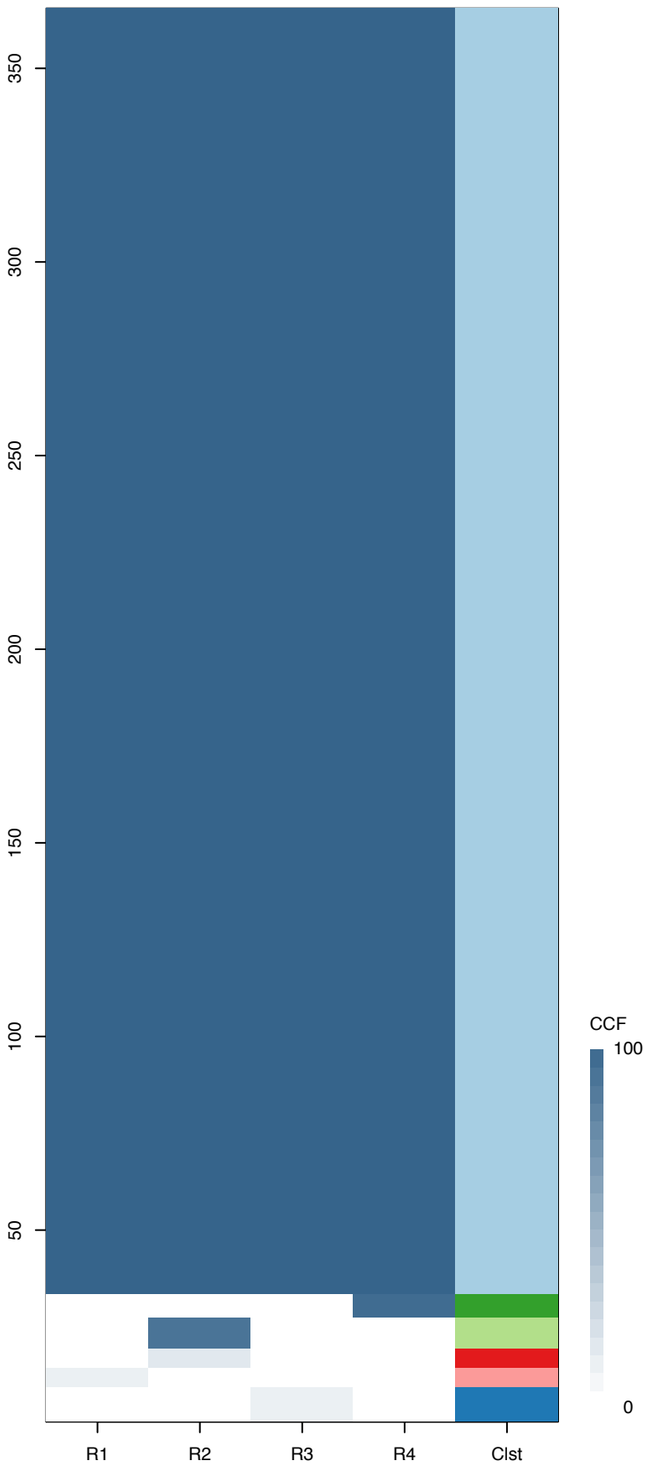
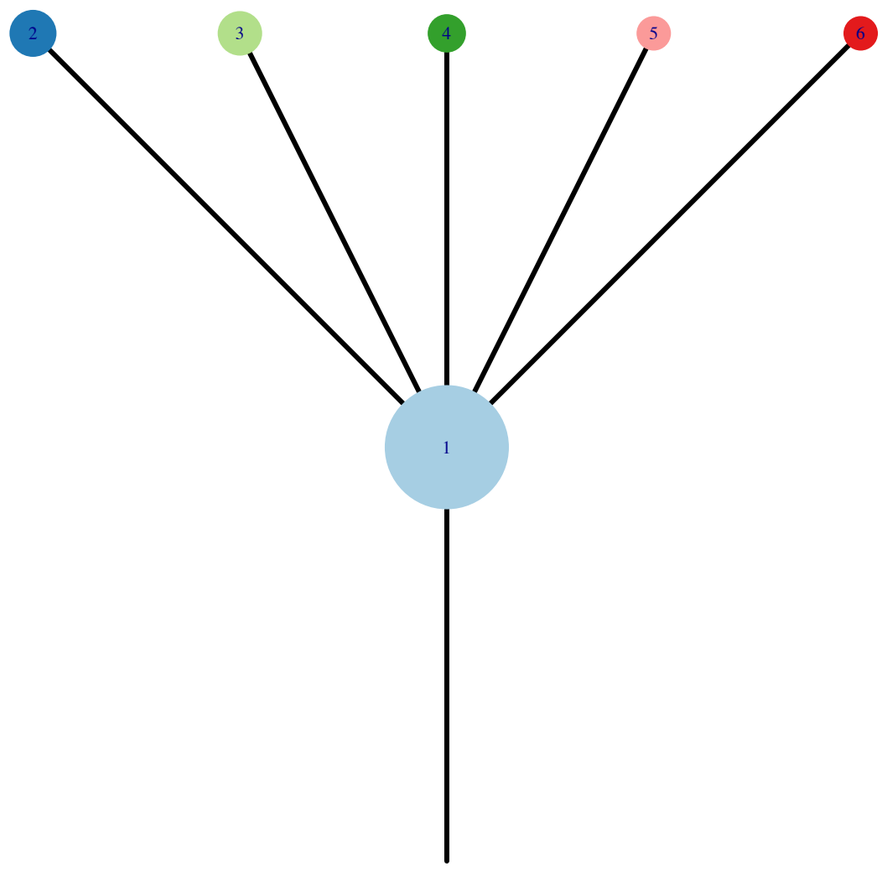


Fig.S12BZ



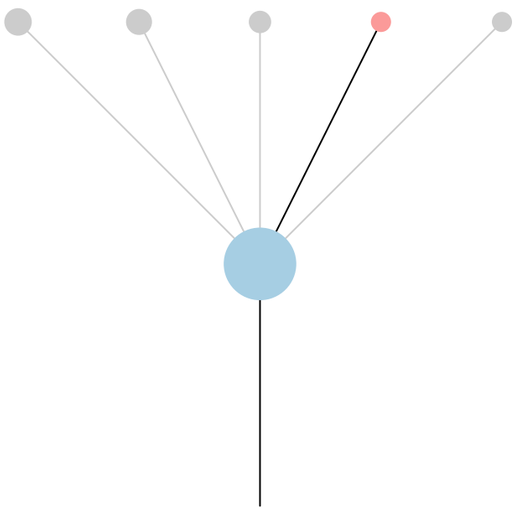
CRUK0084_B



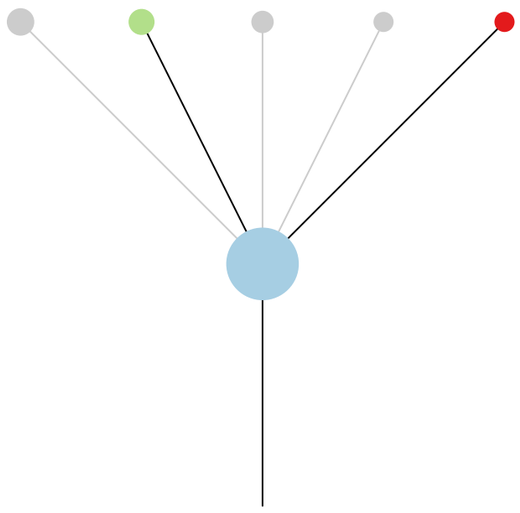
Histology:Squamous, Age:66, PackYears:32.25, Size:32
Stage:1b, Gender:Female, GD:Not GD, Recur:no

Gene	Cluster	Cytoband	Type
CREBBP	1	16p13.3	SNV
KDM5C	?		SNV

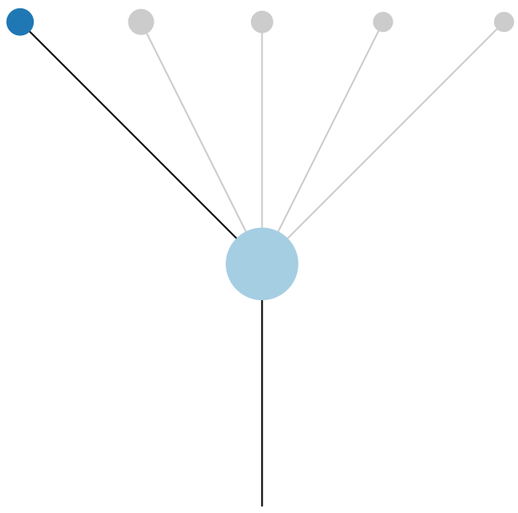
R1



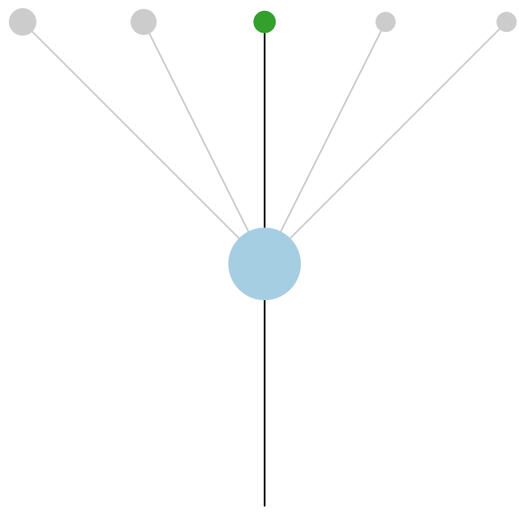
R2



R3



R4



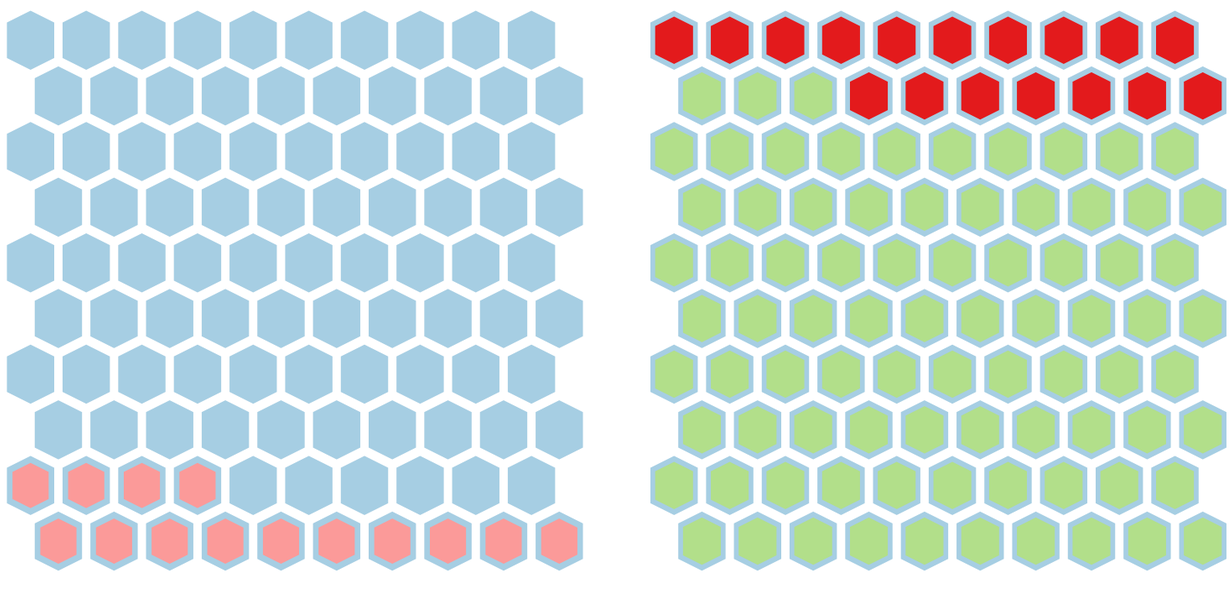
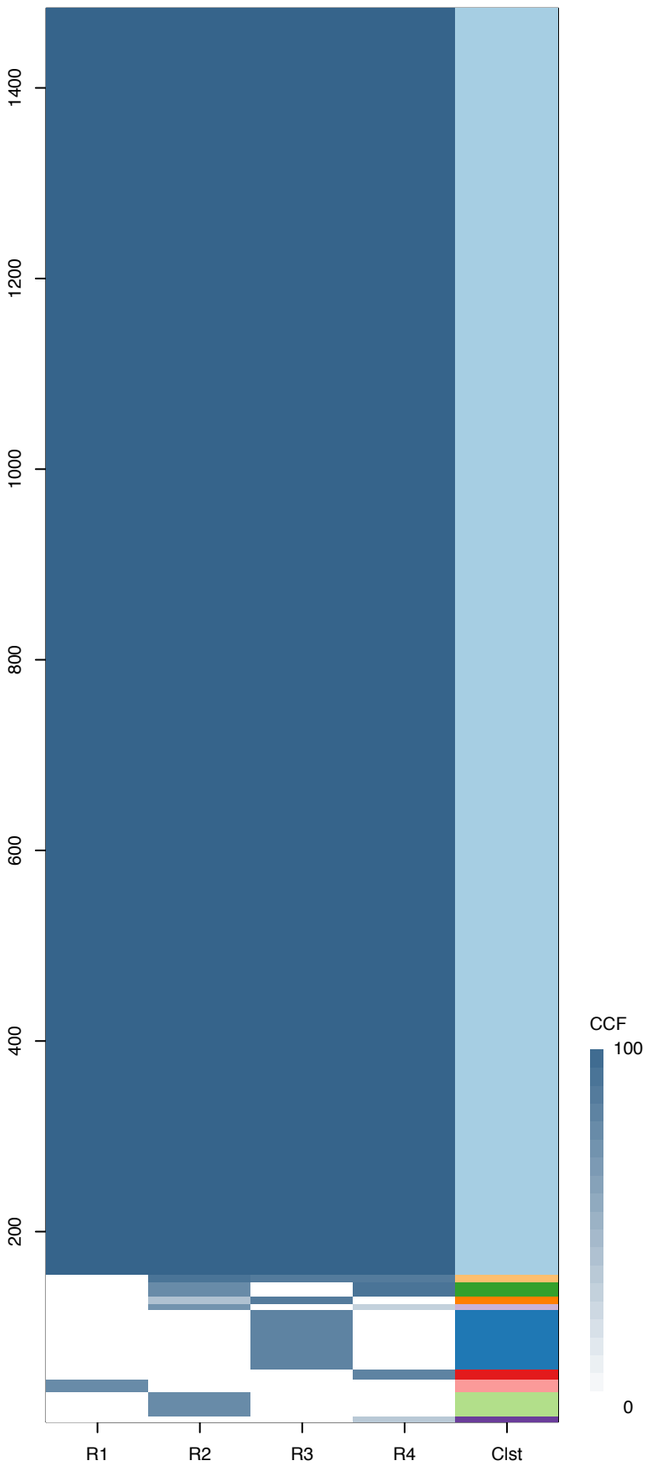
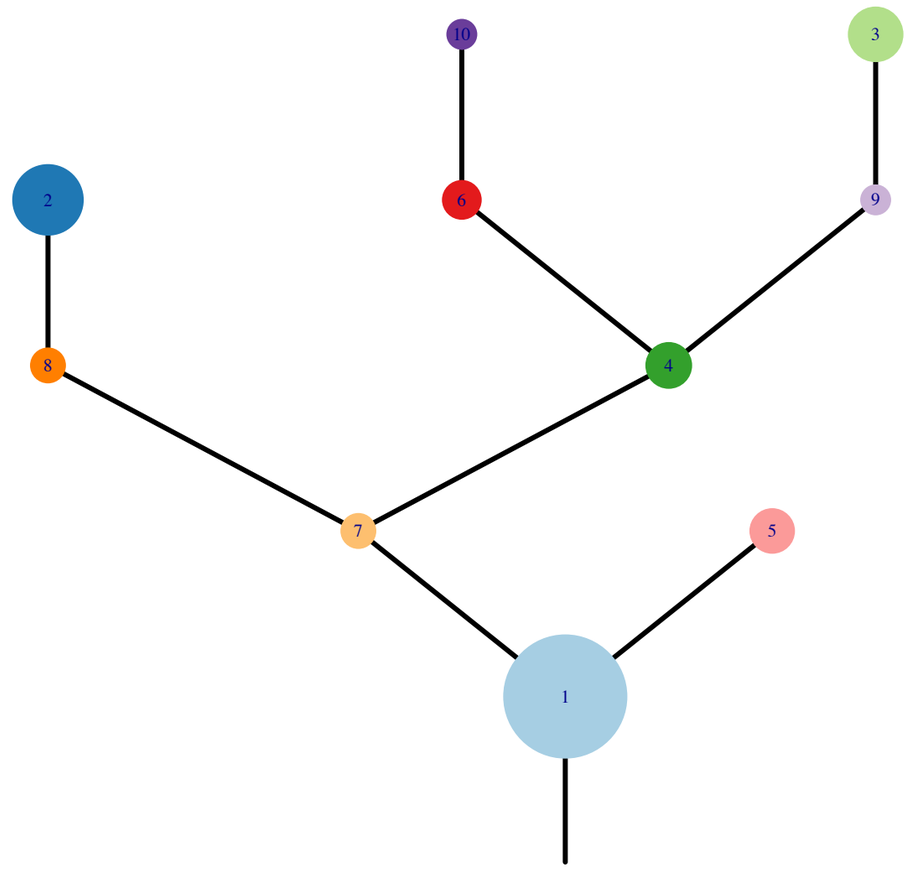


Fig.S12CA



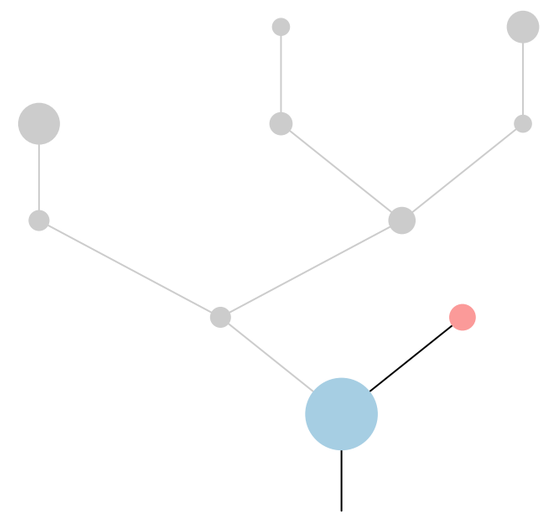
CRUK0085



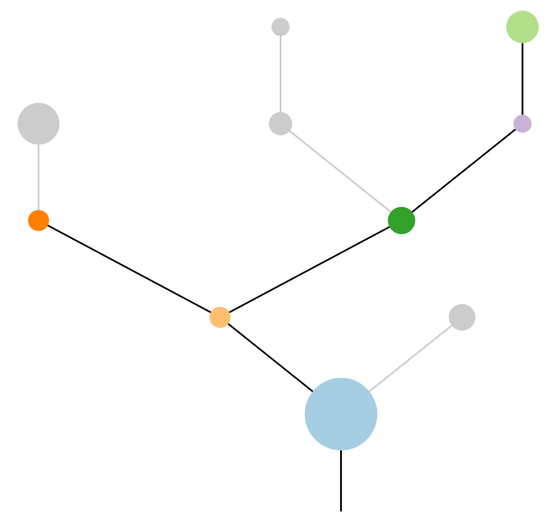
Histology:Squamous, Age:79, PackYears:50.25, Size:29
Stage:1a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
LATS1	1	6q25.1	SNV
FANCM	1	14q21.2	SNV
CREBBP	1	16p13.3	SNV
CHEK2	1	22q12.1	SNV
ARNT	2	1q21.3	Amp
SETDB1	2	1q21.3	Amp
MLLT11	2	1q21.3	Amp
PLAG1	5	8q12.1	Amp
CHCHD7	5	8q12.1	Amp
CDKN2A	7	9p21.3	Del
SDHB	8	1p36.13	Del

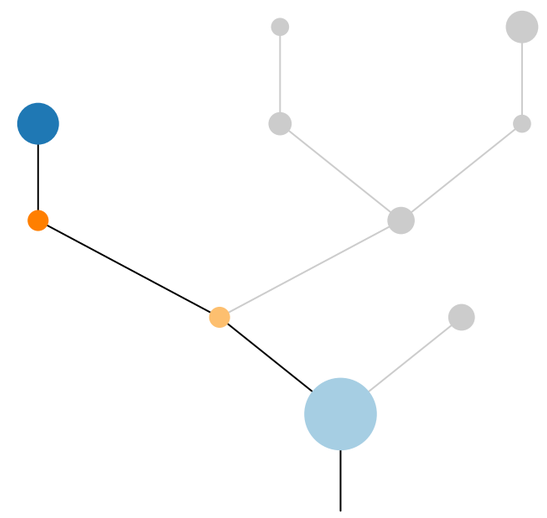
R1



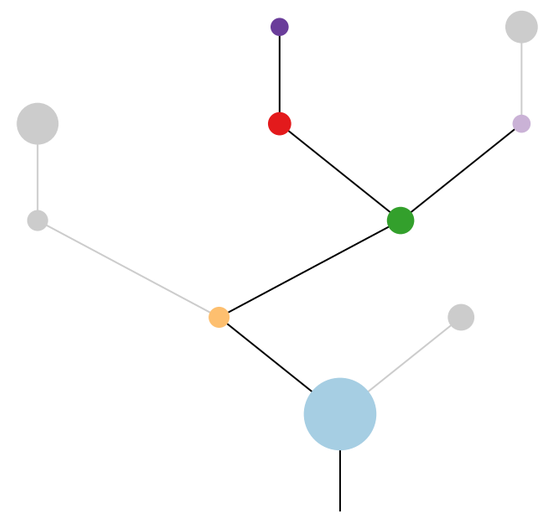
R2

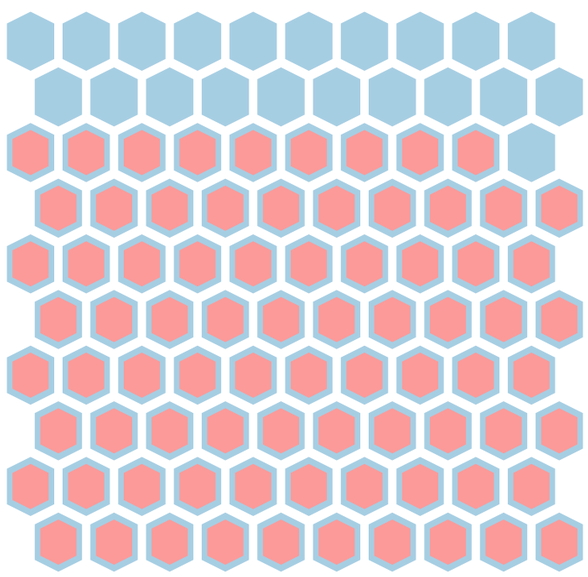


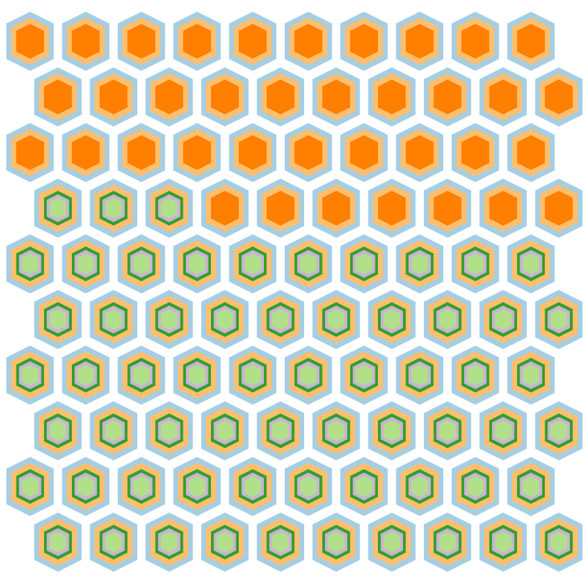
R3

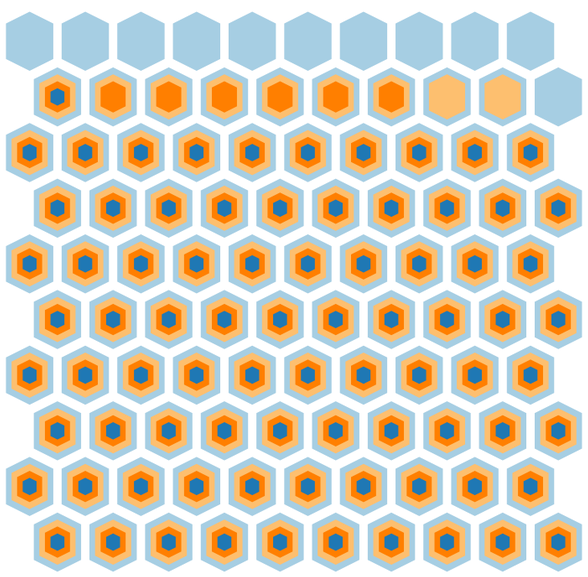


R4









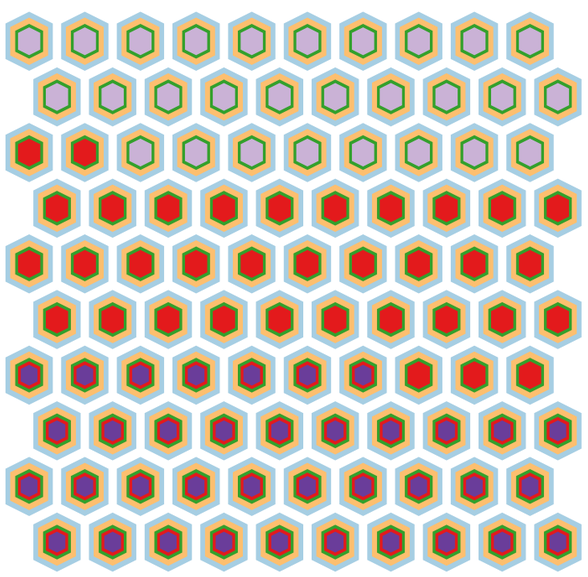
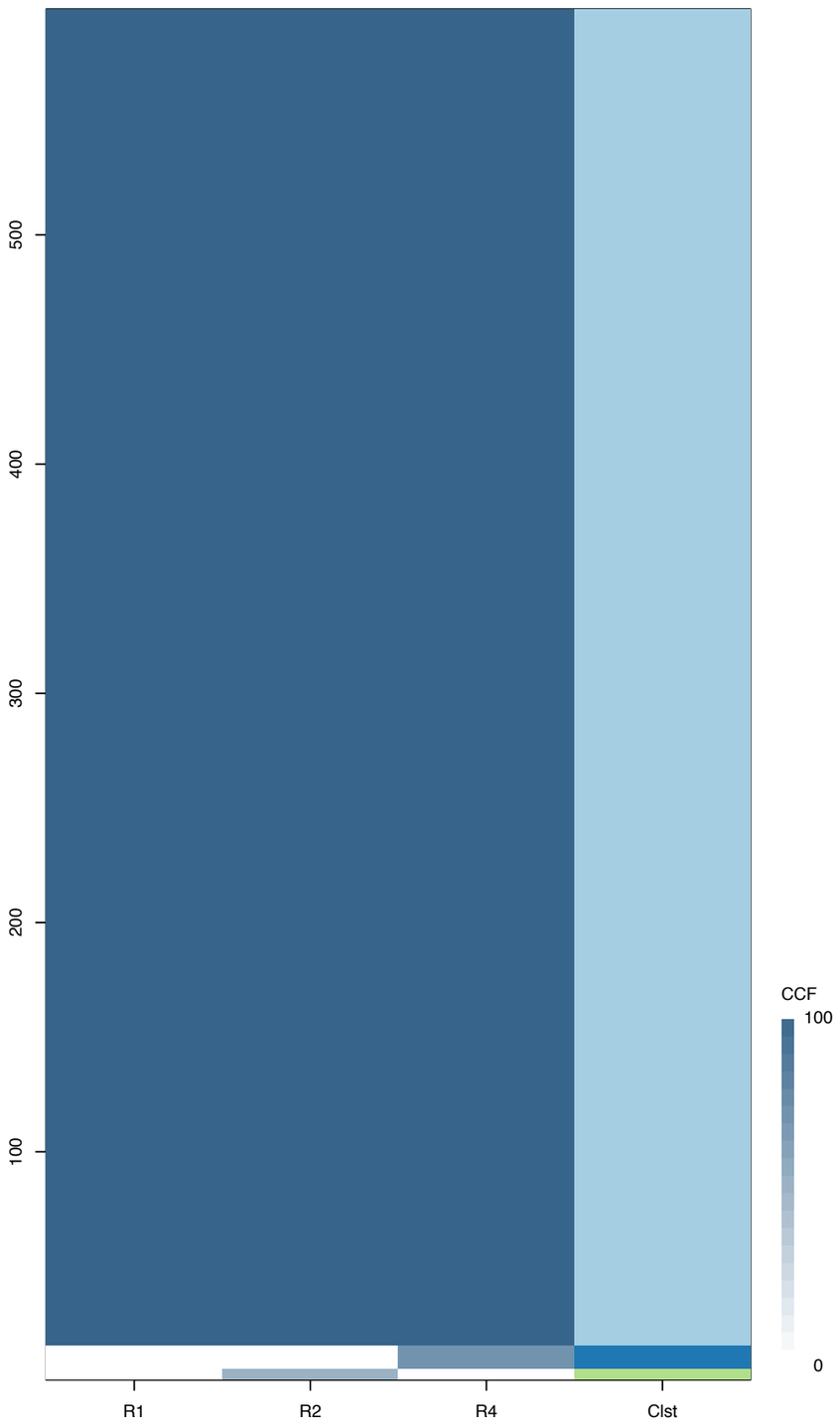
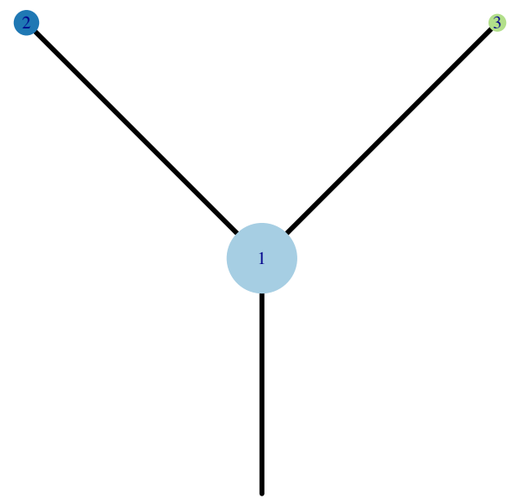


Fig.S12CB



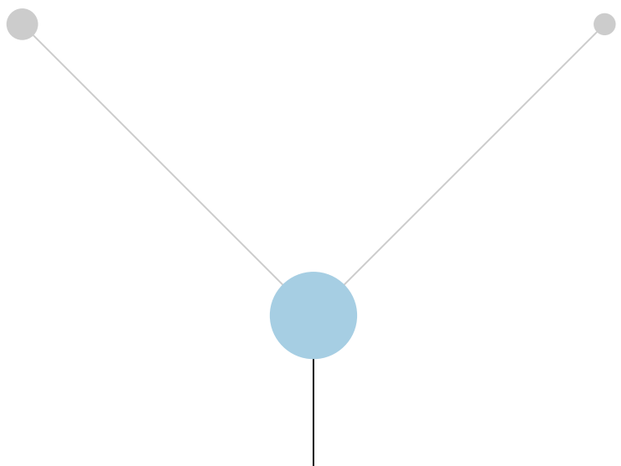
CRUK0086

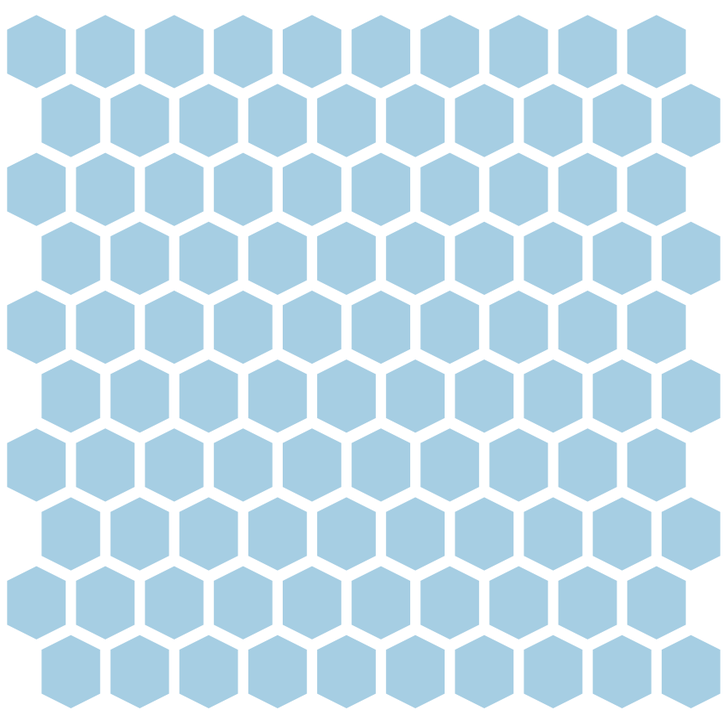


Histology:Squamous, Age:80, PackYears:92, Size:42
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

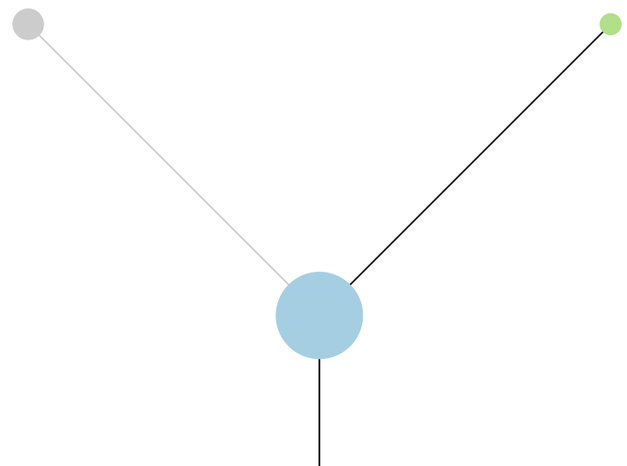
Gene	Cluster	Cytoband	Type
FAT1	1	4q35.2	Del
CARD11	1	7p22.2	Amp
RAC1	1	7p22.1	Amp
ARID2	1	12q12	SNV
TP53	1	17p13.1	SNV

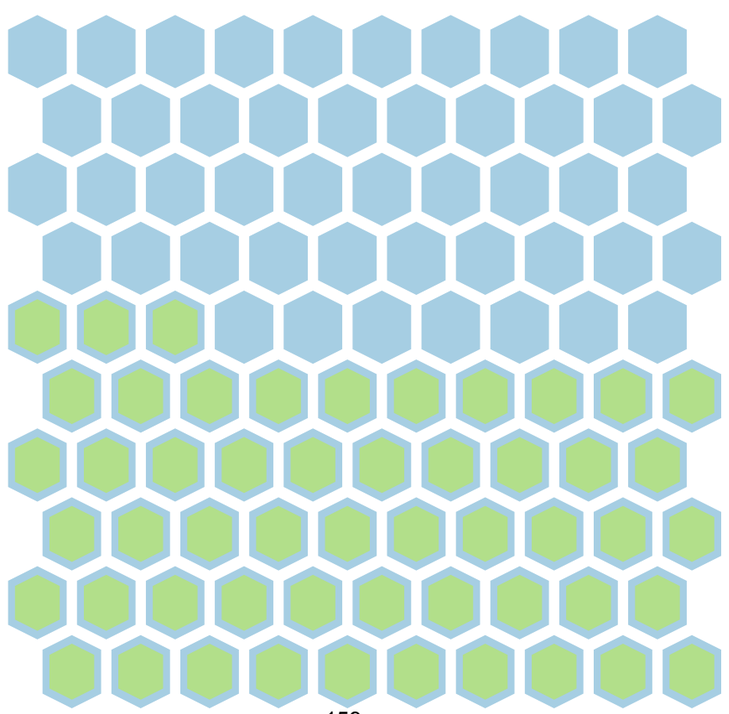
R1



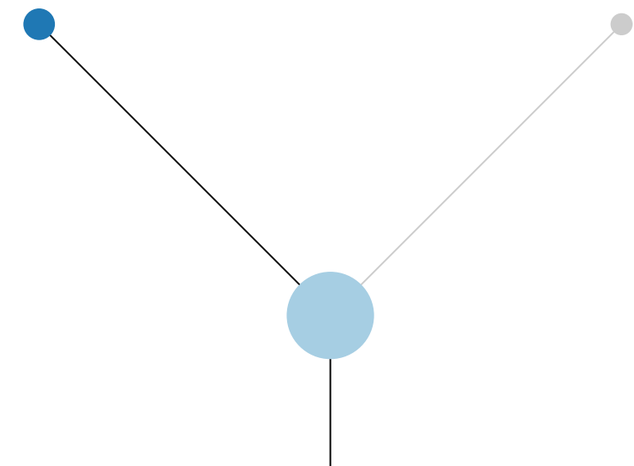


R2





R4



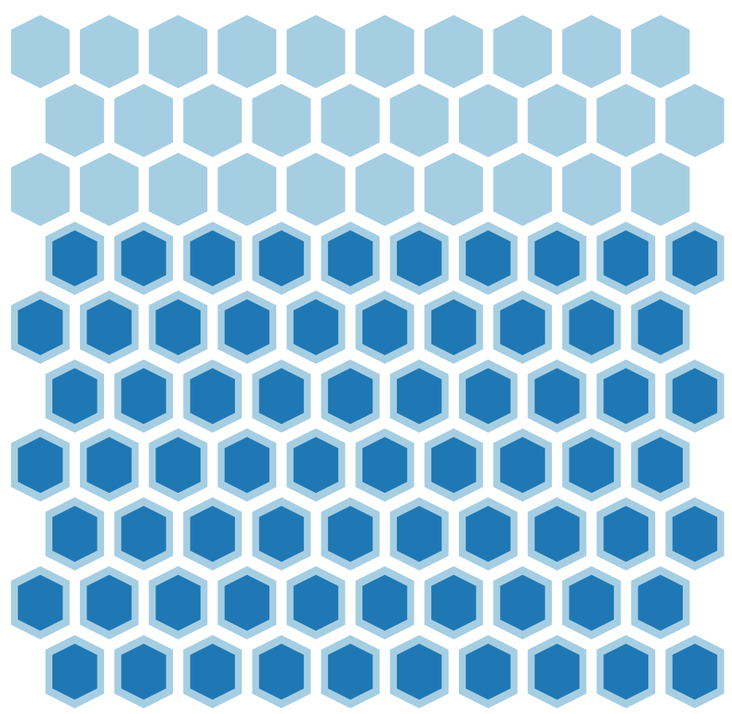
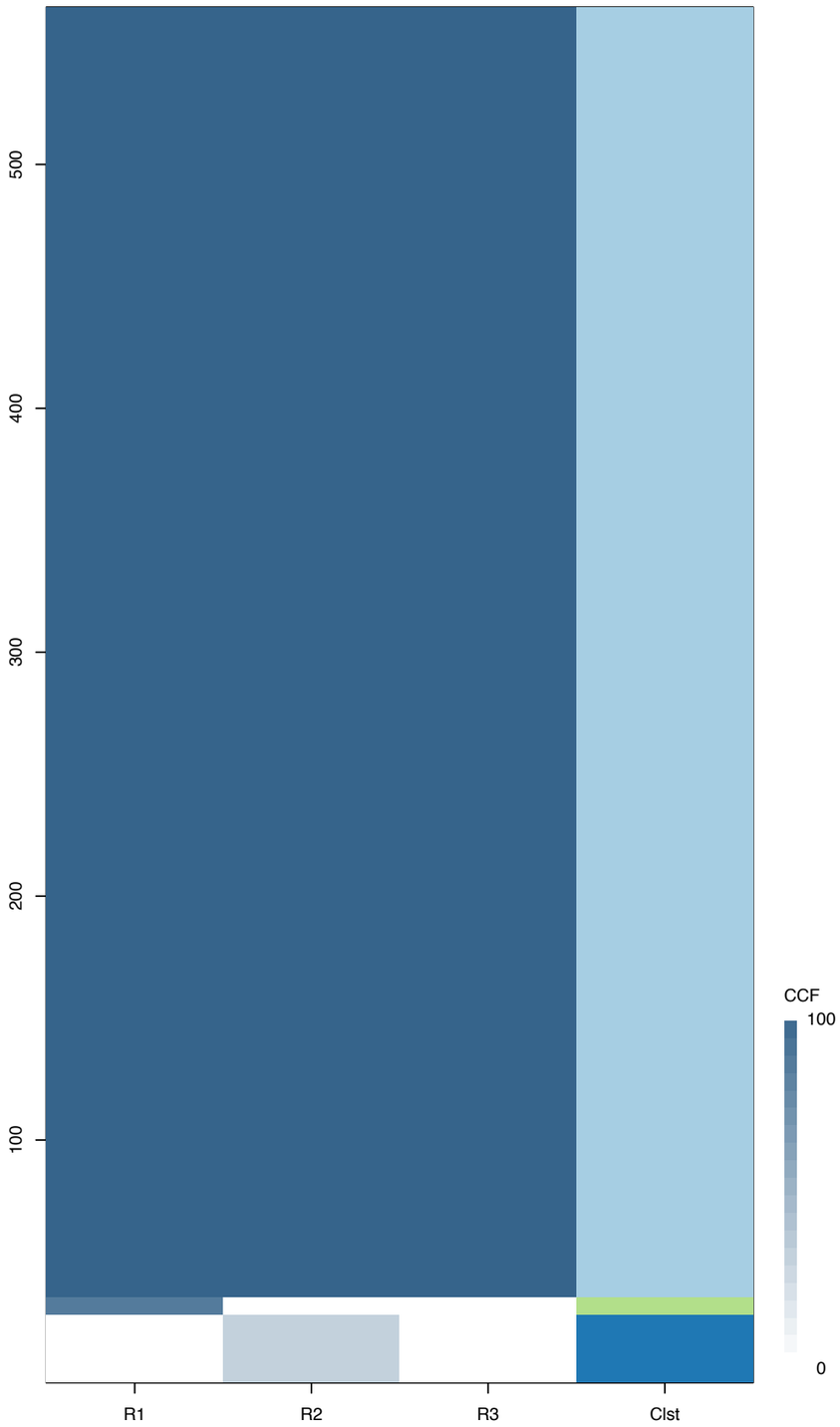
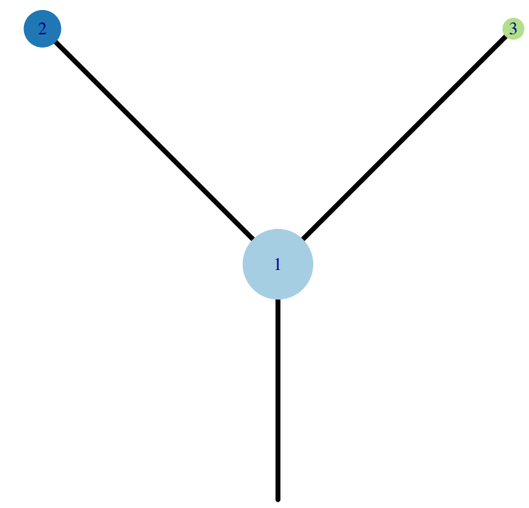


Fig.S12CC



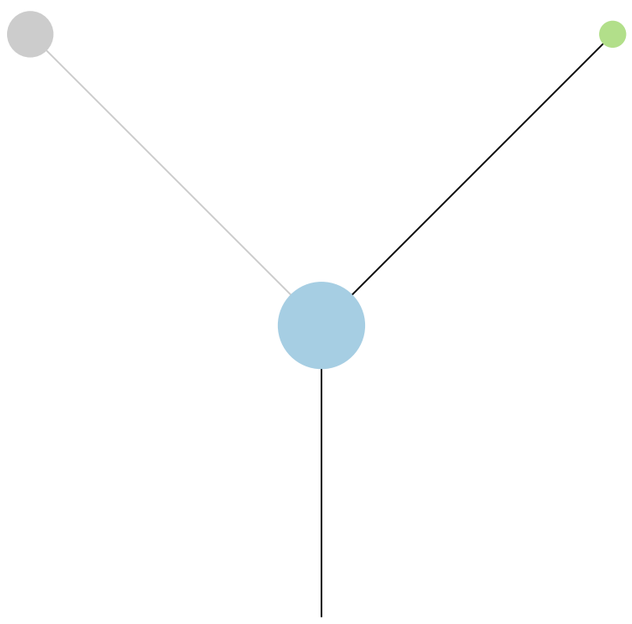
CRUK0087



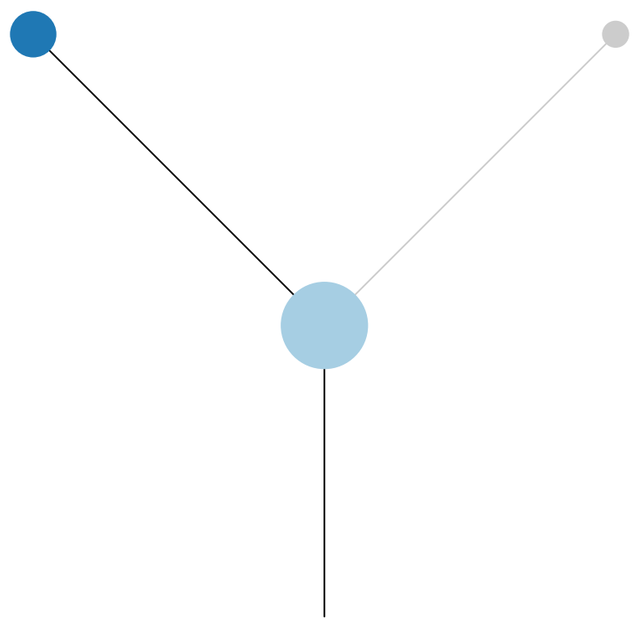
Histology:Squamous, Age:82, PackYears:2.1, Size:110
Stage:3a, Gender:Female, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
NFE2L2	1	2q31.2	SNV
PBRM1	1	3p21.1	SNV
PIK3CA	1	3q26.32	SNV
MAP3K13	1	3q27.2	SNV
TP53	1	17p13.1	SNV
ASXL1	1	20q11.21	SNV
CDKN2A	?	9p21.3	Del
H3F3B	?	17q25.1	Amp

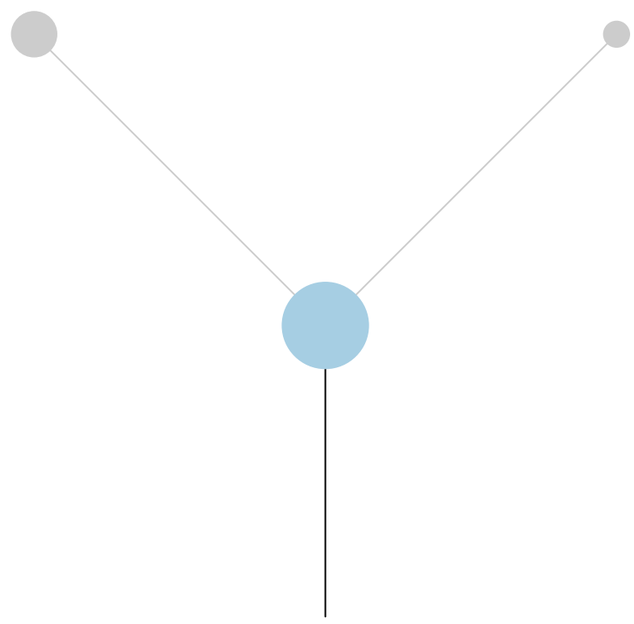
R1

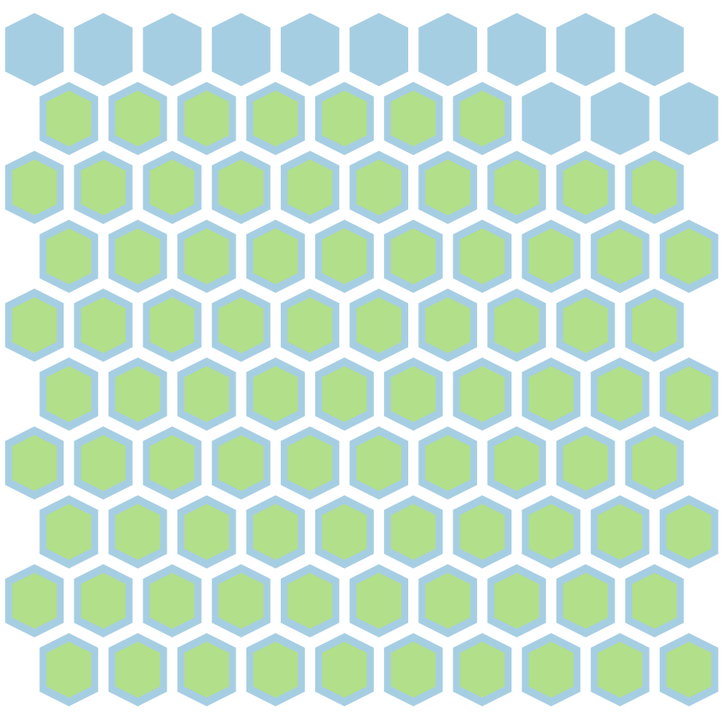


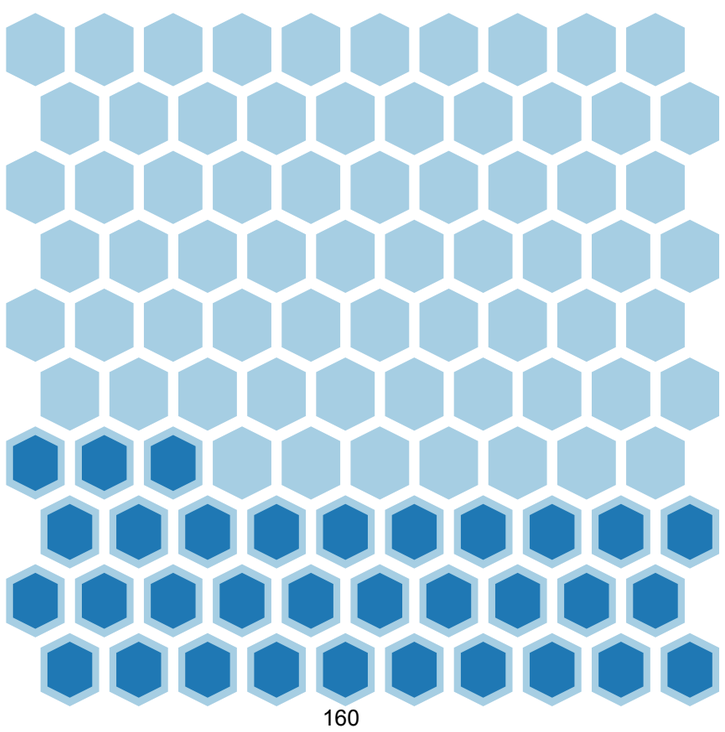
R2



R3







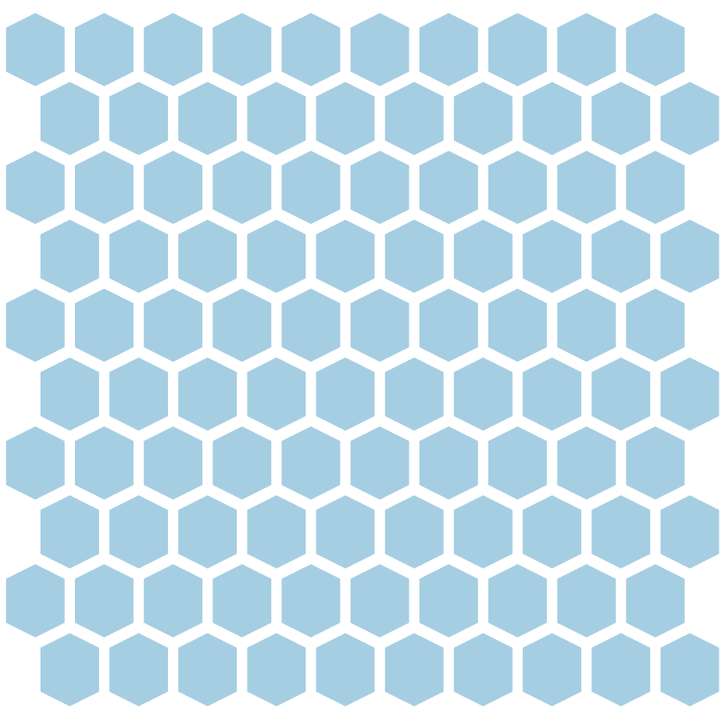
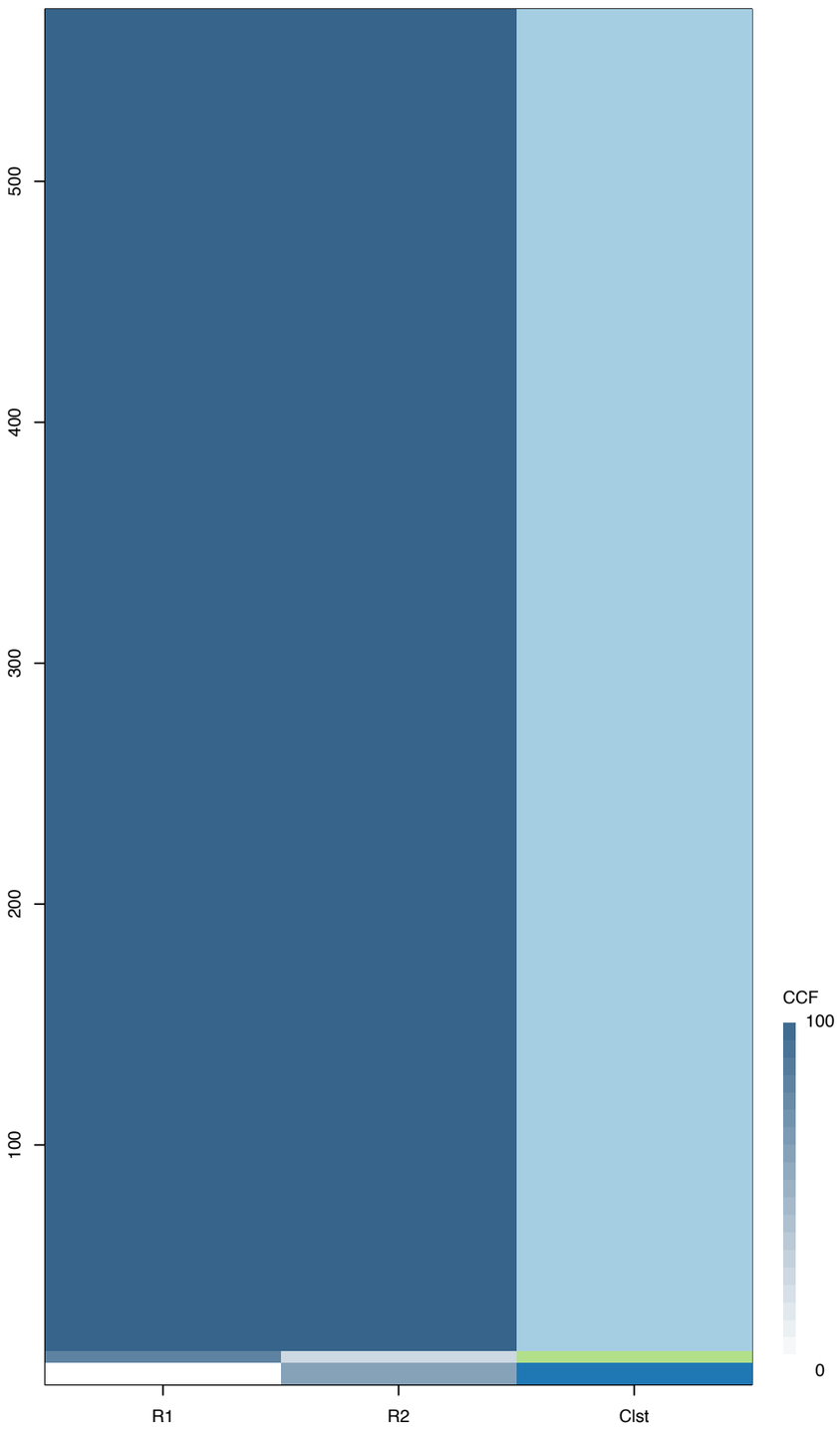
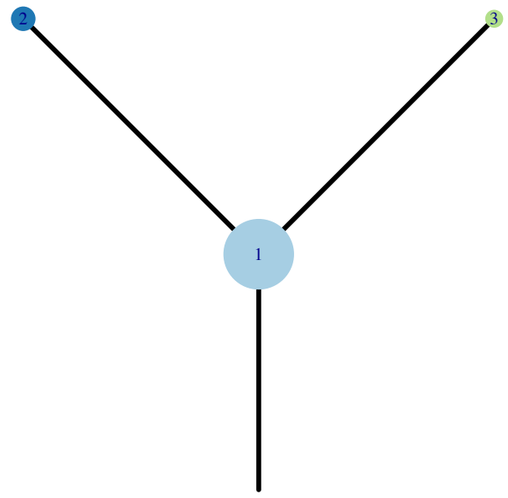


Fig.S12CD



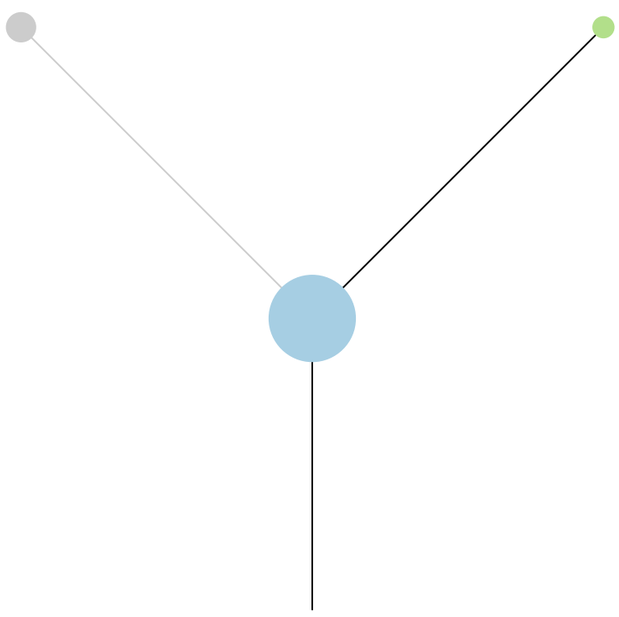
CRUK0088



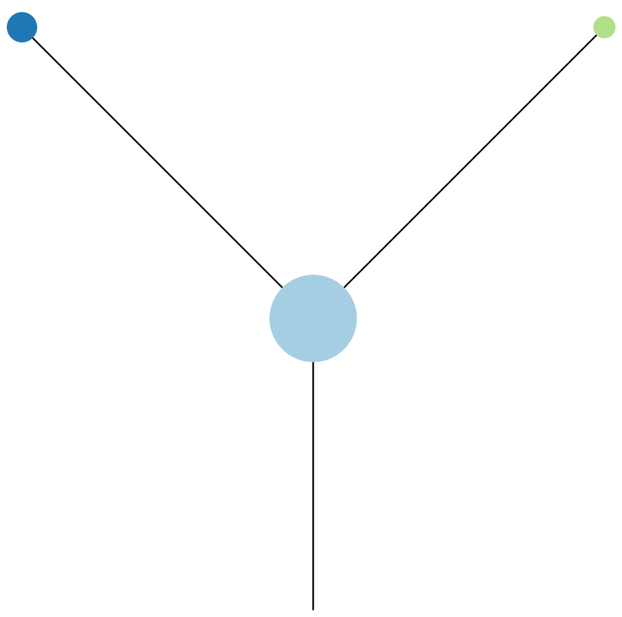
Histology:Squamous, Age:82, PackYears:40, Size:10
Stage:1a, Gender:Male, GD:Clonal GD, Recur:yes

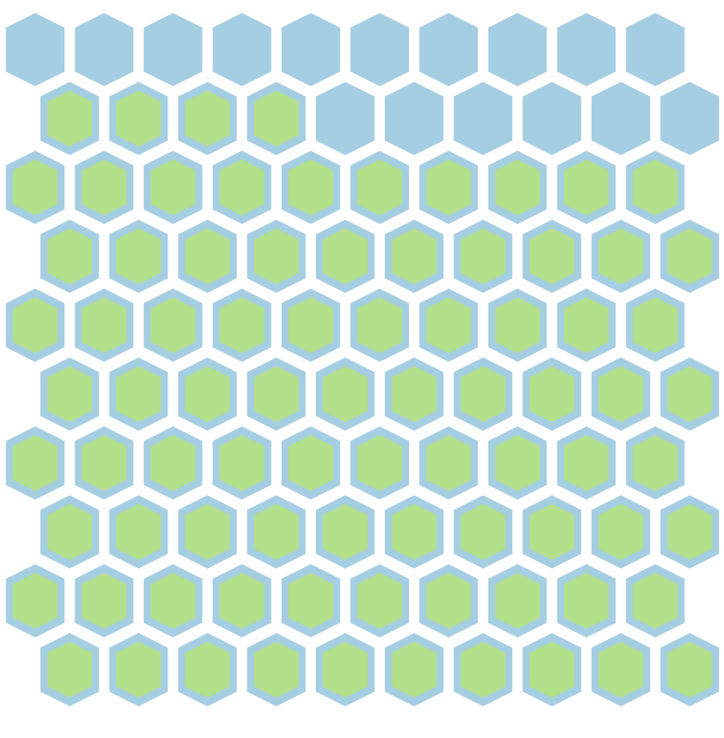
Gene	Cluster	Cytoband	Type
CUX1	1	7q22.1	SNV
TP53	1	17p13.1	SNV
ZNF521	1	18q11.2	Amp
SS18	1	18q11.2	Amp
SETBP1	1	18q12.3	Amp
MALT1	1	18q21.32	Amp
BCL2	1	18q21.33	Amp
KDSR	1	18q21.33	Amp
CD79B	2	17q23.3	Amp
CEBPA	2	19q13.11	Amp
LSM14A	2	19q13.11	Amp

R1



R2





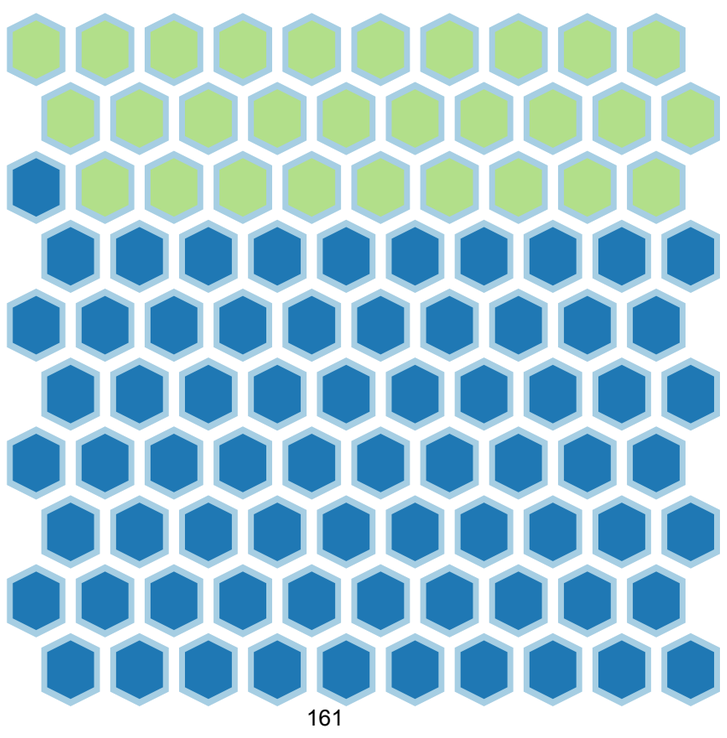
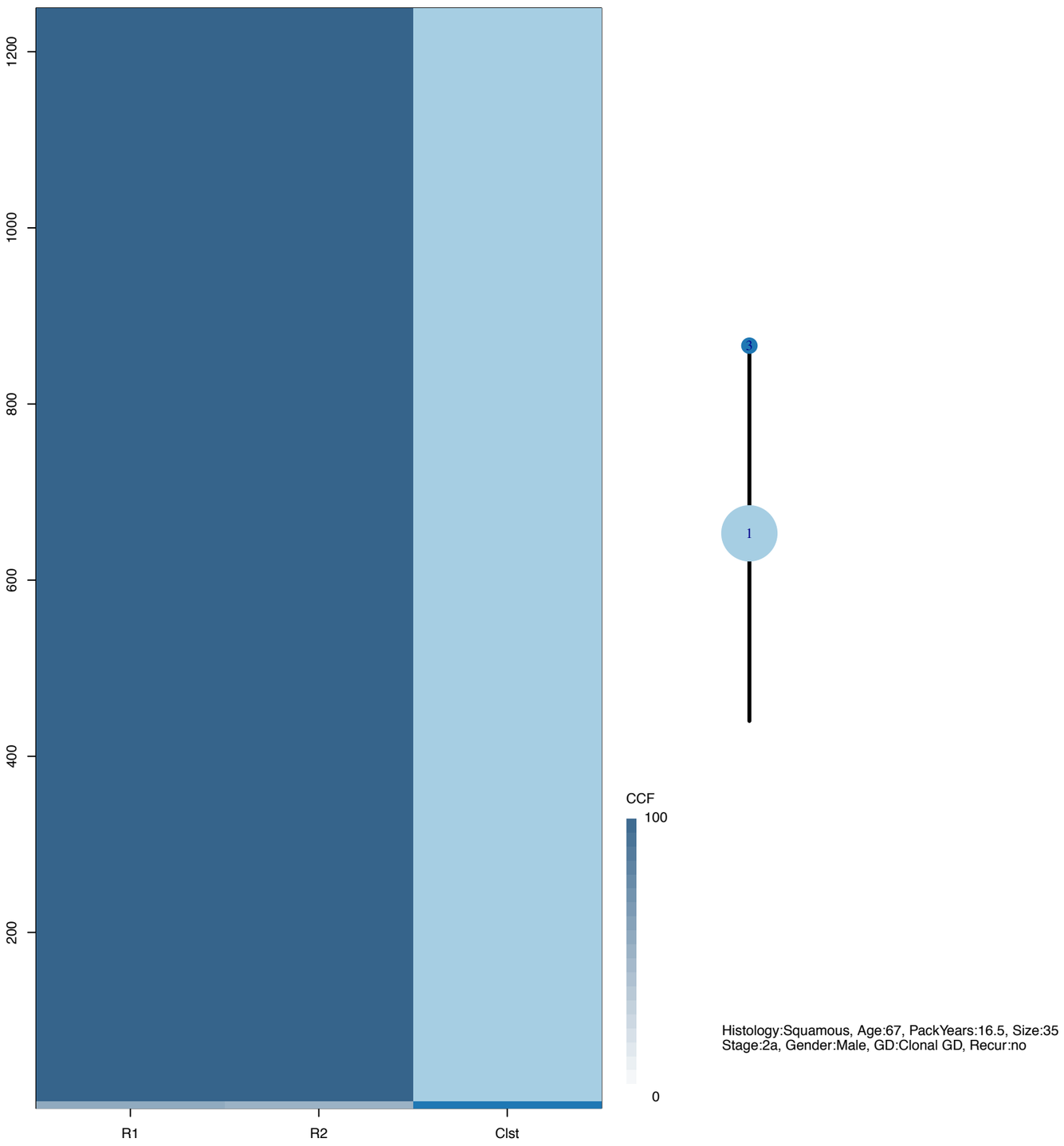


Fig.S12CE

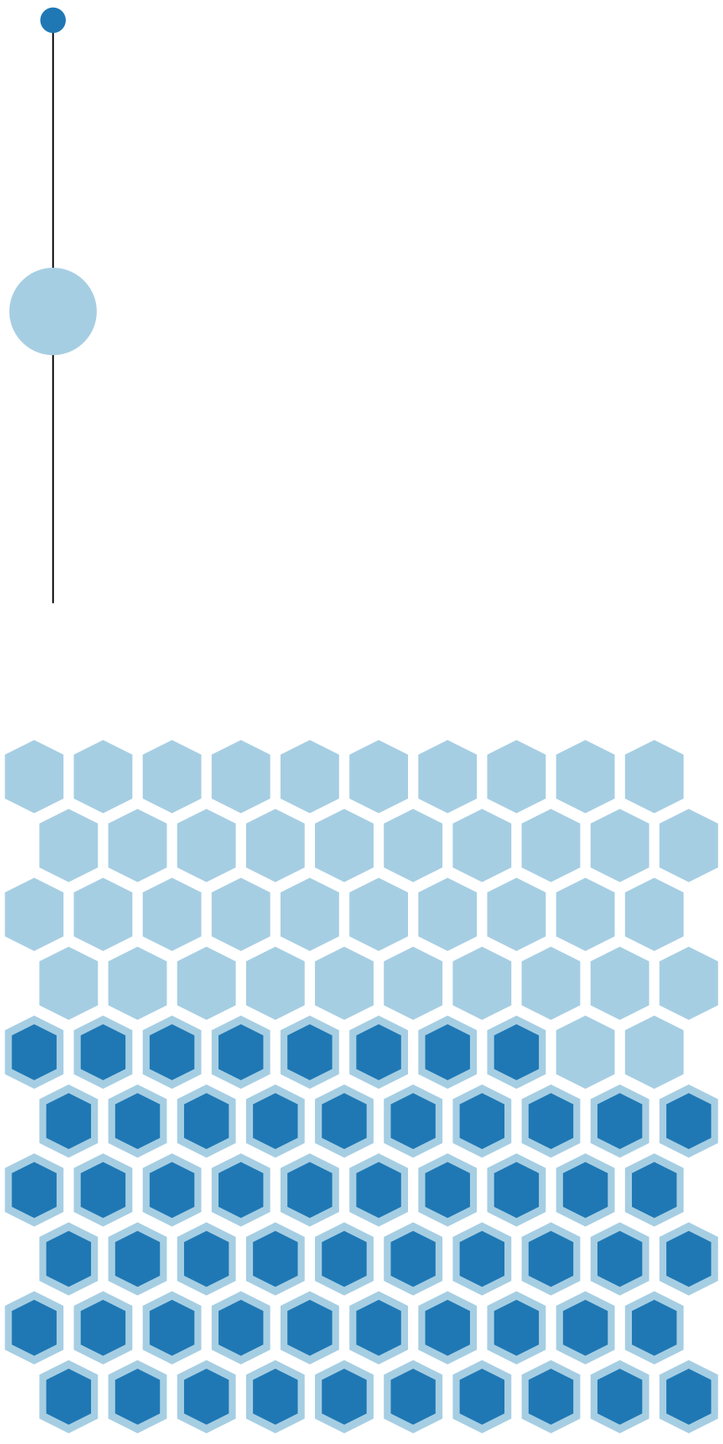
CRUK0090



Histology:Squamous, Age:67, PackYears:16.5, Size:35
Stage:2a, Gender:Male, GD:Clonal GD, Recur:no

Gene	Cluster	Cytoband	Type
NRAS	1	1p13.2	SNV
CUX1	1	7q22.1	SNV
CDKN2A	1	9p21.3	SNV
COL2A1	1	12q13.11	SNV
NCOA6	1	20q11.22	SNV

R1



R2

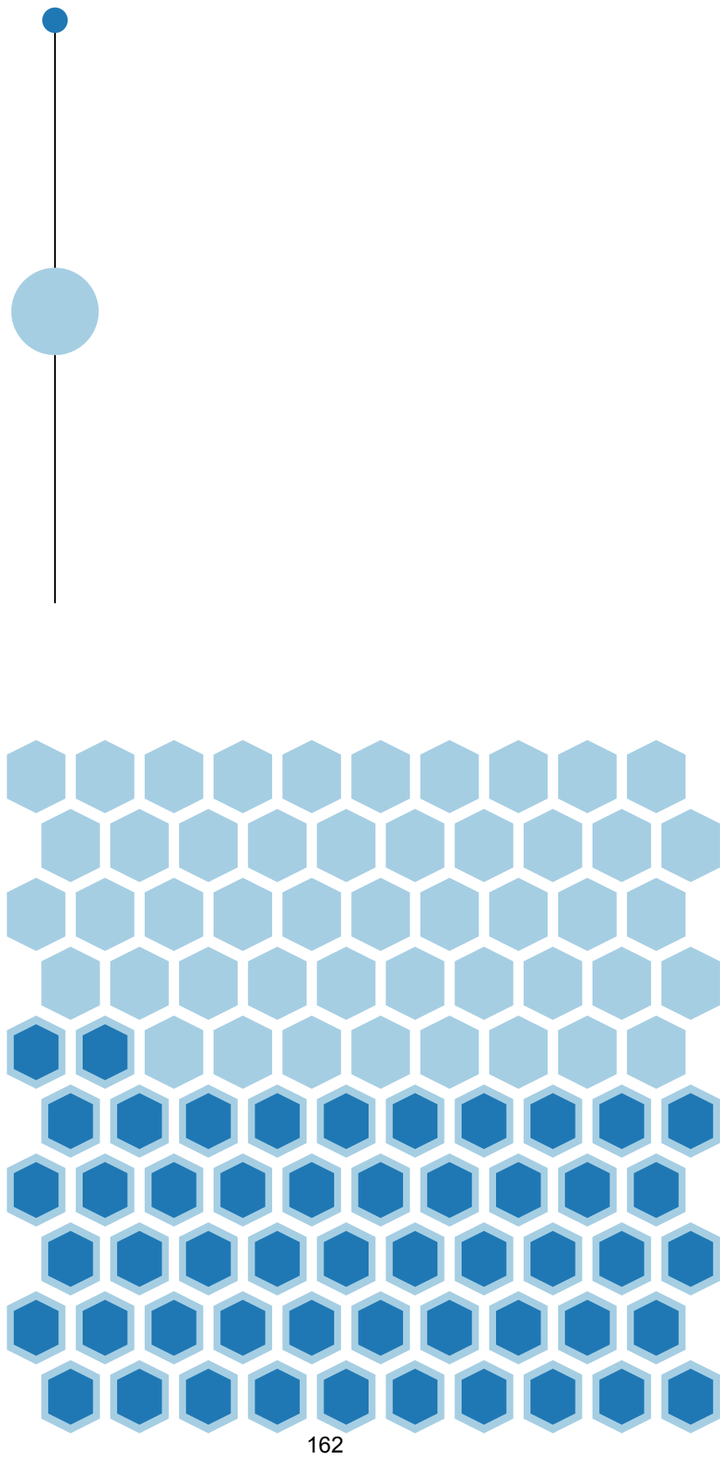
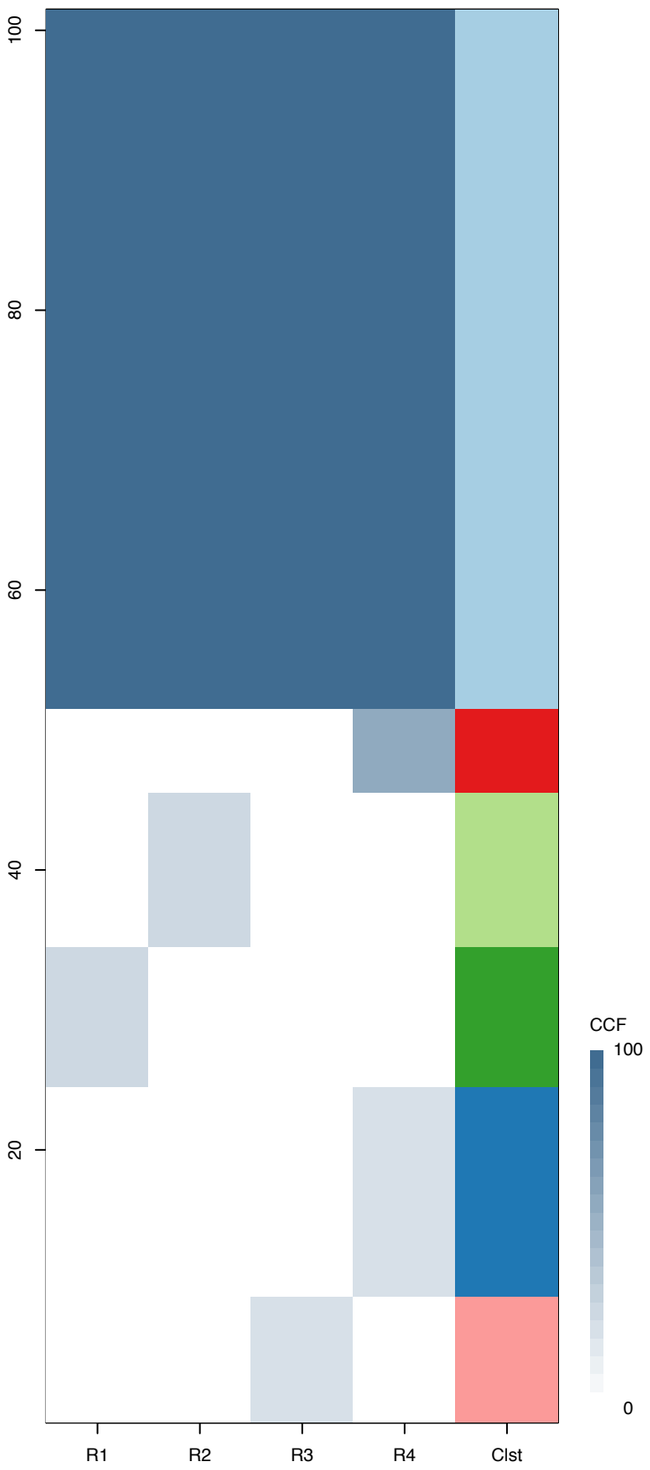
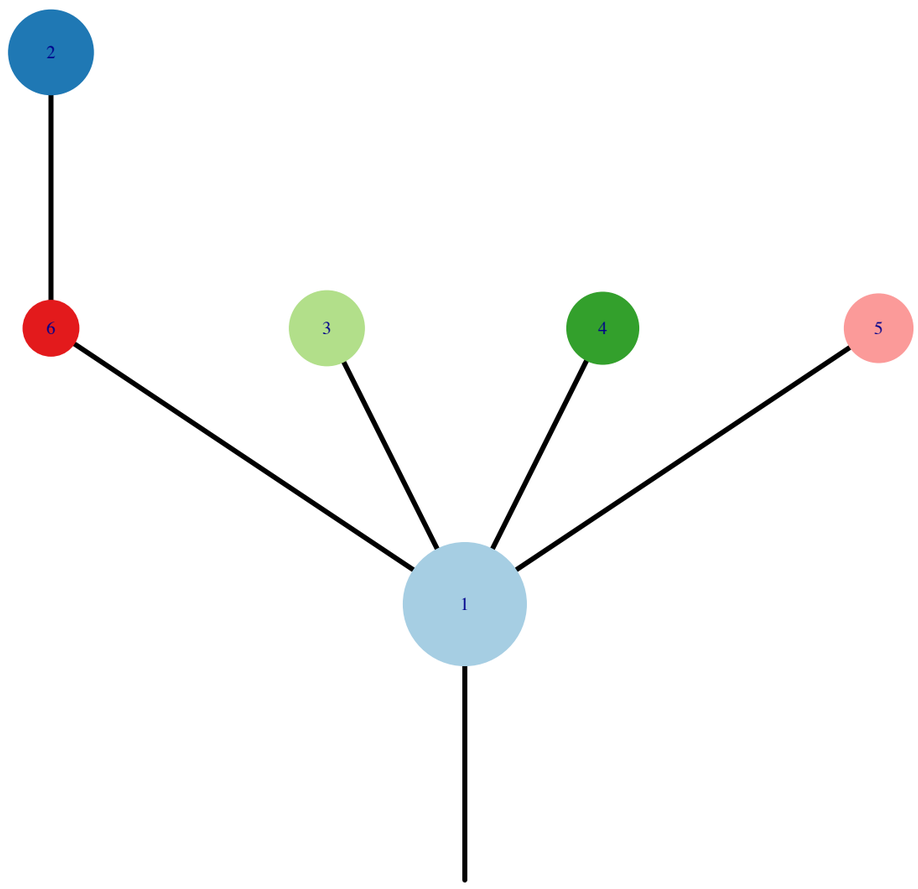


Fig.S12CF



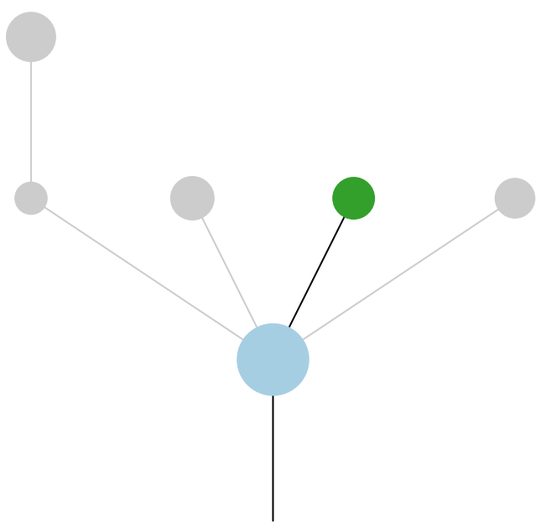
CRUK0094_A



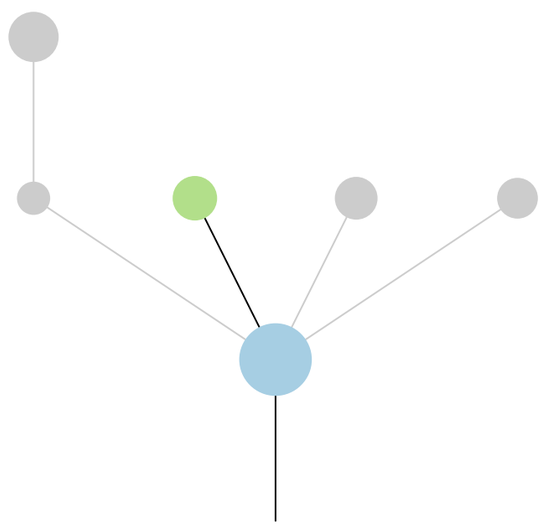
Histology:Other, Age:35, PackYears:NA, Size:35
Stage:2a, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
TERT	1	5p15.33	Amp
SMARCA4	1	19p13.2	SNV
ERCC5	2	13q33.1	SNV

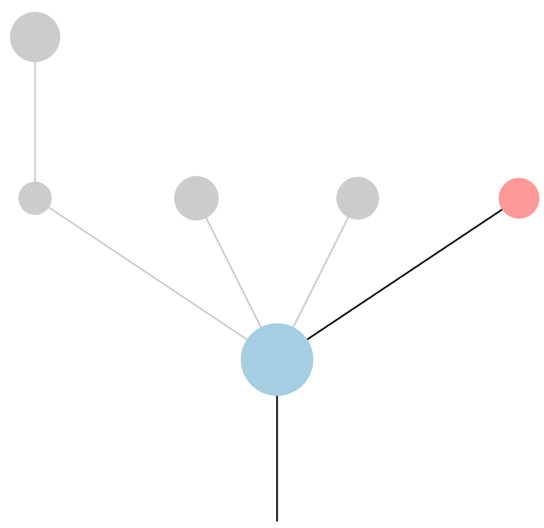
R1



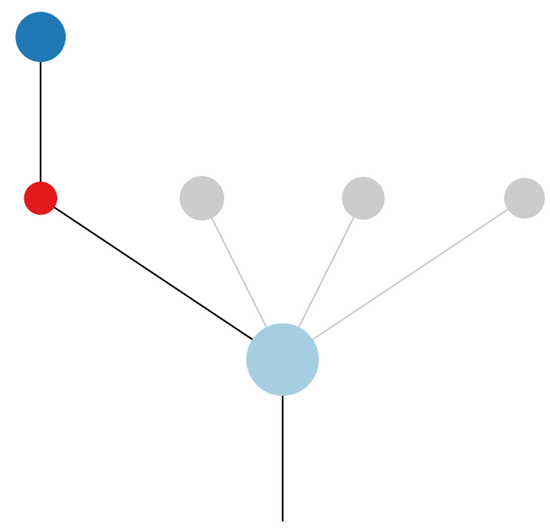
R2



R3



R4



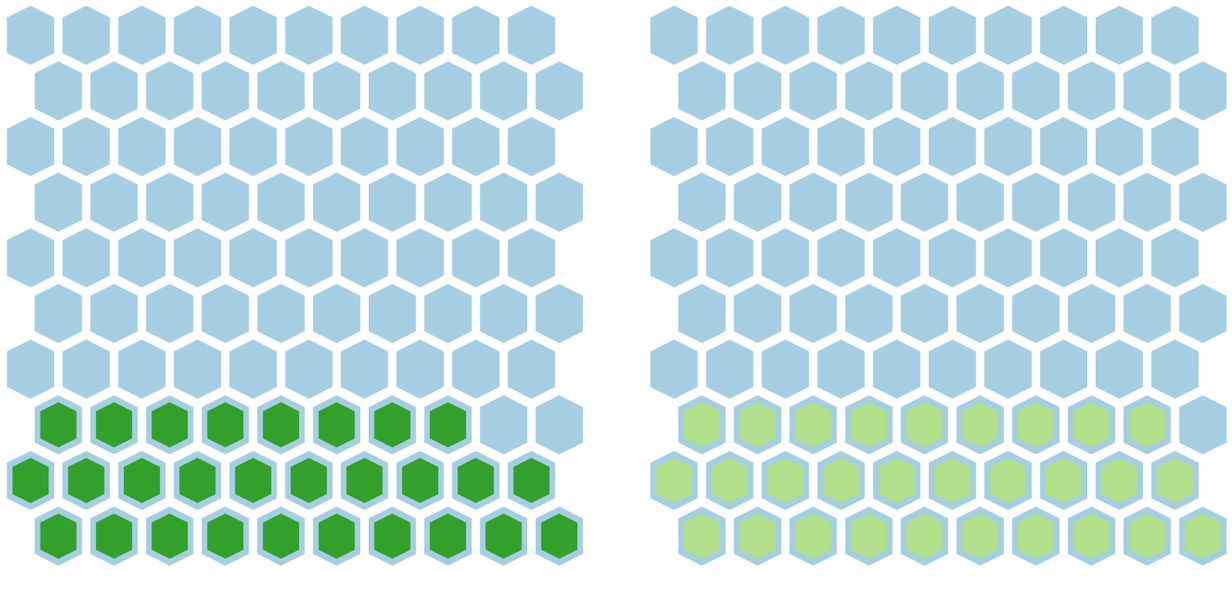
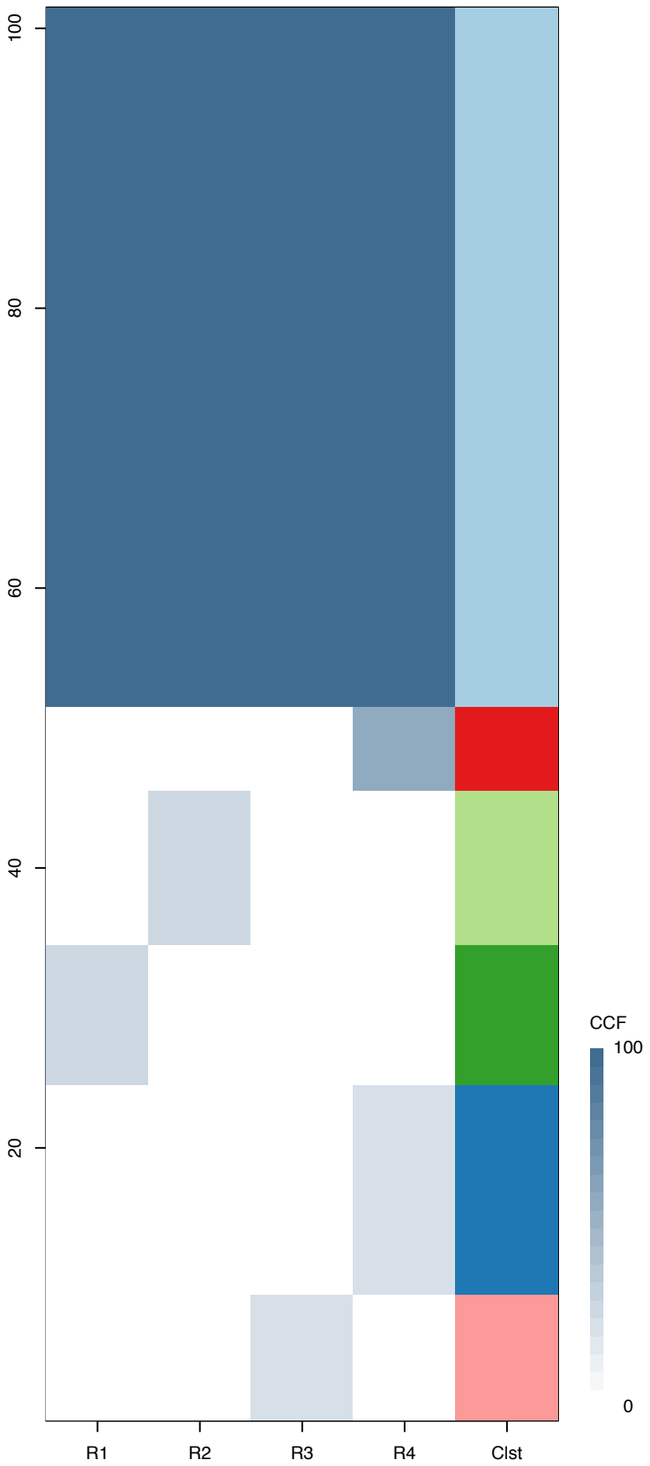
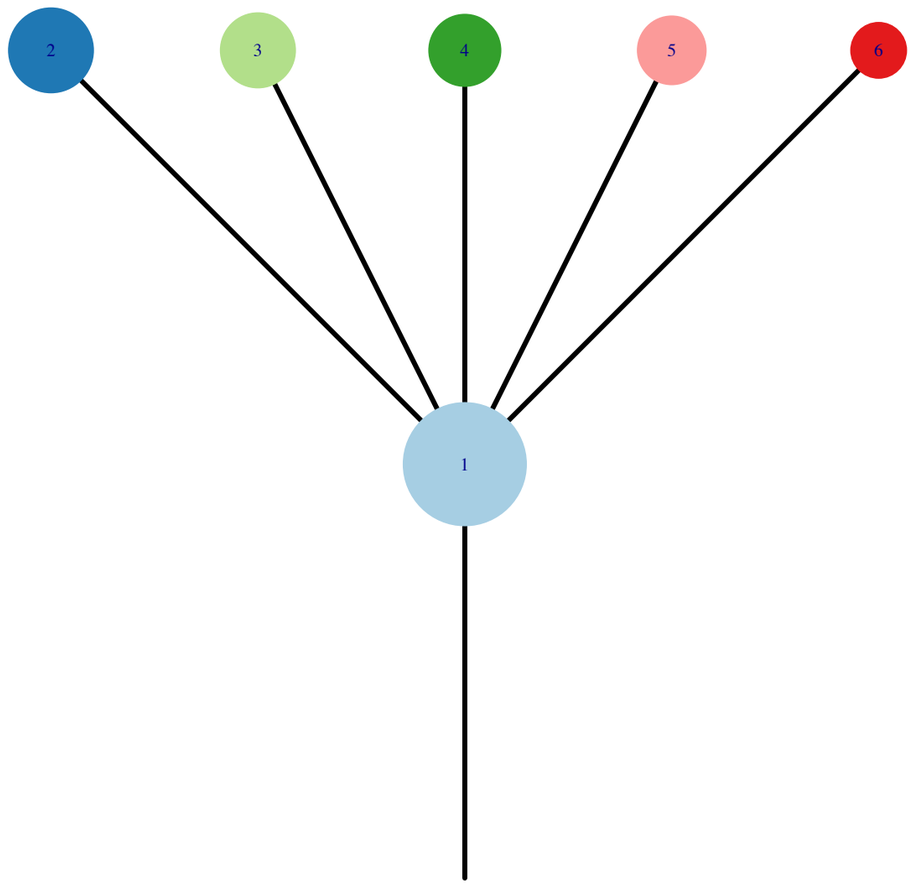


Fig.S12CF



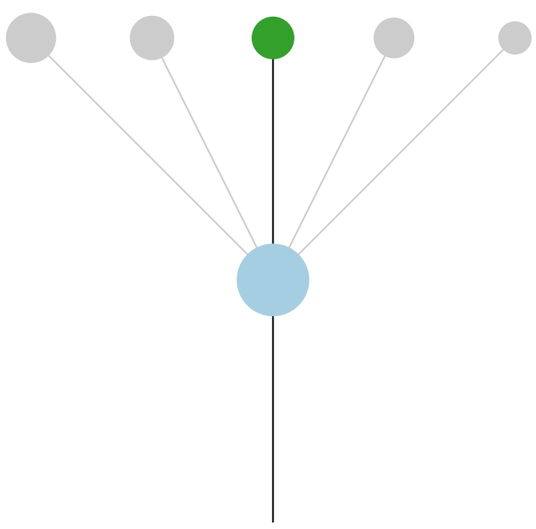
CRUK0094_B



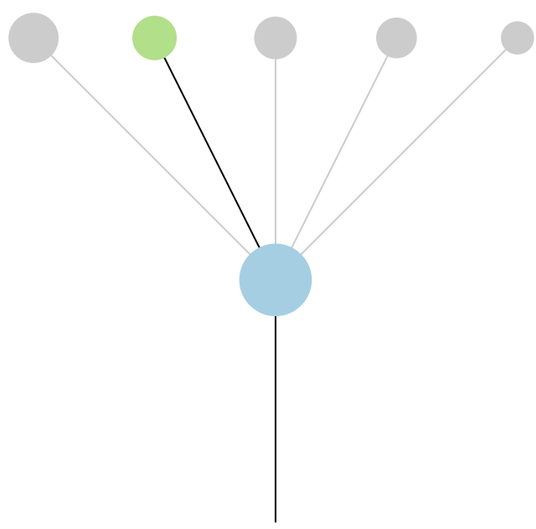
Histology:Other, Age:35, PackYears:NA, Size:35
Stage:2a, Gender:Male, GD:Not GD, Recur:yes

Gene	Cluster	Cytoband	Type
TERT	1	5p15.33	Amp
SMARCA4	1	19p13.2	SNV
ERCC5	2	13q33.1	SNV

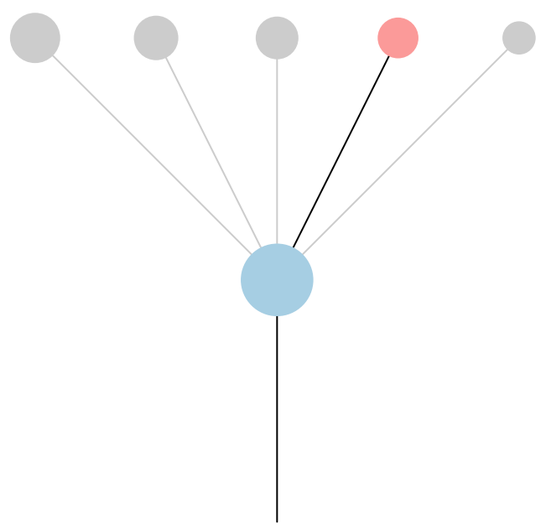
R1



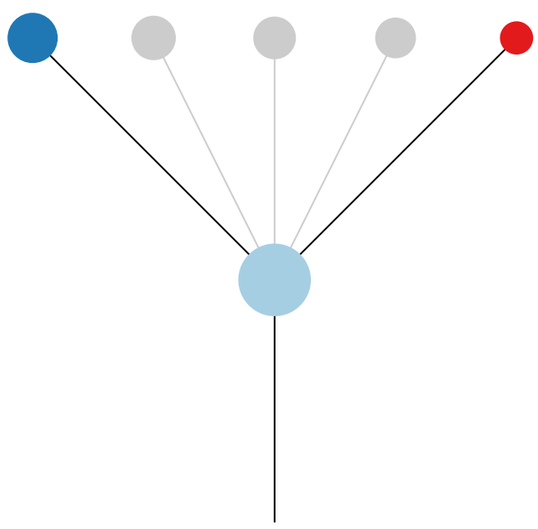
R2



R3



R4



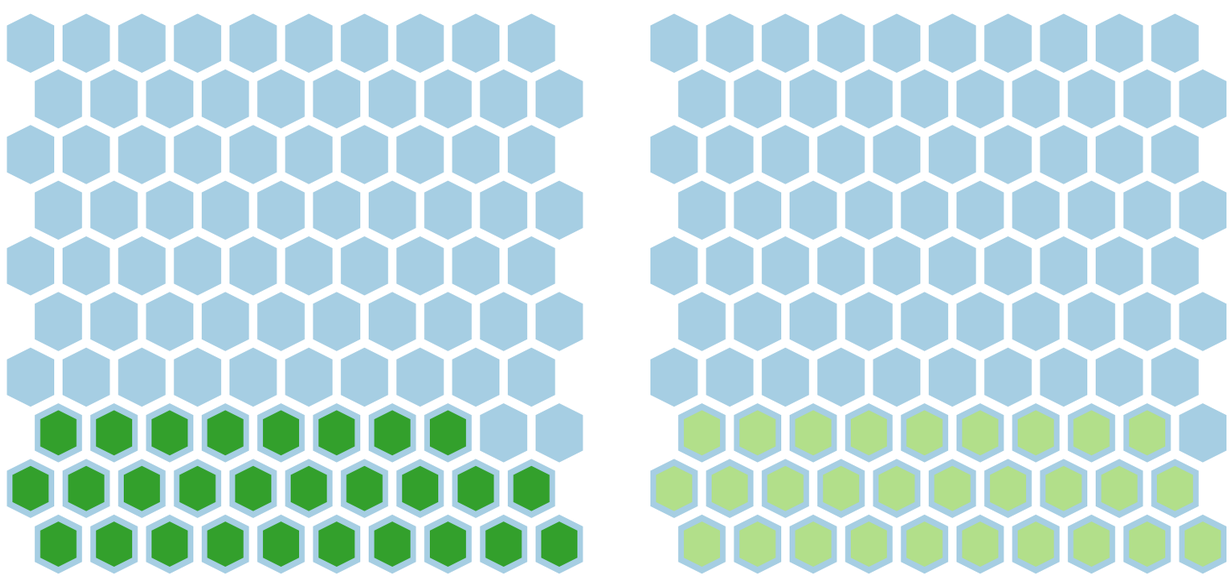
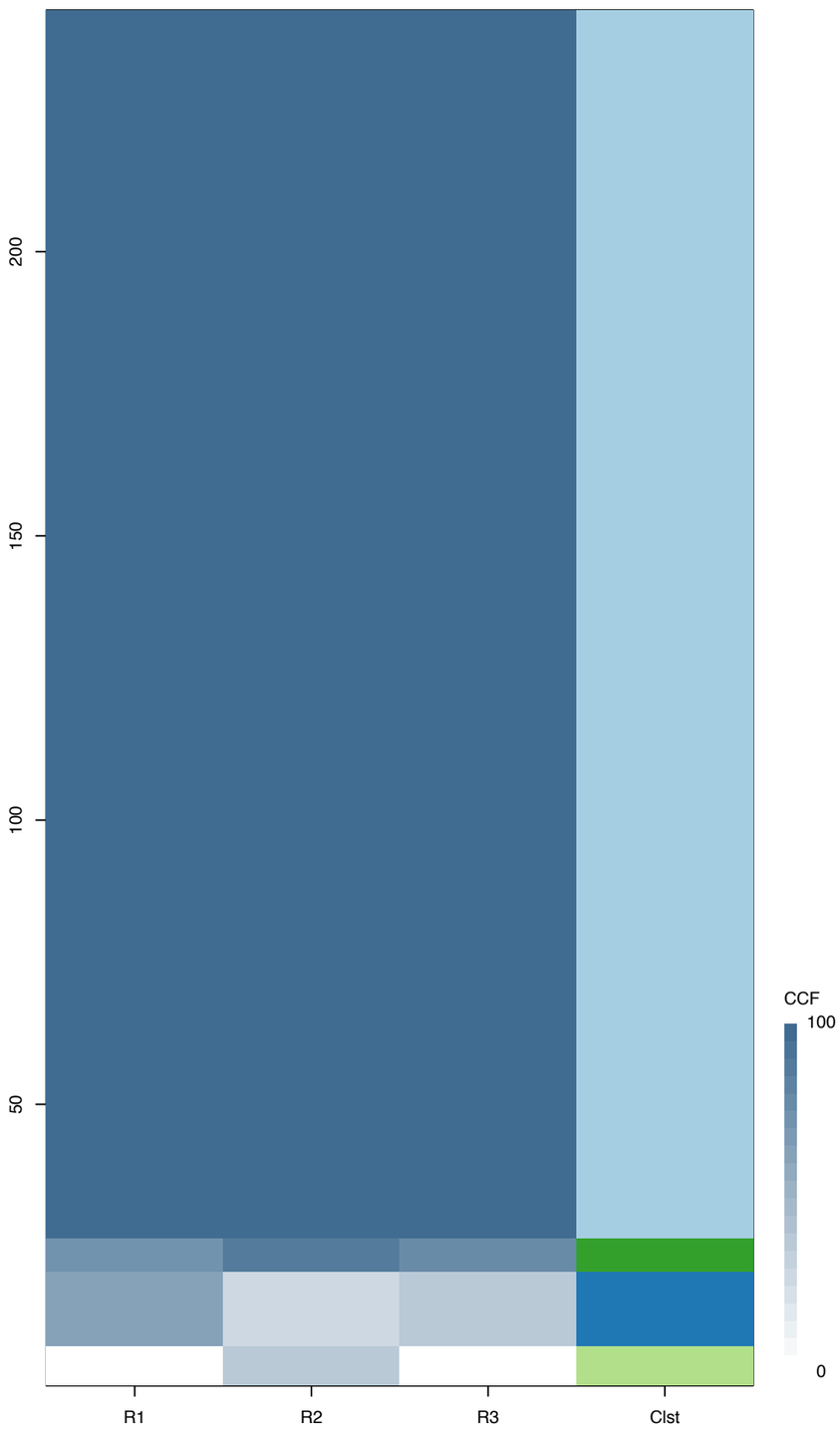
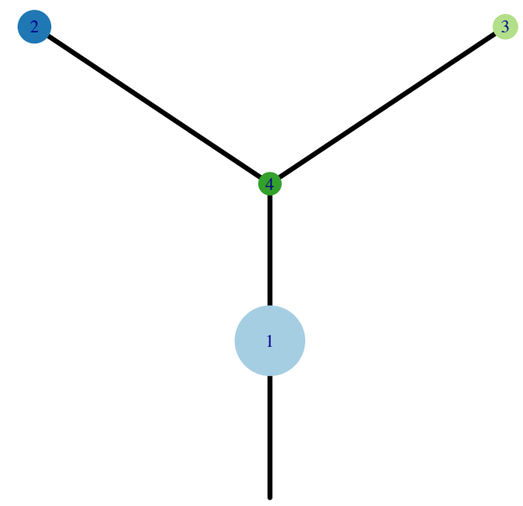


Fig.S12CG



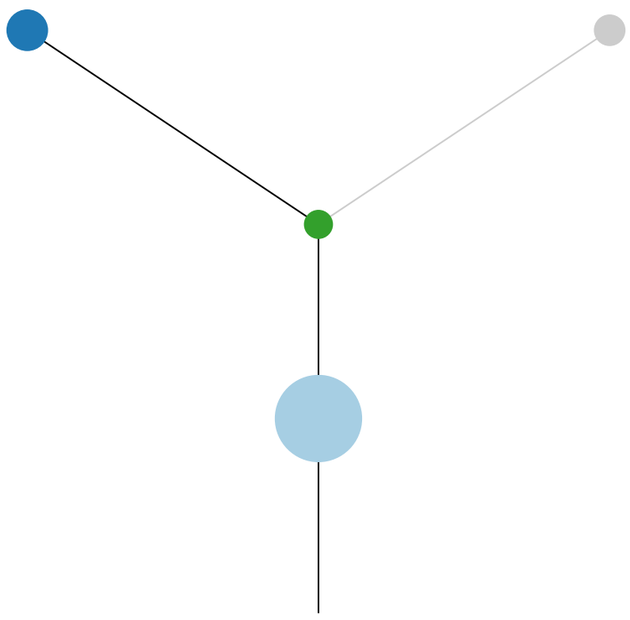
CRUK0095_A



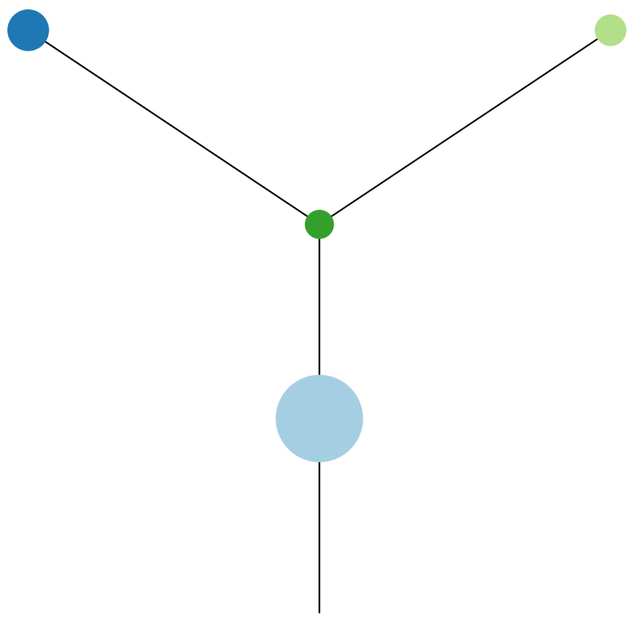
Gene	Cluster	Cytoband	Type
RASA1	1	5q14.3	SNV
TP53	1	17p13.1	SNV
NF1	1	17q11.2	SNV

Histology:Other, Age:60, PackYears:22.8, Size:20
Stage:1a, Gender:Female, GD:Not GD, Recur:no

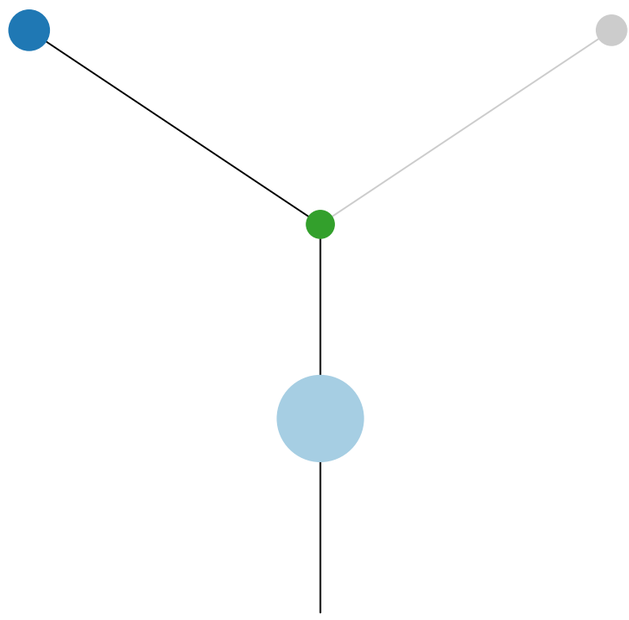
R1

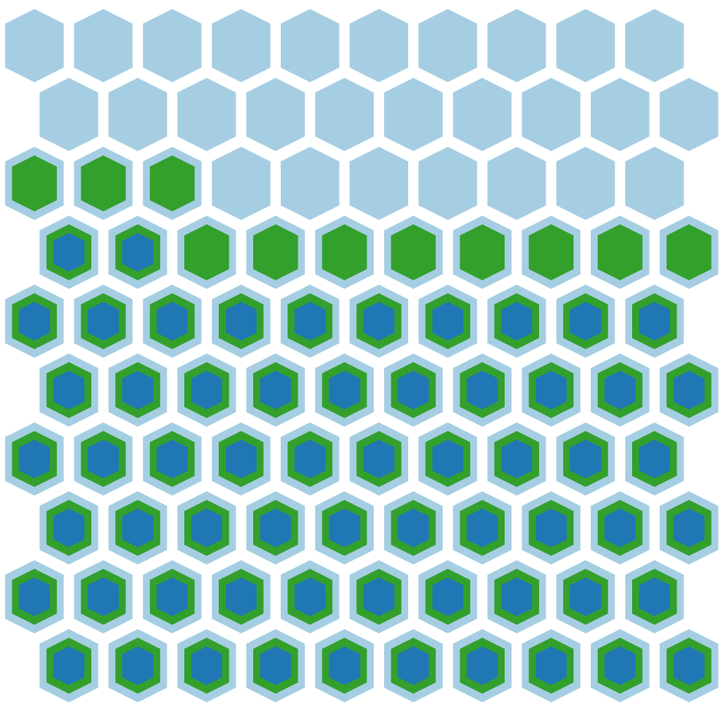


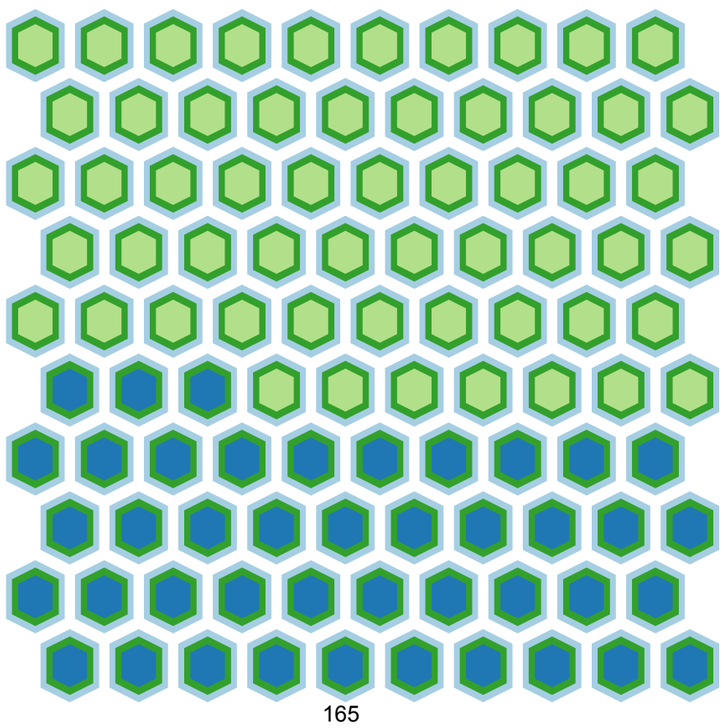
R2



R3







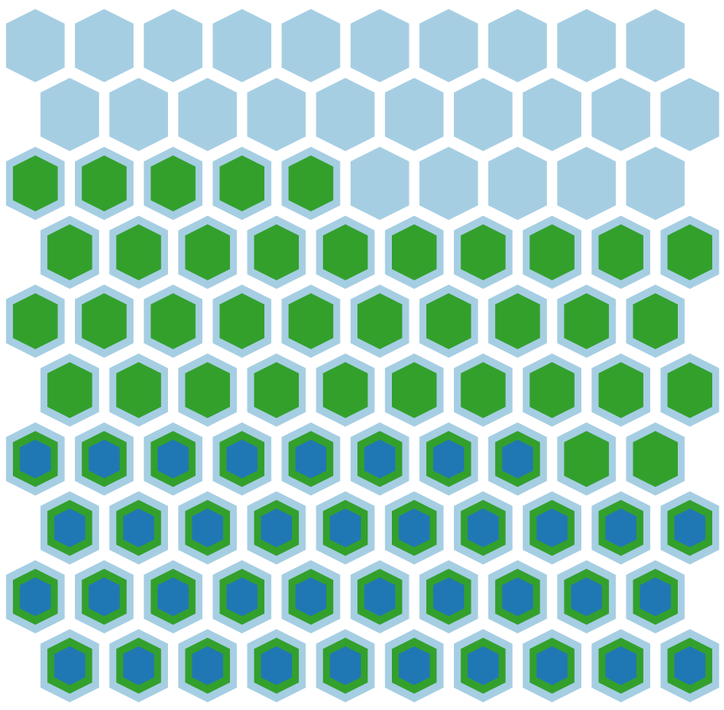
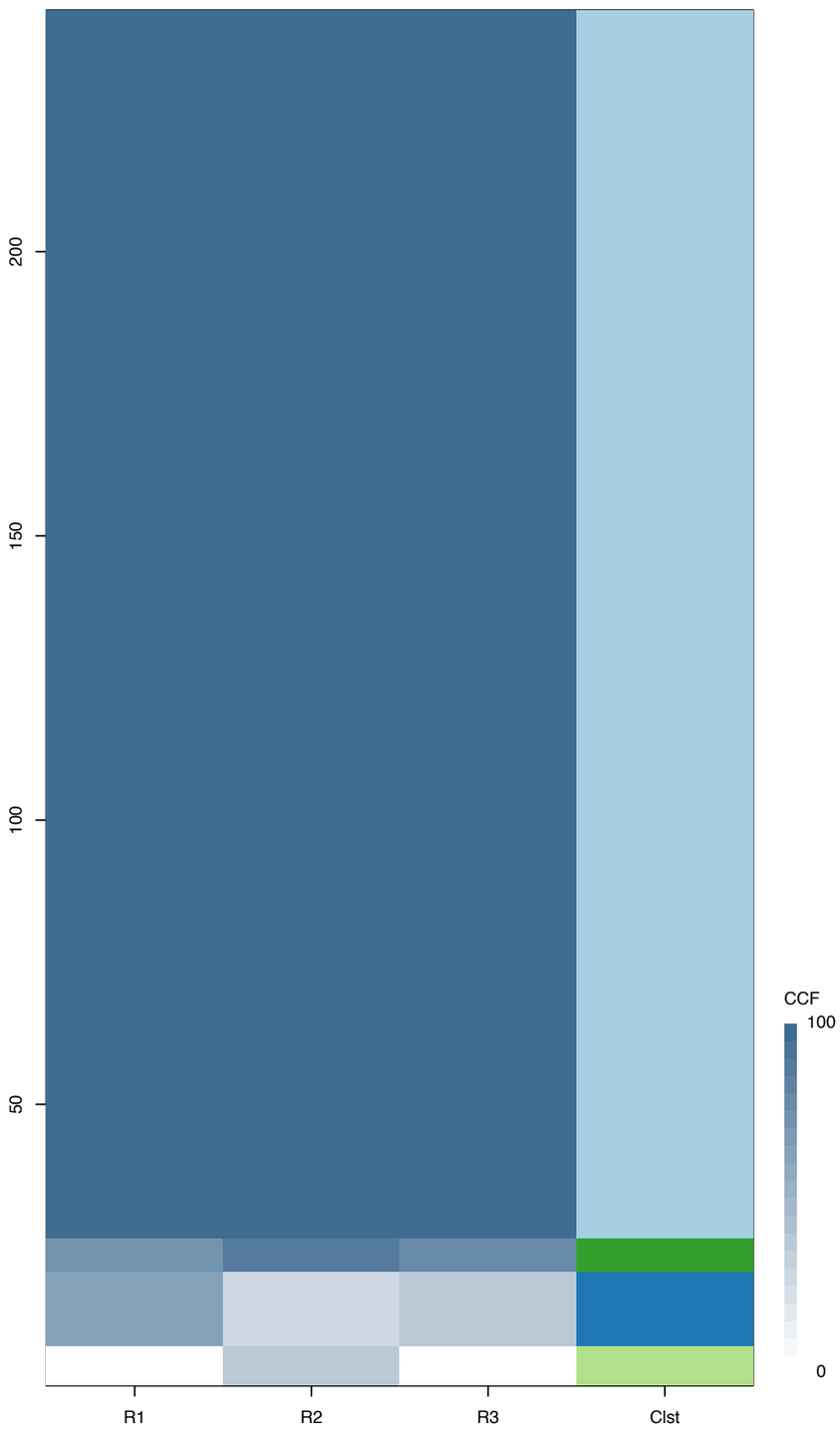
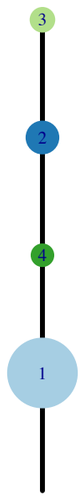


Fig.S12CG



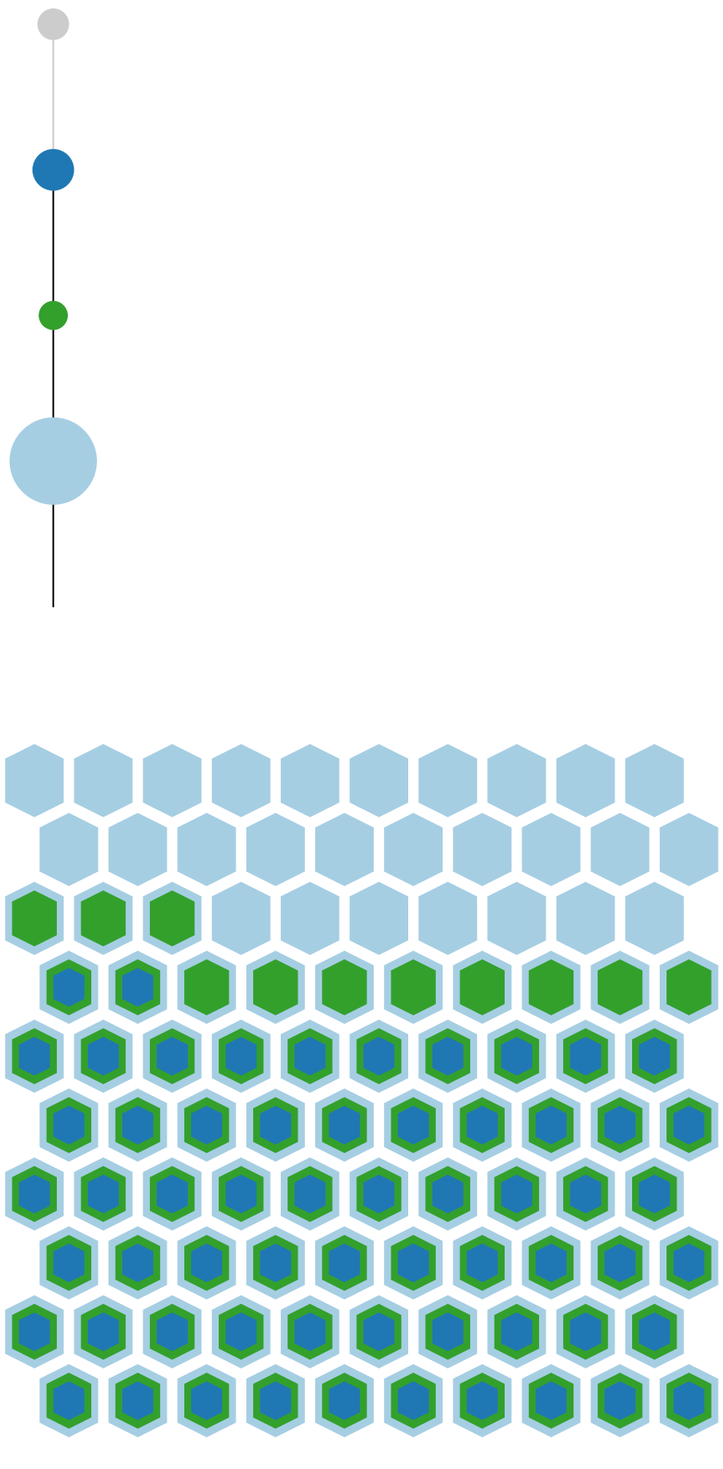
CRUK0095_B



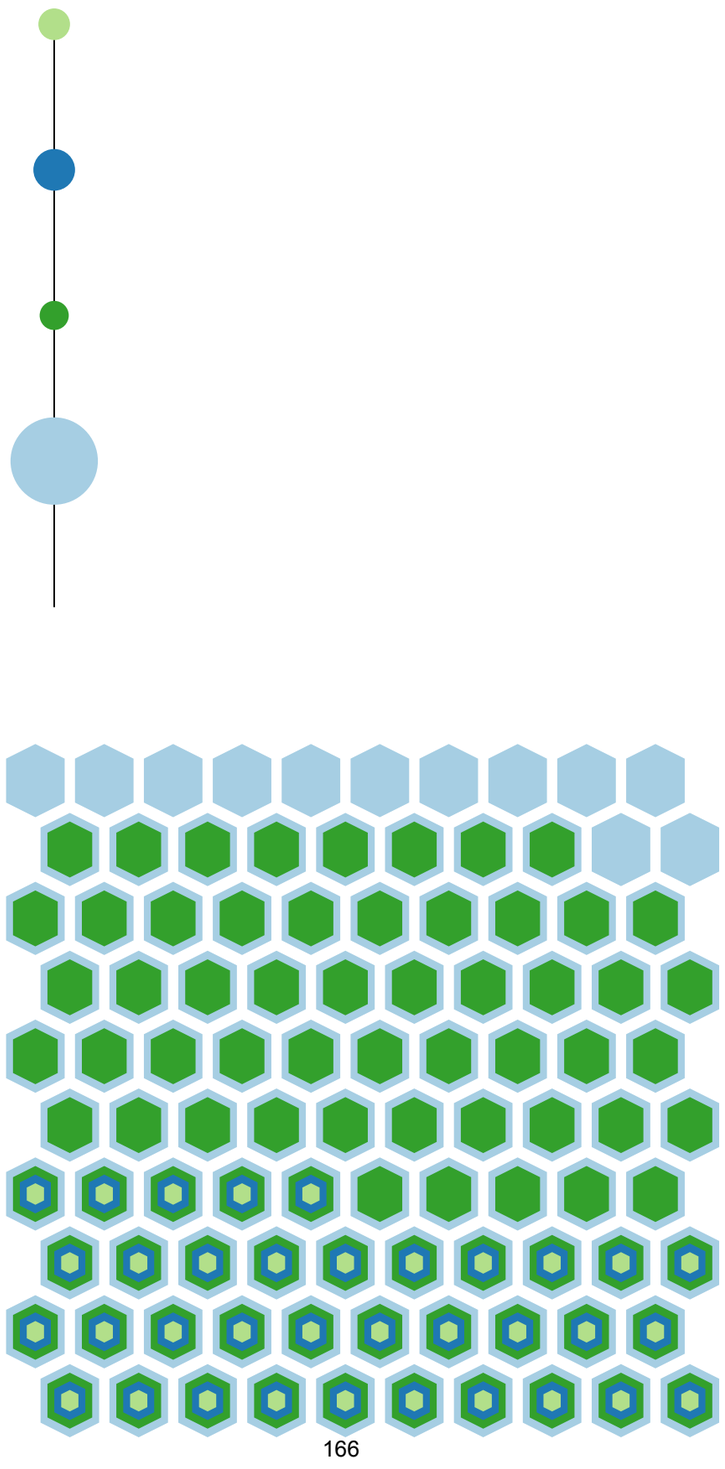
Gene	Cluster	Cytoband	Type
RASA1	1	5q14.3	SNV
TP53	1	17p13.1	SNV
NF1	1	17q11.2	SNV

Histology:Other, Age:60, PackYears:22.8, Size:20
Stage:1a, Gender:Female, GD:Not GD, Recur:no

R1



R2



R3

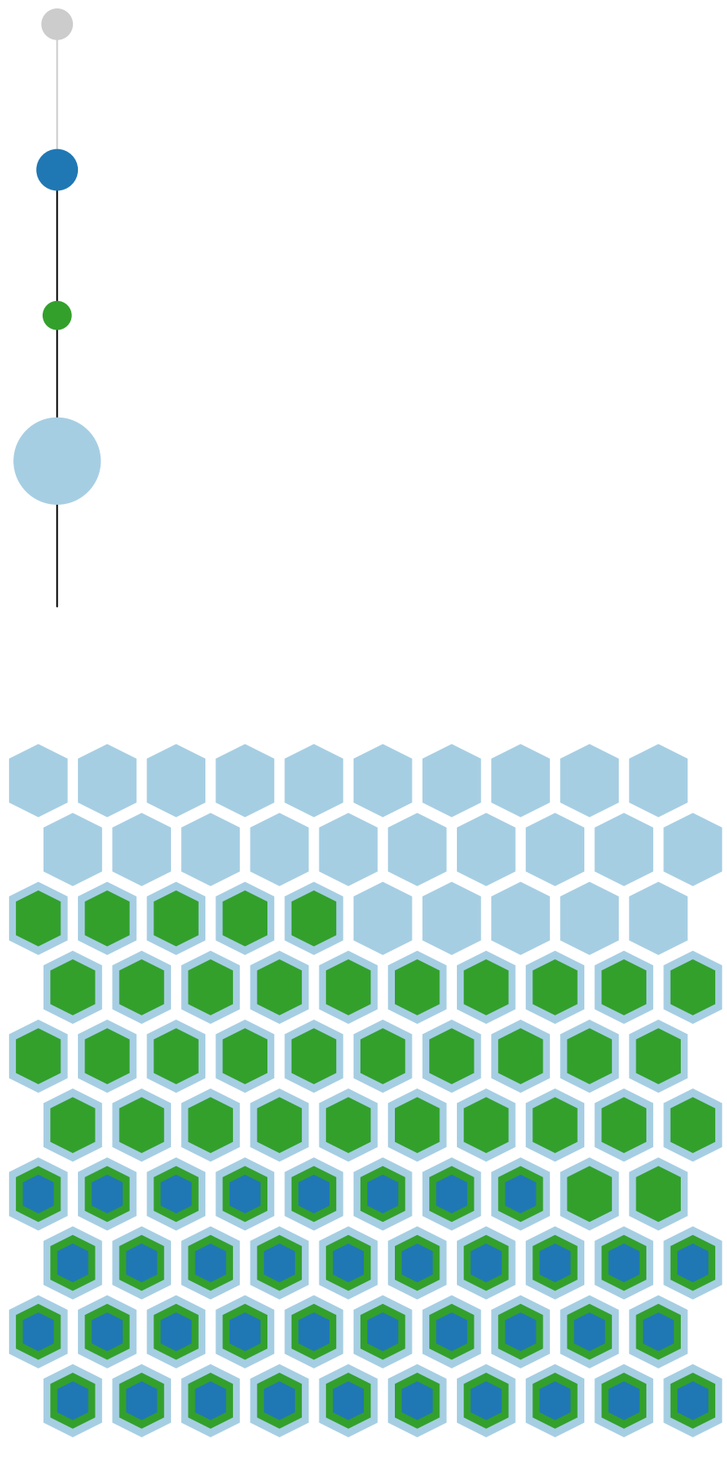
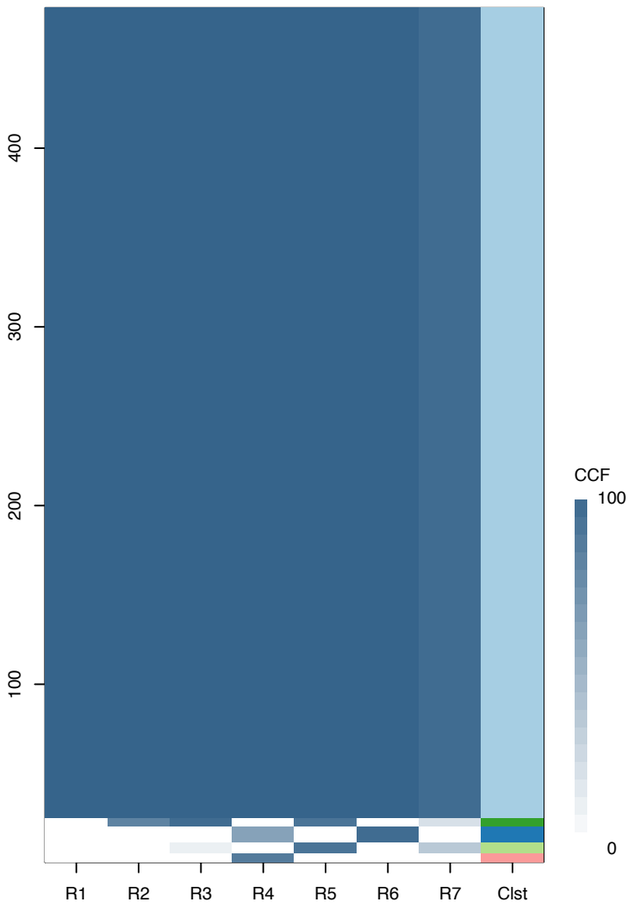
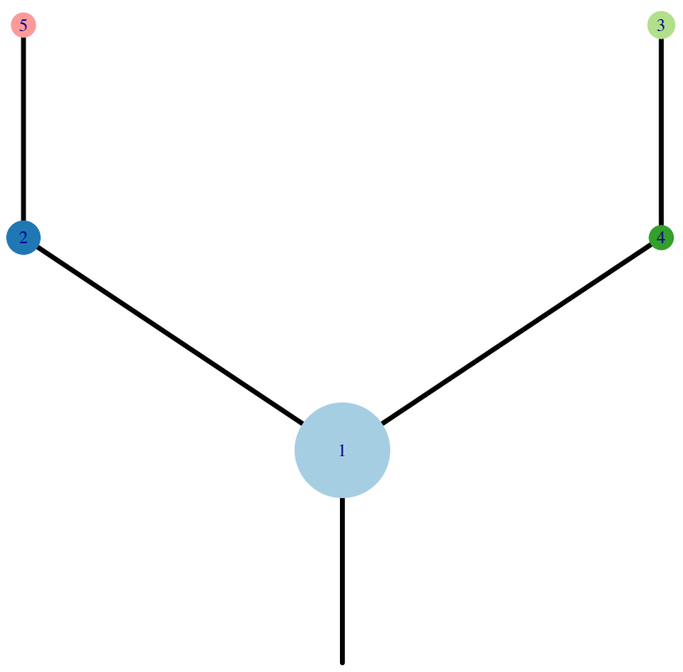


Fig.S12CH



CRUK0096



Histology:Other, Age:75, PackYears:53, Size:85
Stage:2b, Gender:Female, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
MAP3K1	1	5q11.2	SNV
SGK223	1	8p23.1	SNV
KRAS	1	12p12.1	SNV

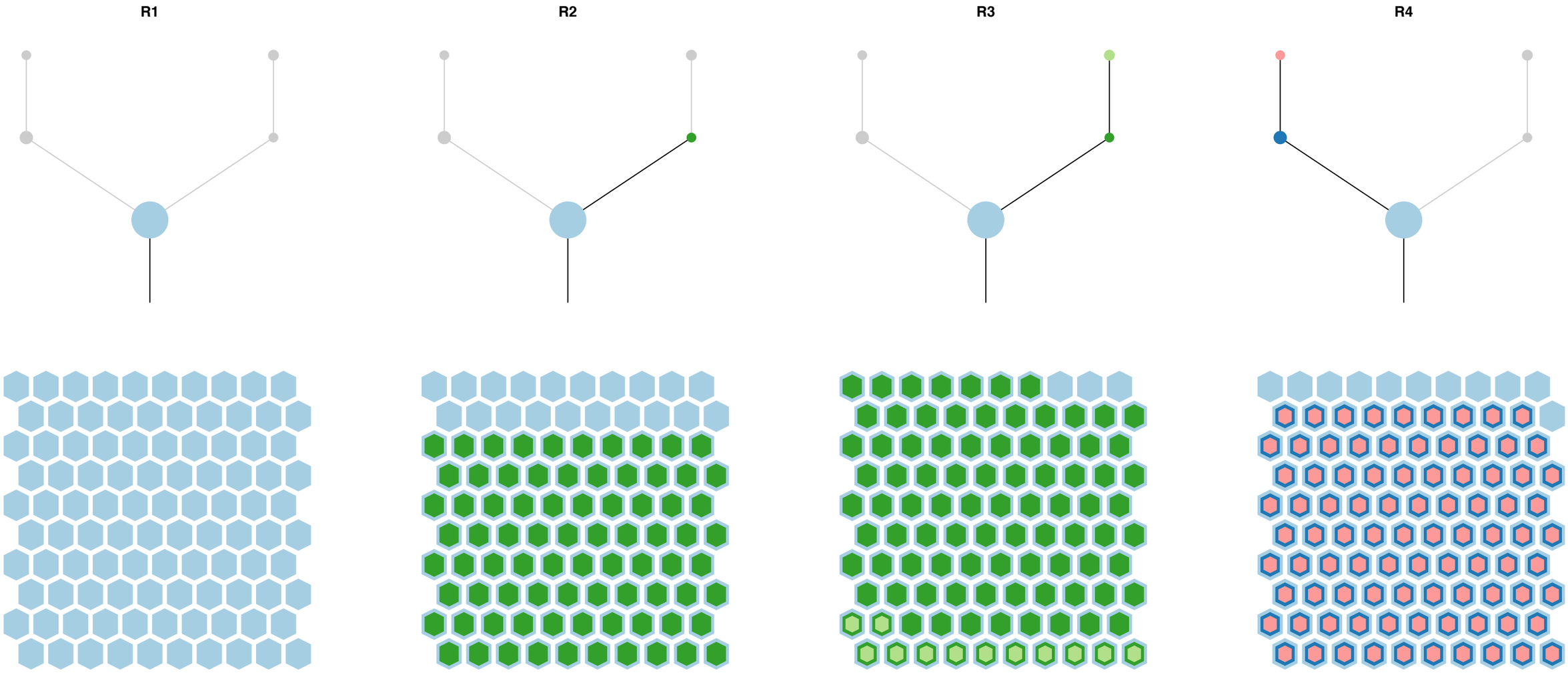
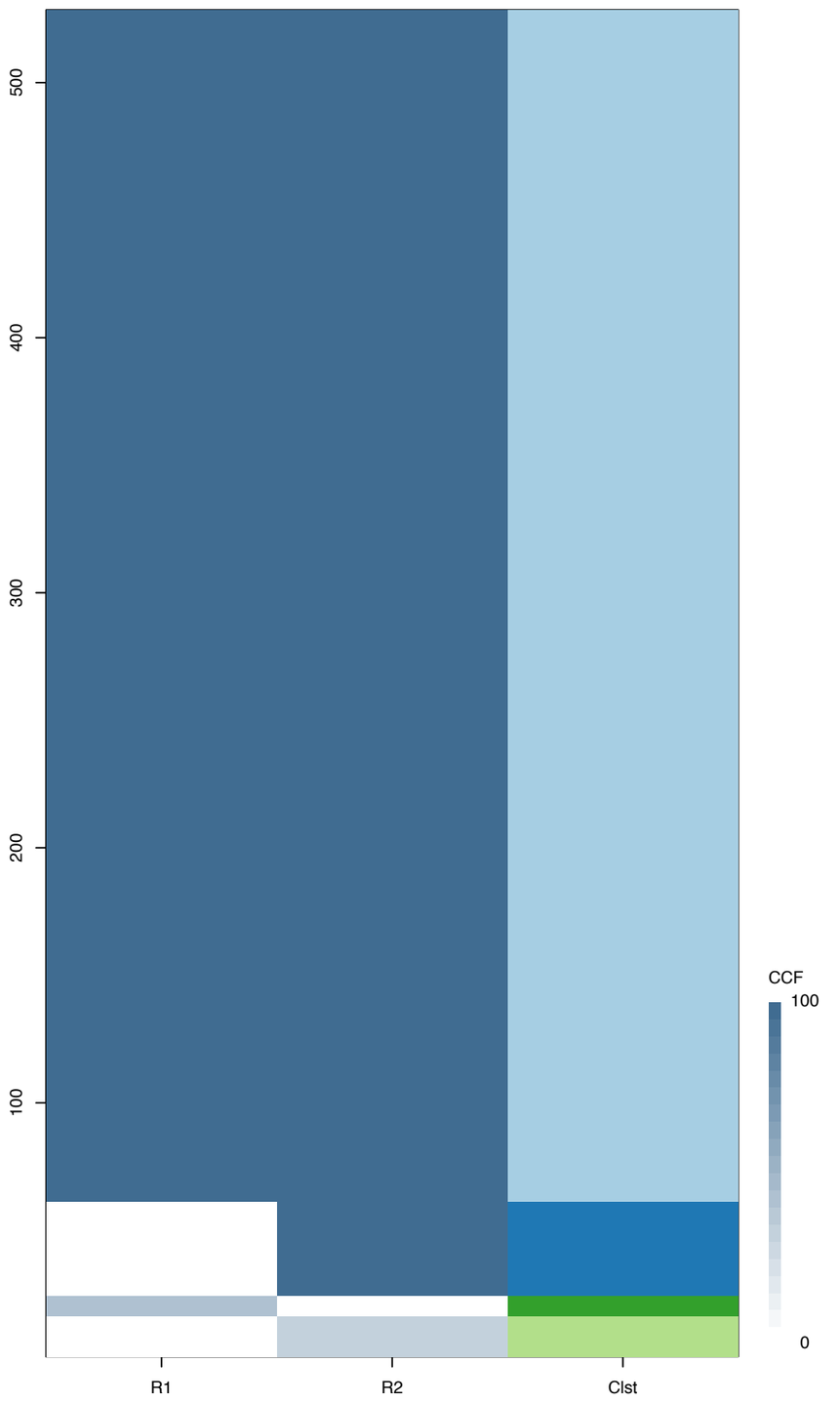
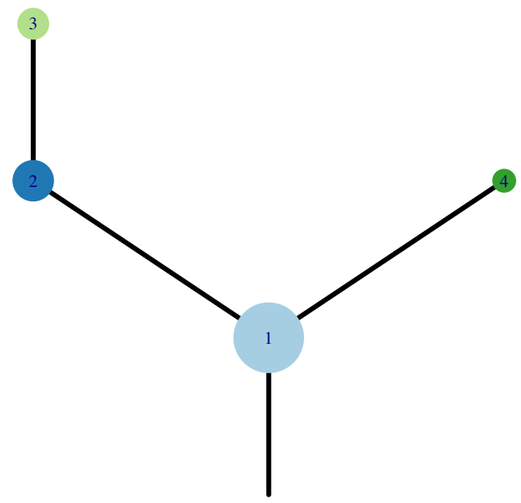


Fig.S12CI



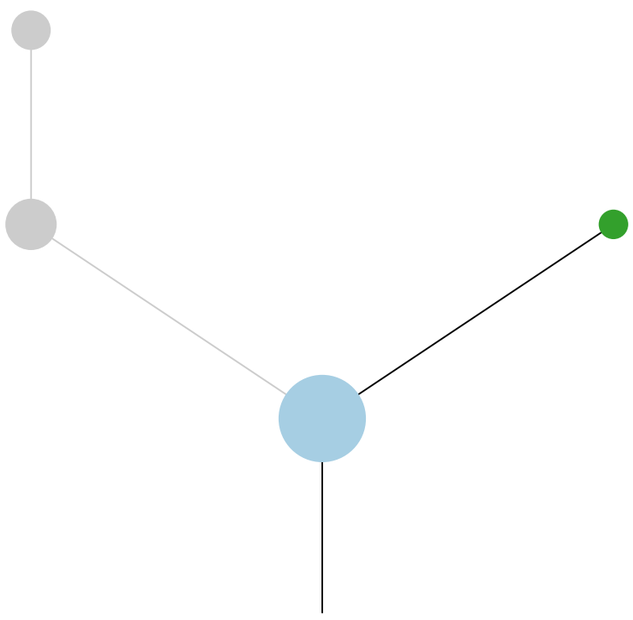
CRUK0097



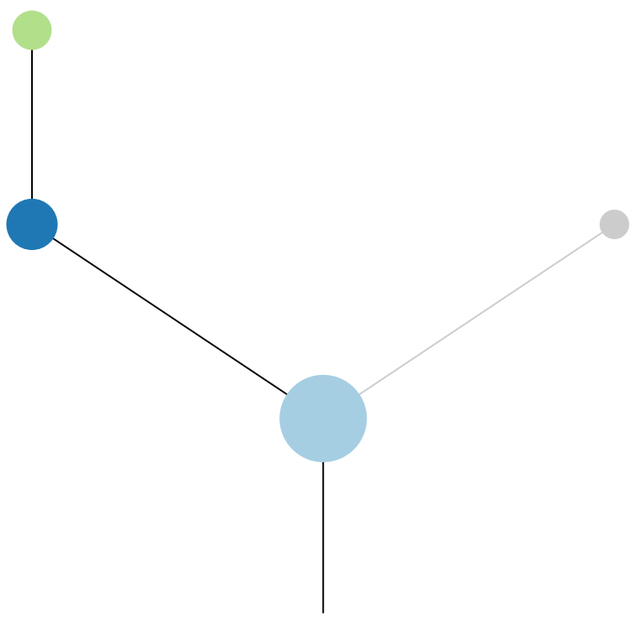
Histology:Other, Age:77, PackYears:118, Size:25
Stage:1b, Gender:Male, GD:Clonal GD, Recur:no

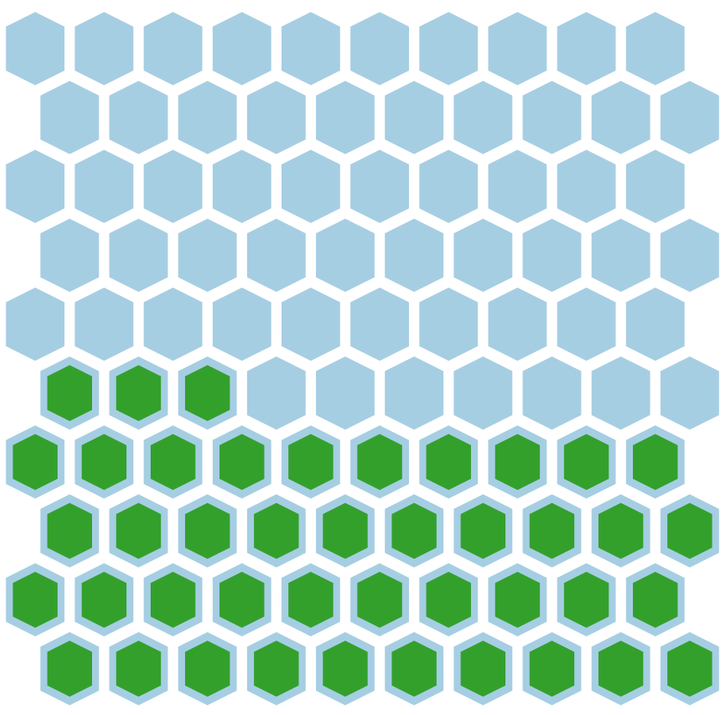
Gene	Cluster	Cytoband	Type
PTEN	1	10q23.31	SNV
TP53	1	17p13.1	SNV

R1



R2





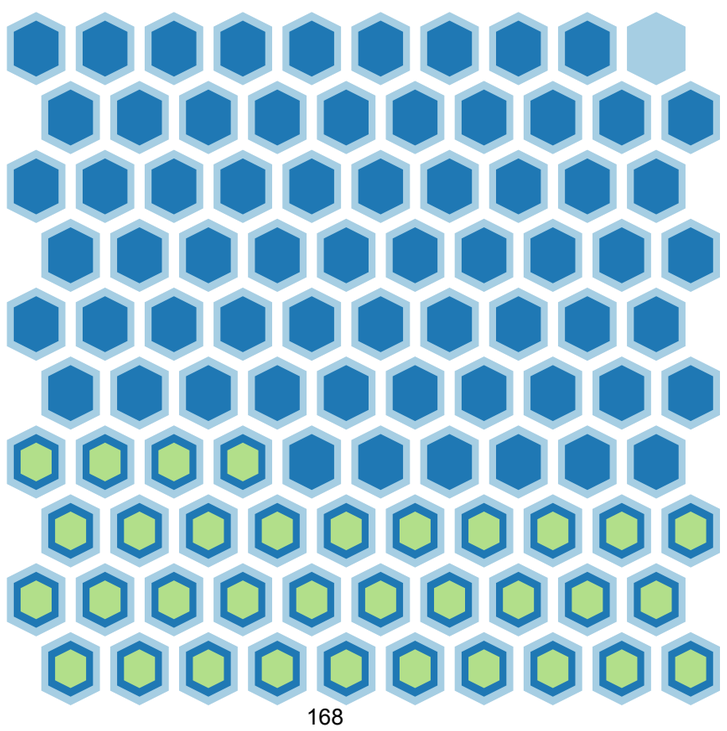
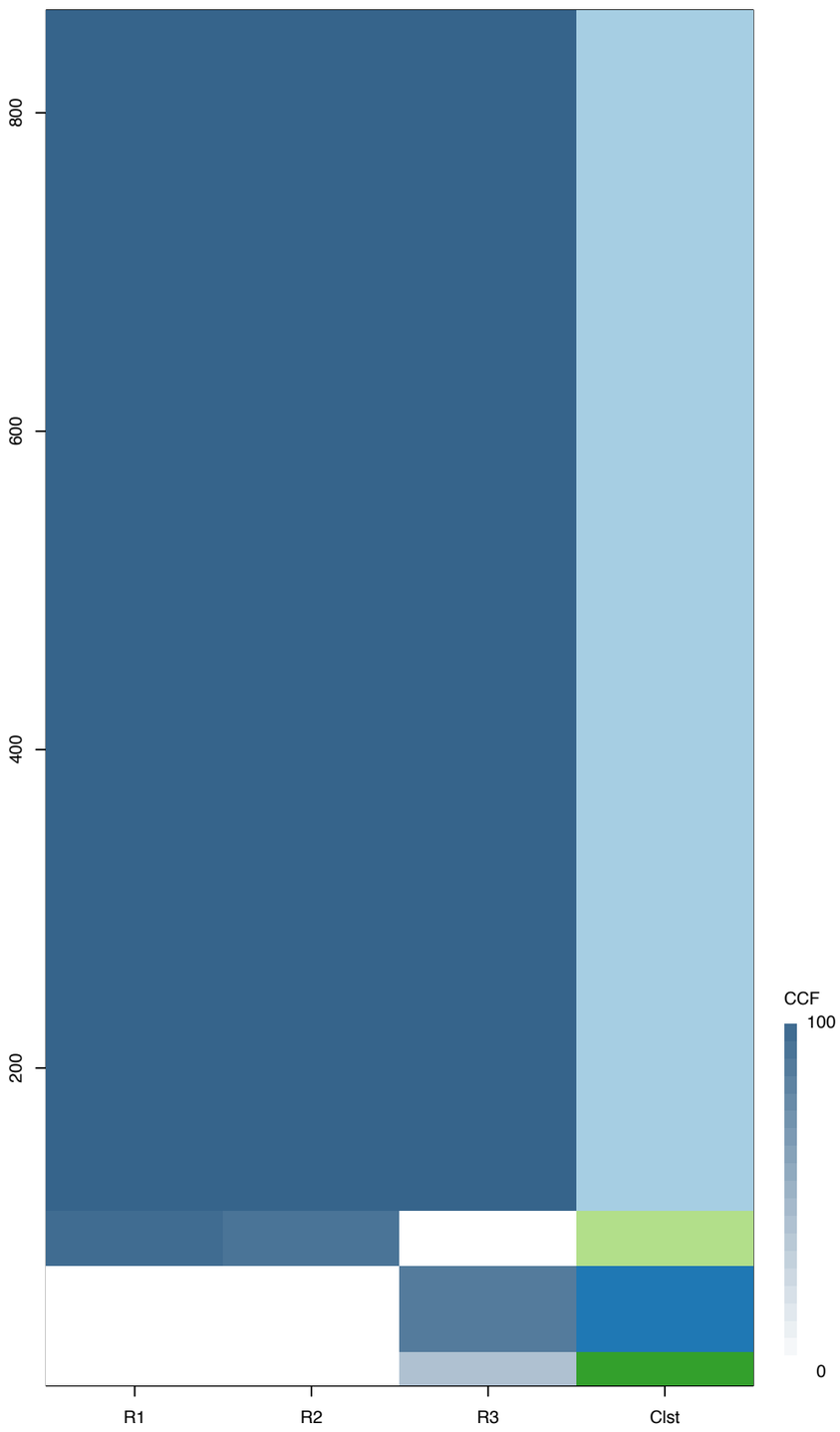
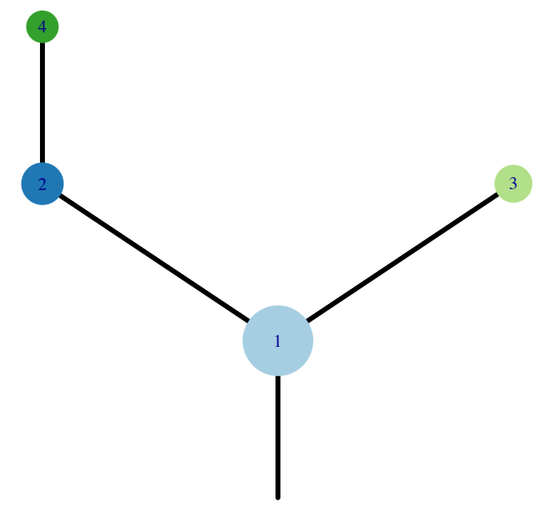


Fig.S12CJ



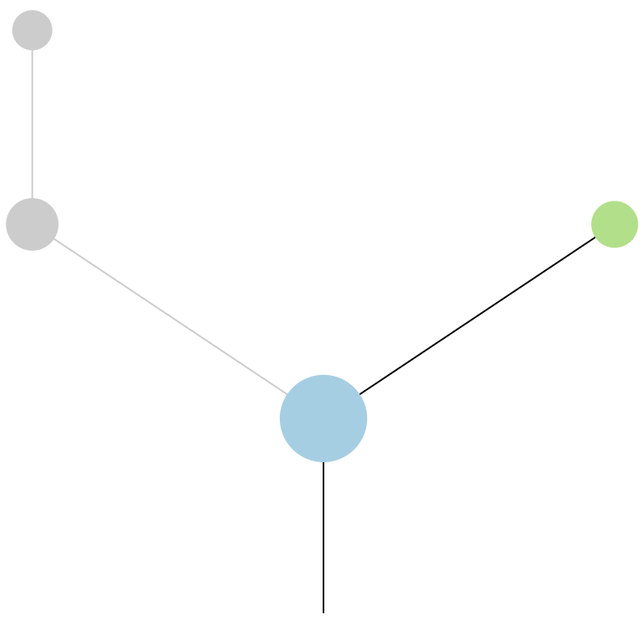
CRUK0098



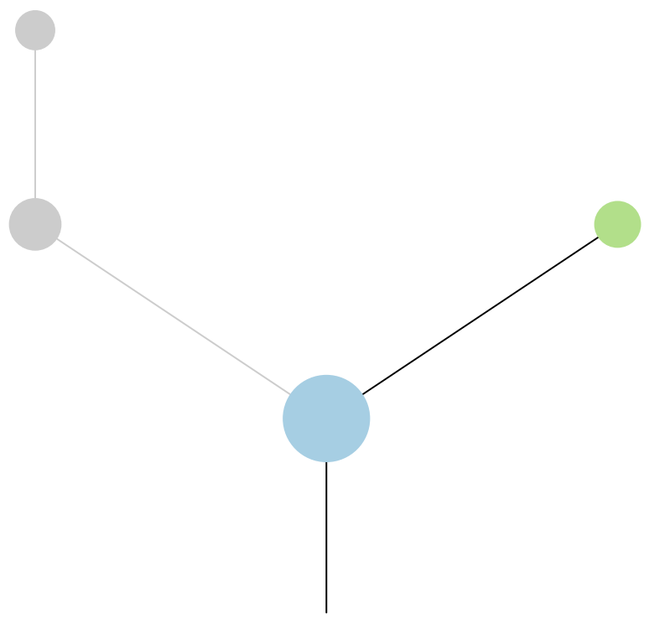
Histology:Other, Age:56, PackYears:48.75, Size:42
Stage:1b, Gender:Male, GD:Clonal GD, Recur:yes

Gene	Cluster	Cytoband	Type
PTEN	1	10q23.31	Del
FAS	1	10q23.31	Del
TP53	1	17p13.1	SNV
UBR5	3	8q22.3	SNV
HIST1H3B	?	6p22.2	Amp
TFPT	?	19q13.42	Amp

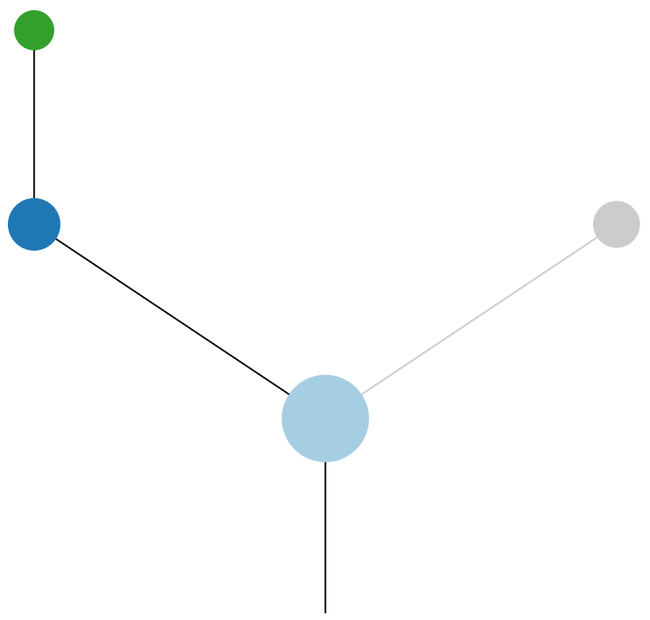
R1

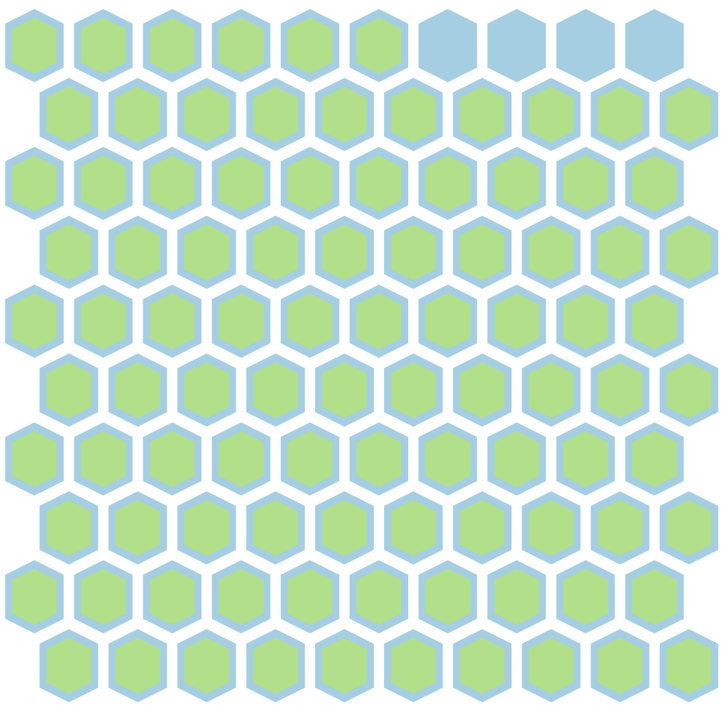


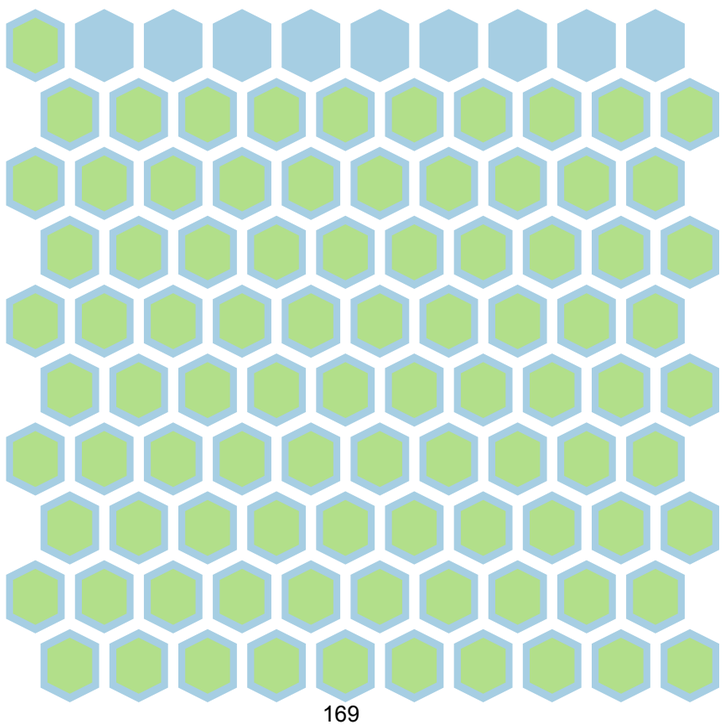
R2



R3







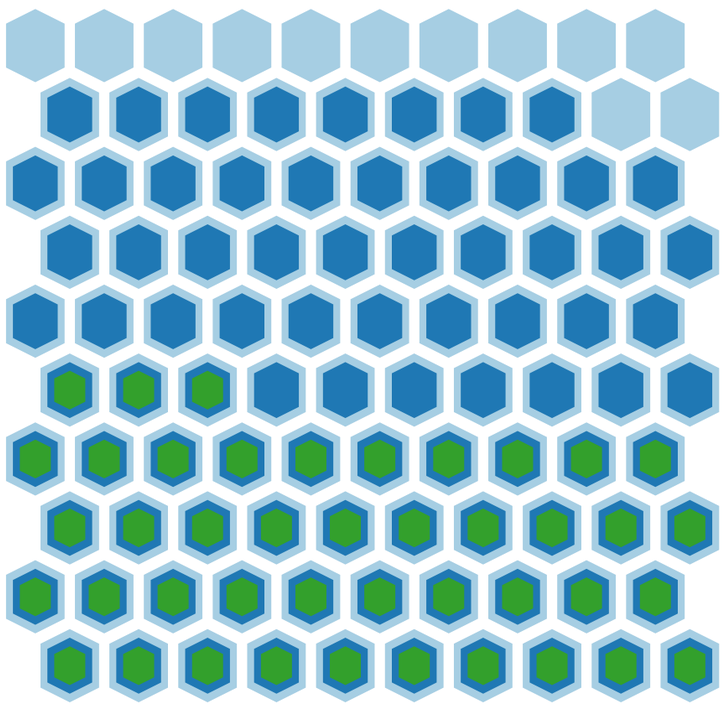
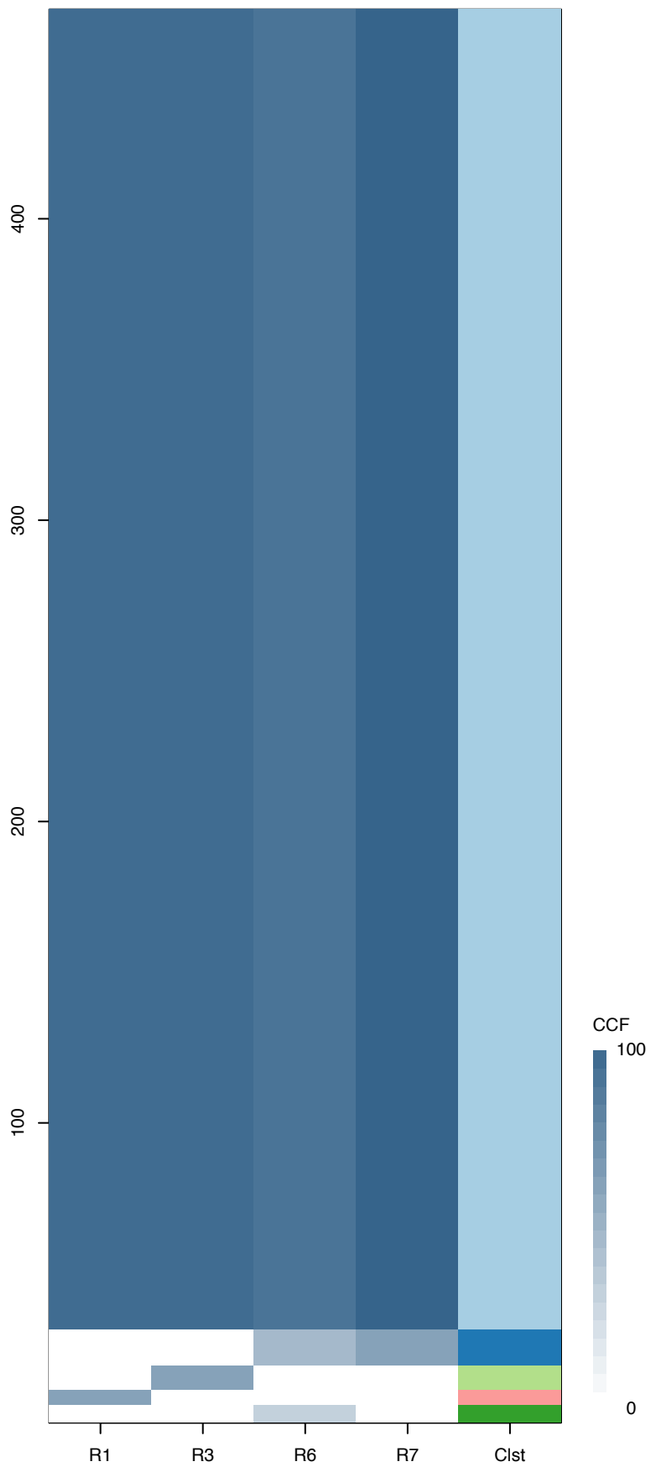
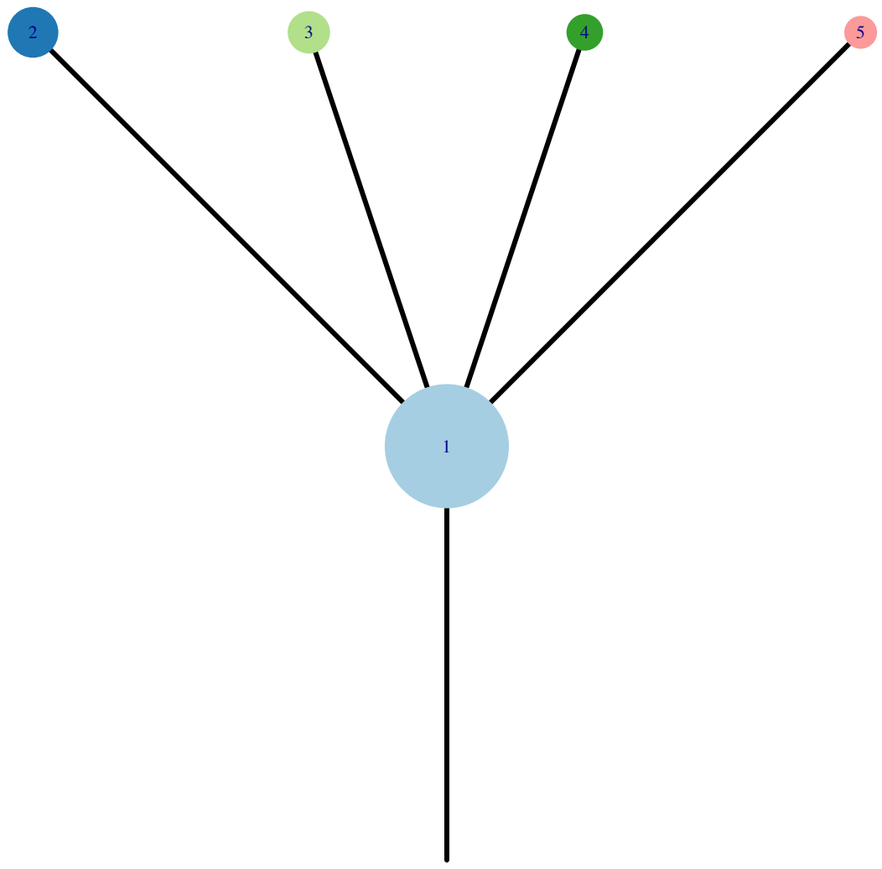


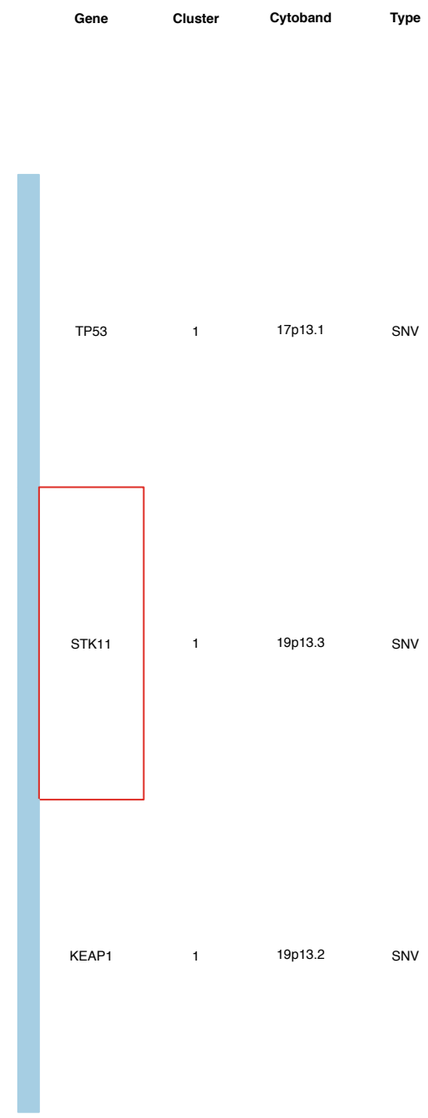
Fig.S12CK

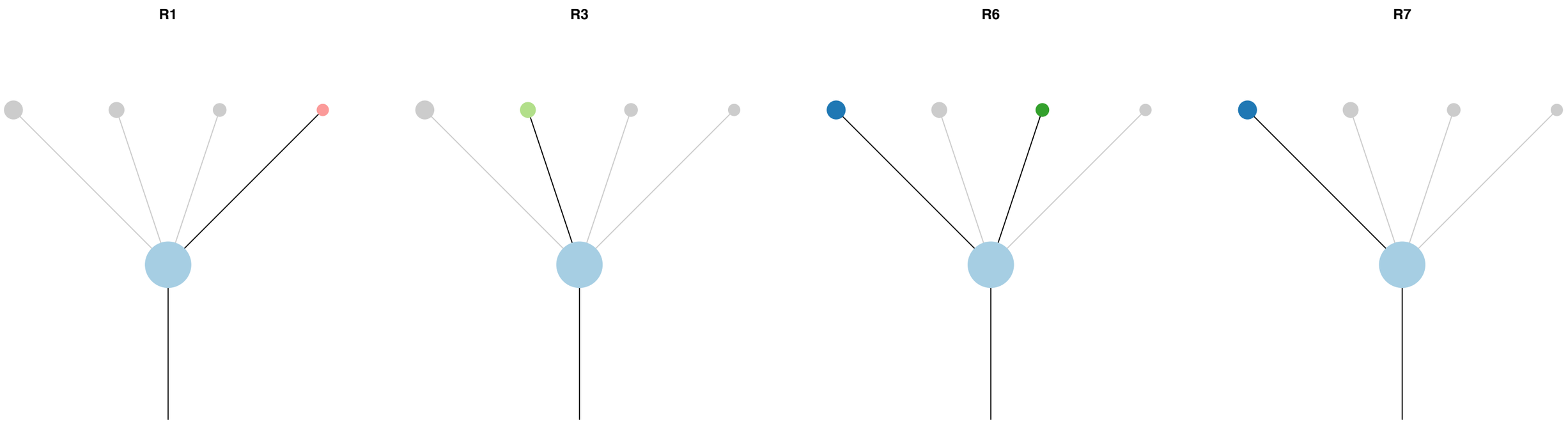


CRUK0099_A



Histology:Other, Age:78, PackYears:48, Size:72
Stage:3a, Gender:Male, GD:Not GD, Recur:yes





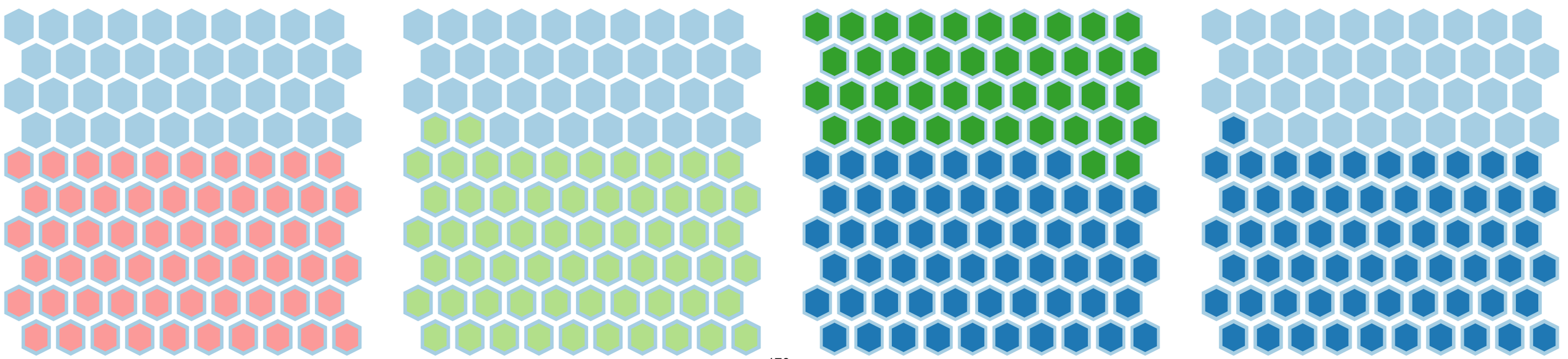
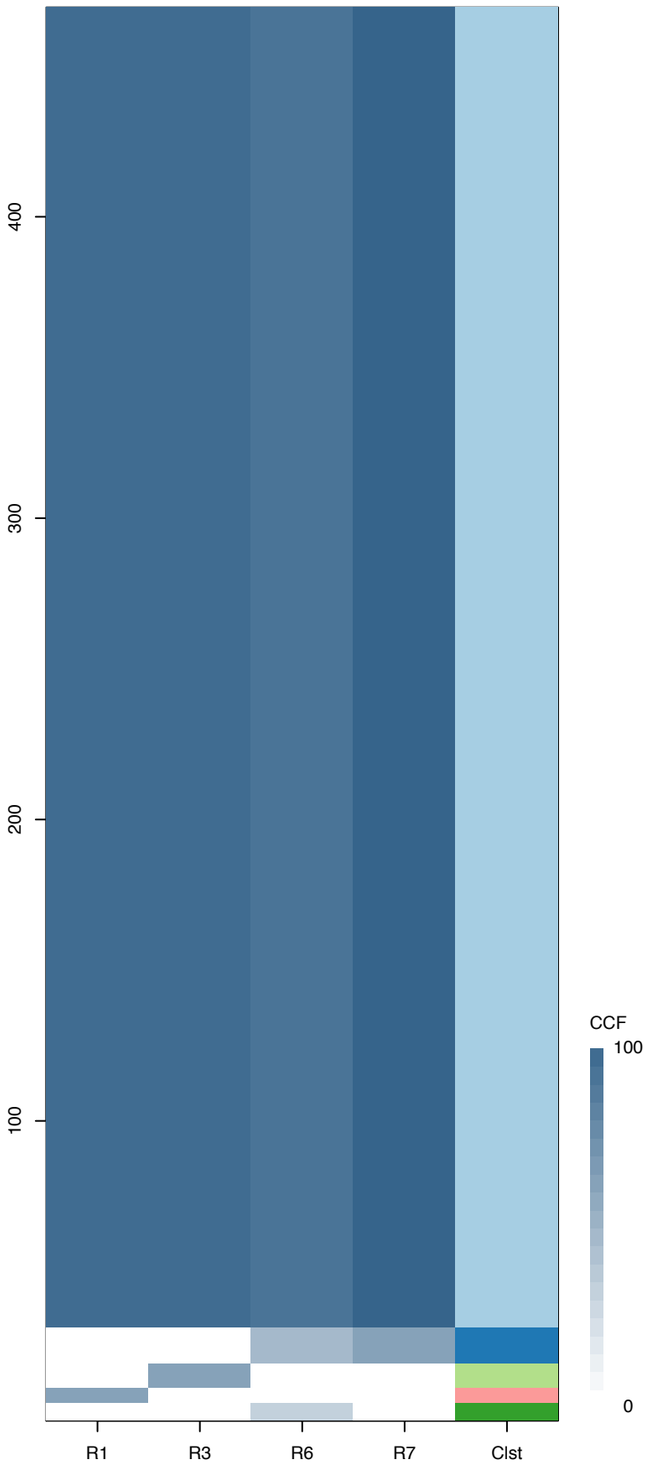
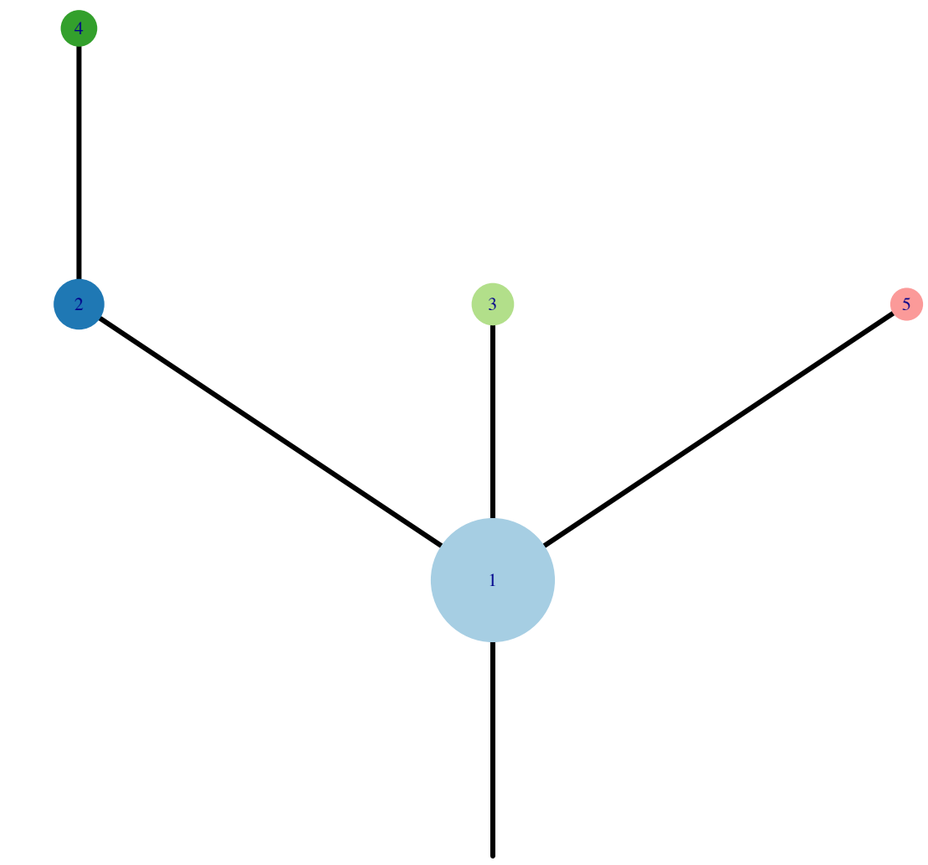


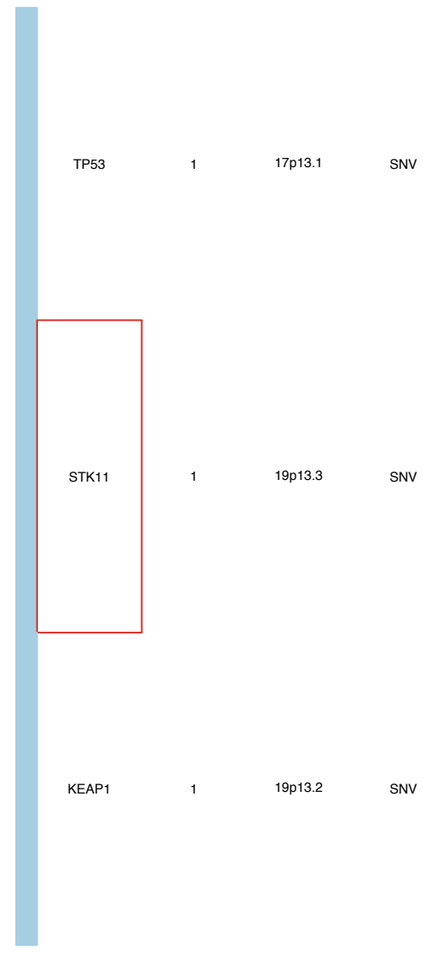
Fig.S12CK



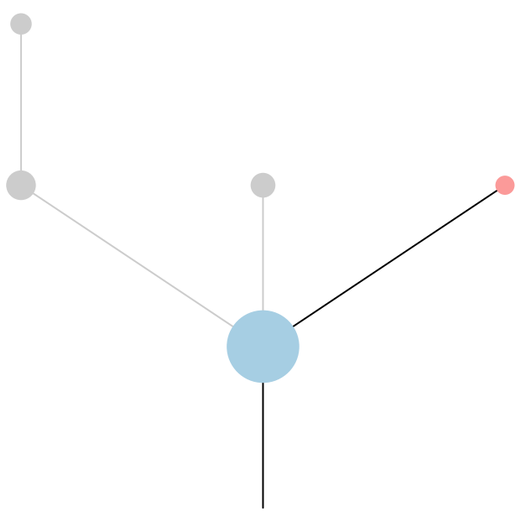
CRUK0099_B



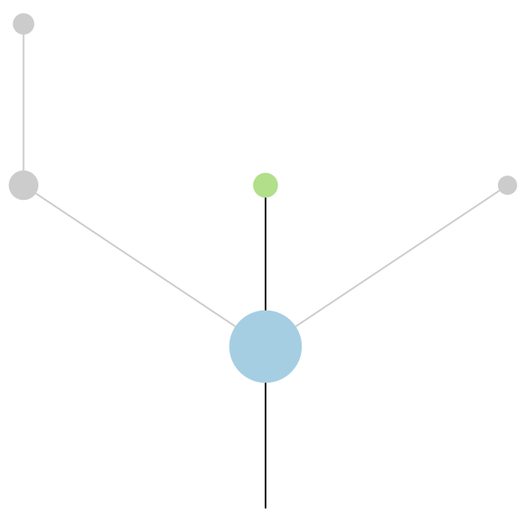
Histology:Other, Age:78, PackYears:48, Size:72
Stage:3a, Gender:Male, GD:Not GD, Recur:yes



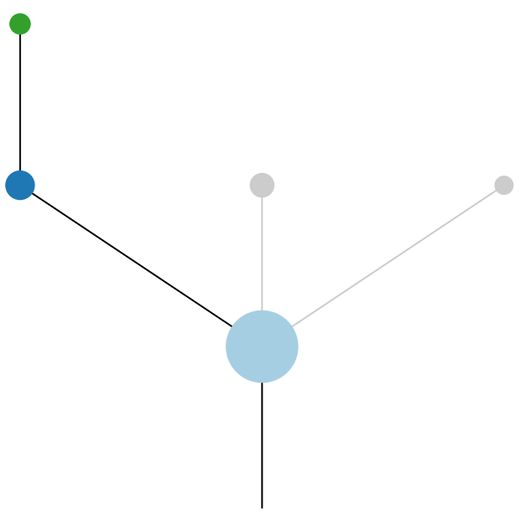
R1



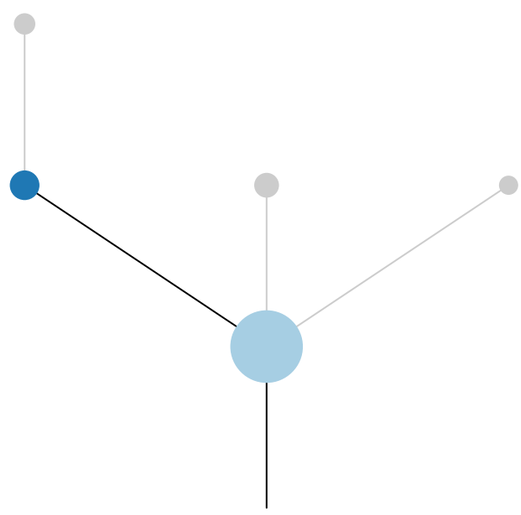
R3



R6



R7



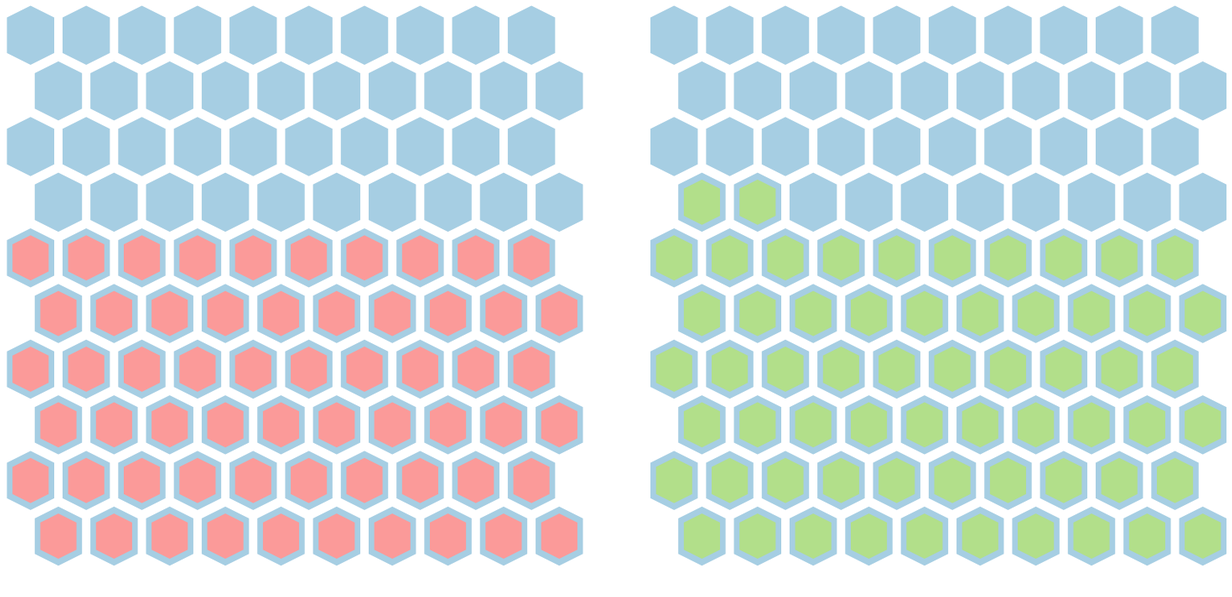
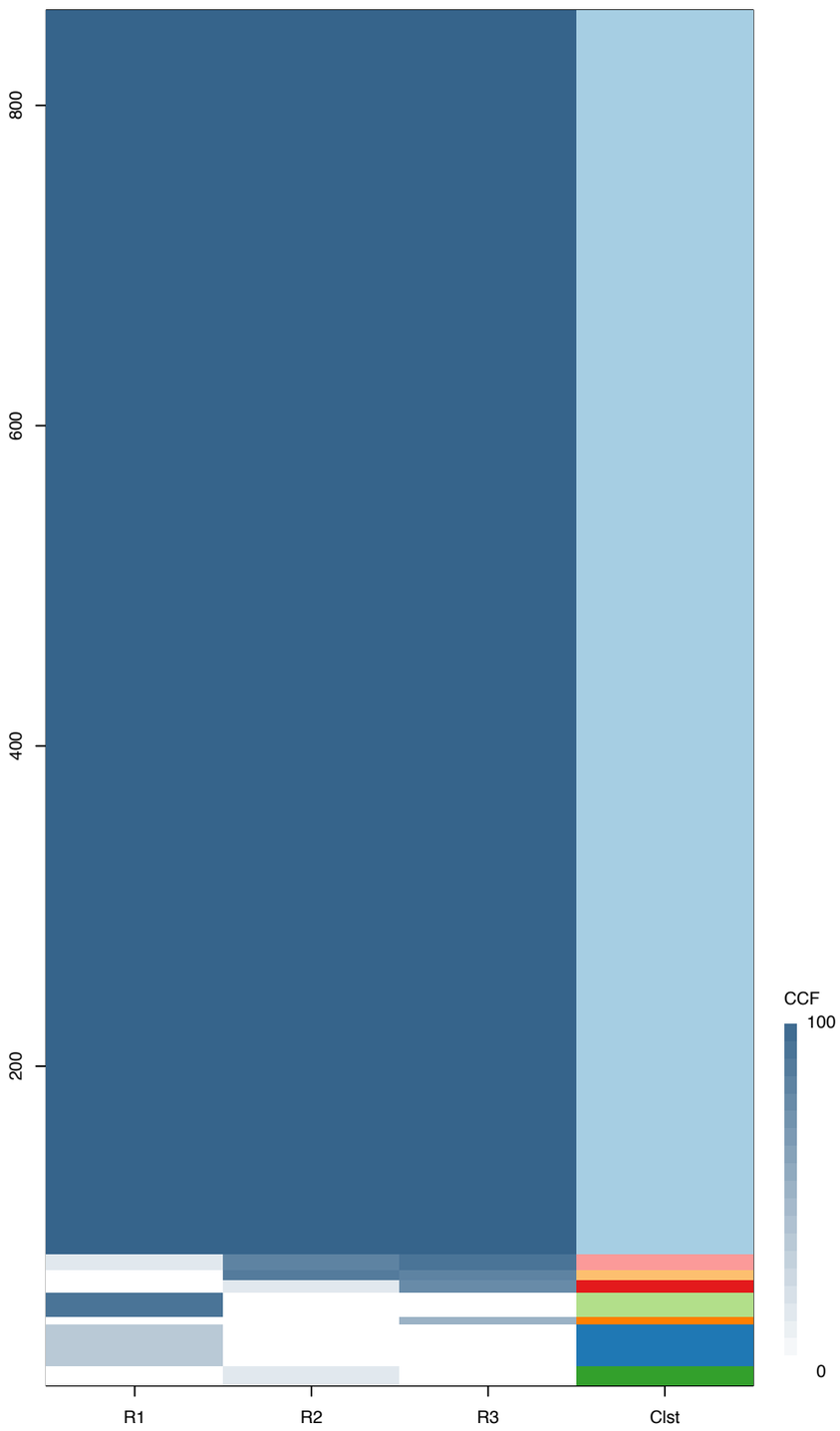
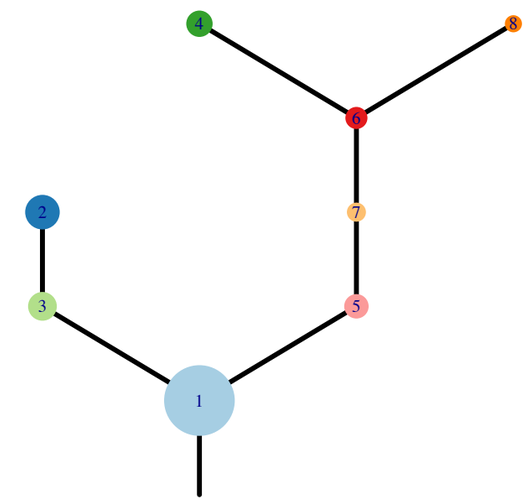


Fig.S12CL



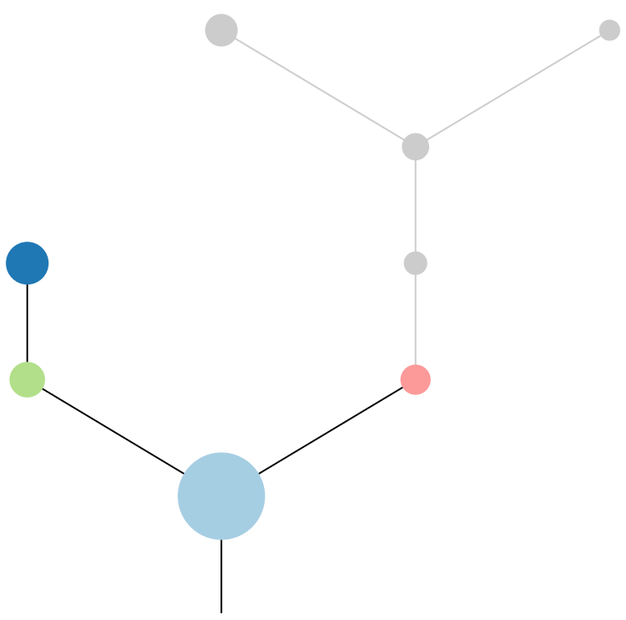
CRUK0100_A

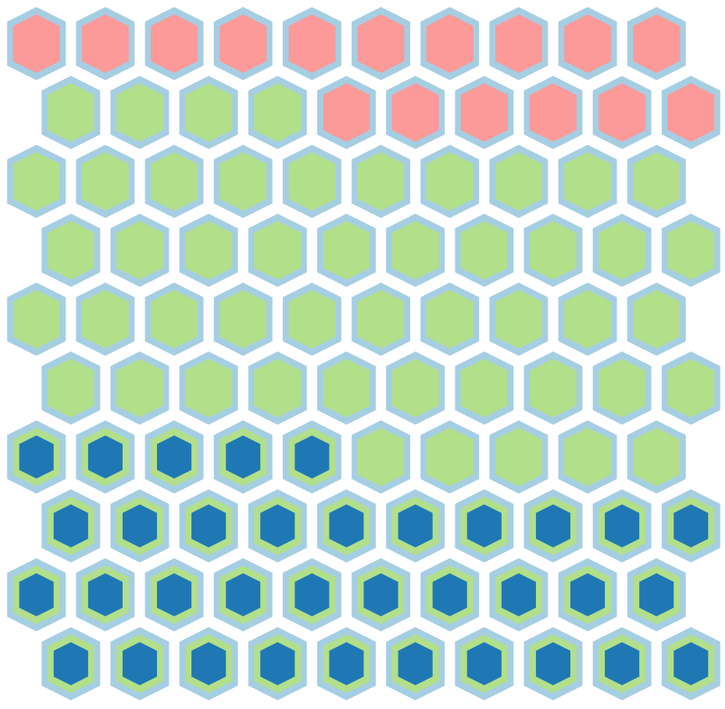


Gene	Cluster	Cytoband	Type
SPEN	1	1p36.13	SNV
COL5A2	1	2q32.2	SNV
PHOX2B	1	4p13	SNV
CDKN2A	1	9p21.3	Del
TP53	1	17p13.1	SNV
STK11	1	19p13.3	SNV
GOLGA5	6	14q32.12	Amp
ZNF521	6	18q11.2	Amp
SS18	6	18q11.2	Amp
AMER1	?		SNV

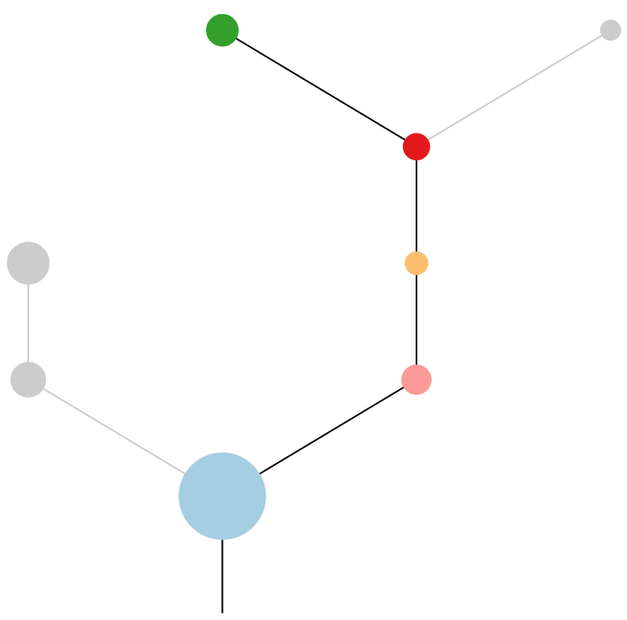
Histology:Other, Age:50, PackYears:25, Size:29
Stage:3a, Gender:Female, GD:Clonal GD, Recur:yes

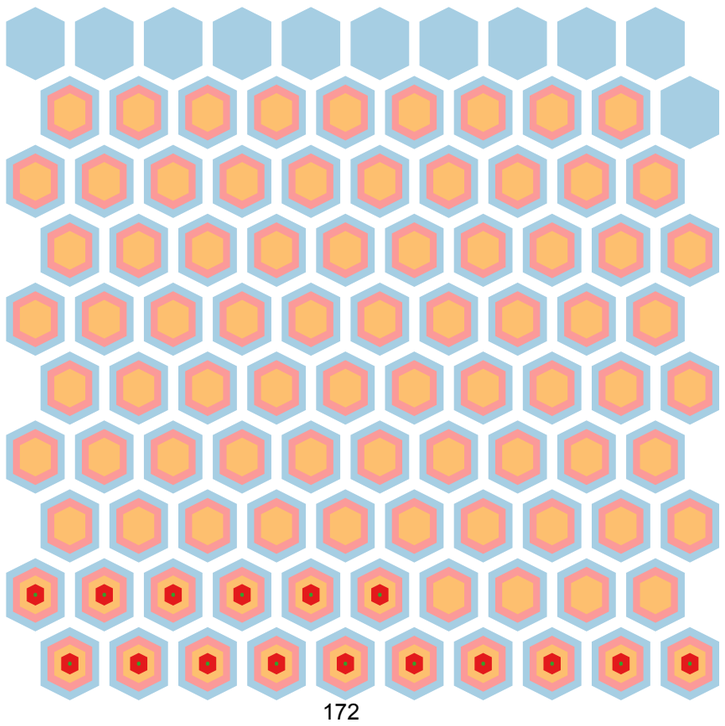
R1



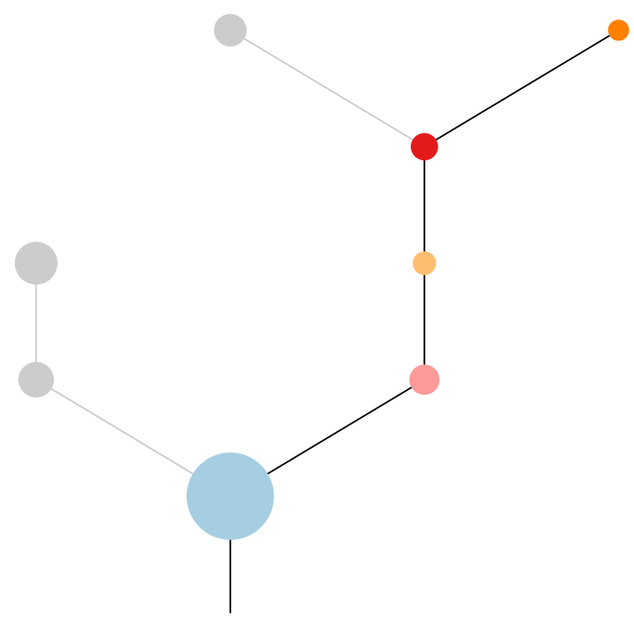


R2





R3



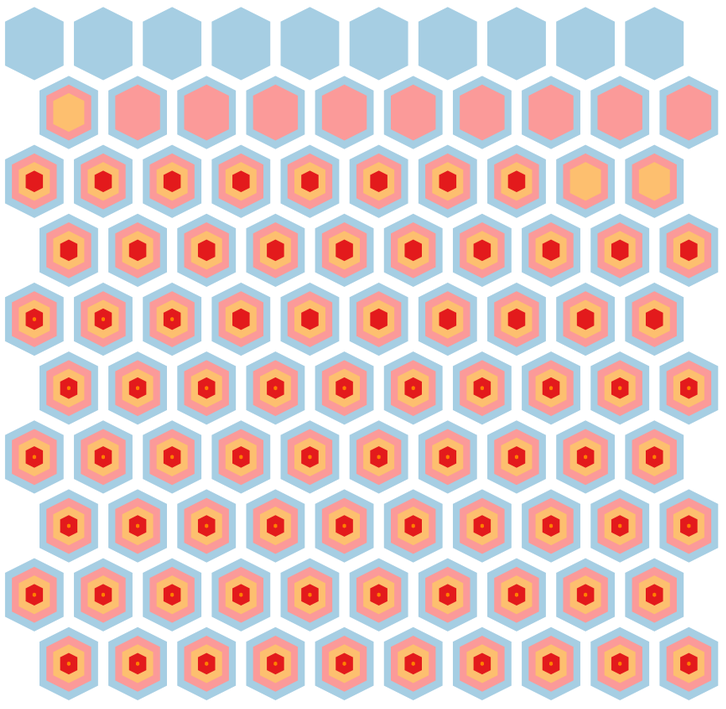
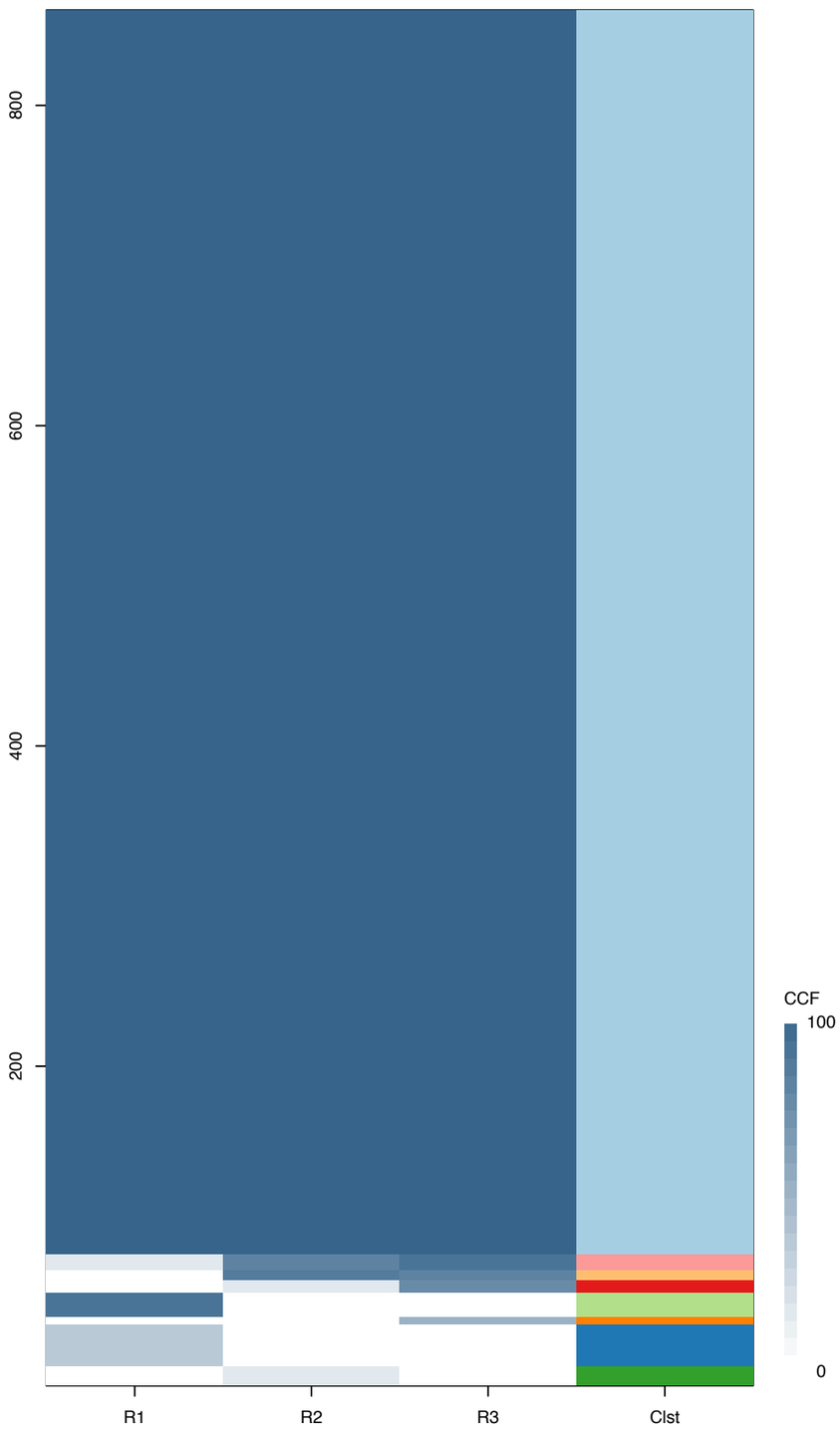
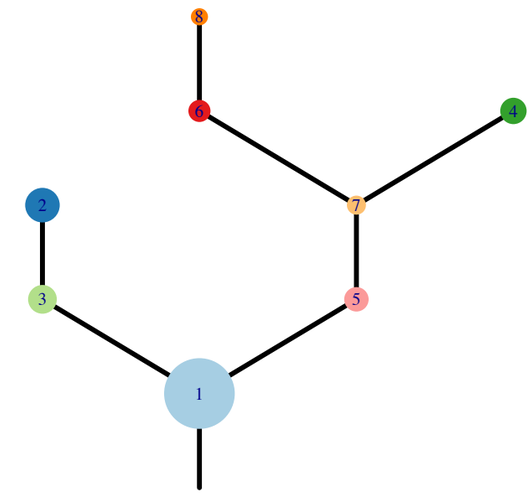


Fig.S12CL



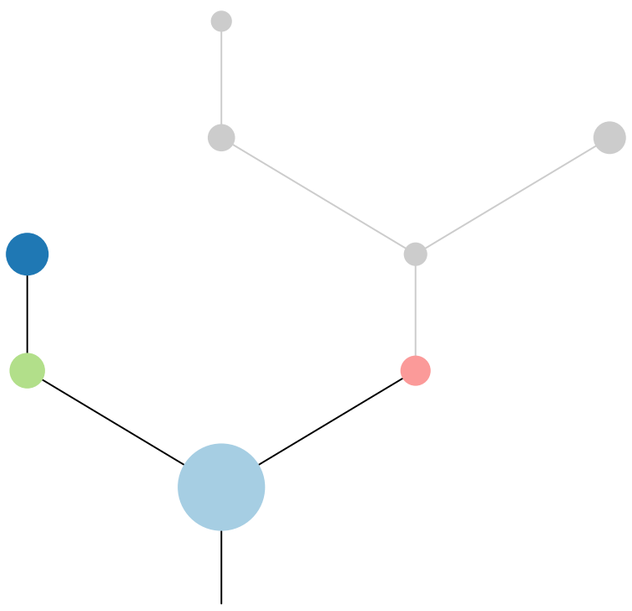
CRUK0100_B



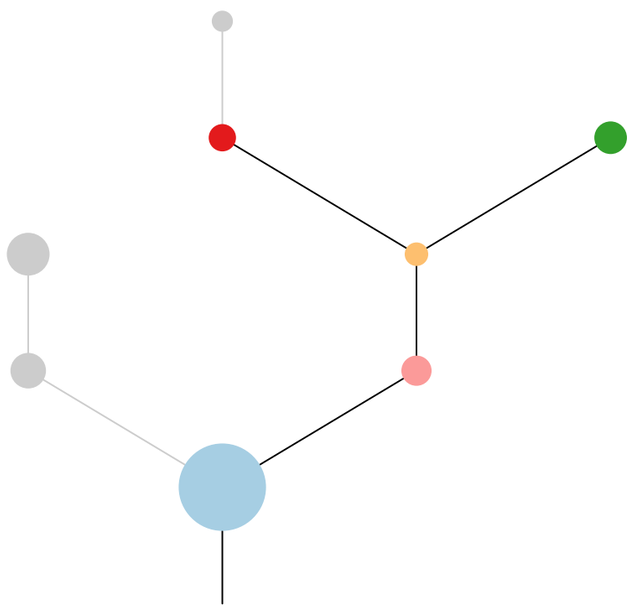
Gene	Cluster	Cytoband	Type
SPEN	1	1p36.13	SNV
COL5A2	1	2q32.2	SNV
PHOX2B	1	4p13	SNV
CDKN2A	1	9p21.3	Del
TP53	1	17p13.1	SNV
STK11	1	19p13.3	SNV
GOLGA5	6	14q32.12	Amp
ZNF521	6	18q11.2	Amp
SS18	6	18q11.2	Amp
AMER1	?		SNV

Histology:Other, Age:50, PackYears:25, Size:29
Stage:3a, Gender:Female, GD:Clonal GD, Recur:yes

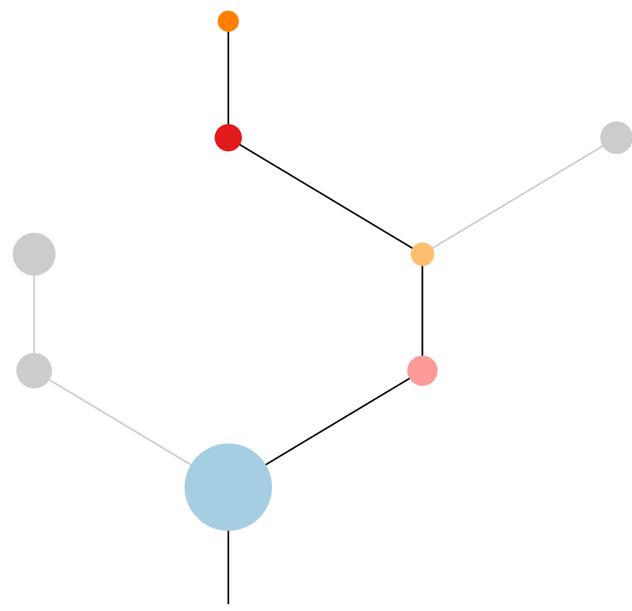
R1

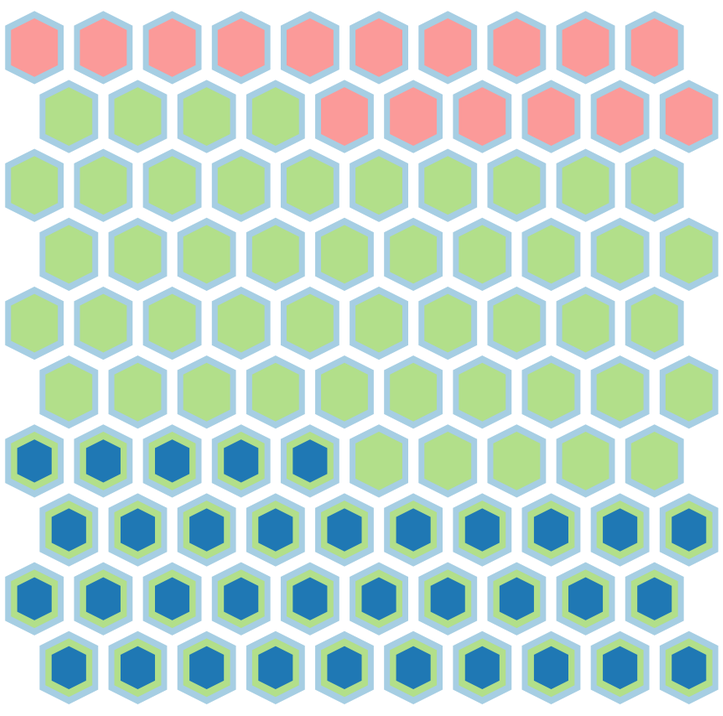


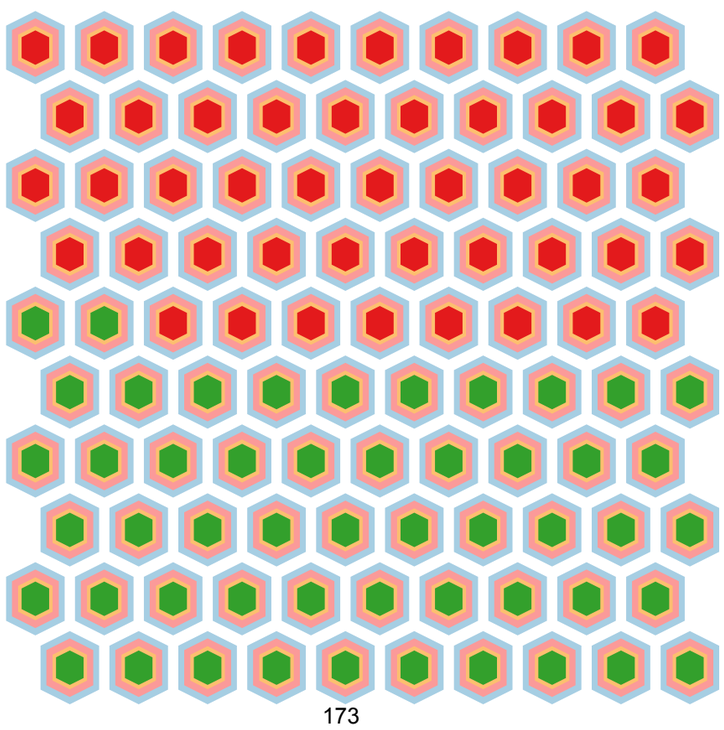
R2



R3







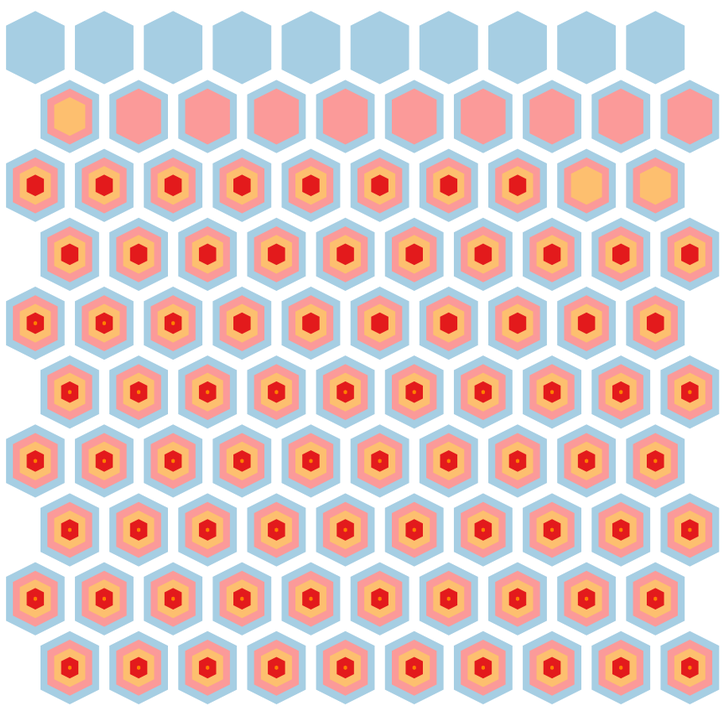
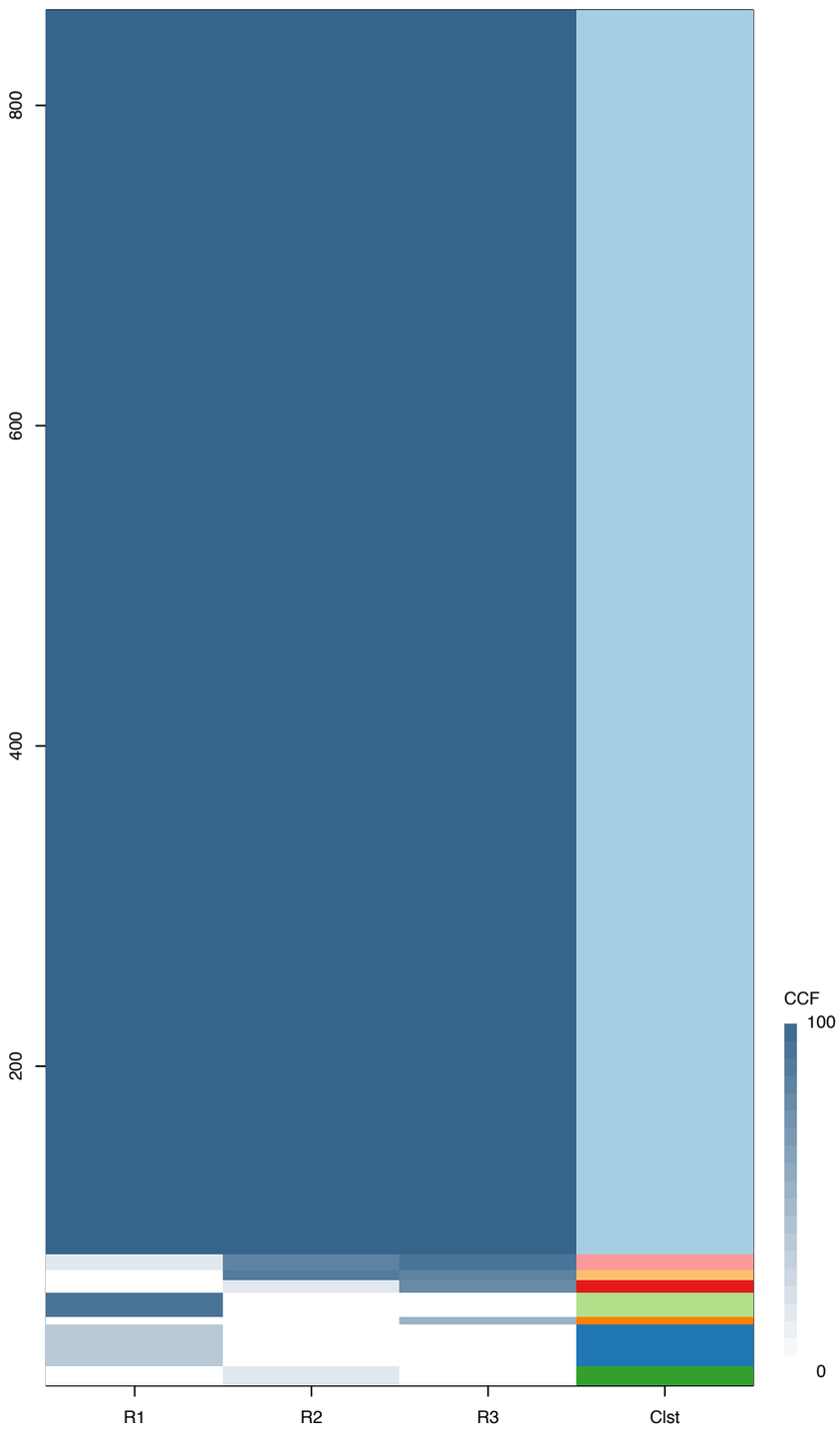
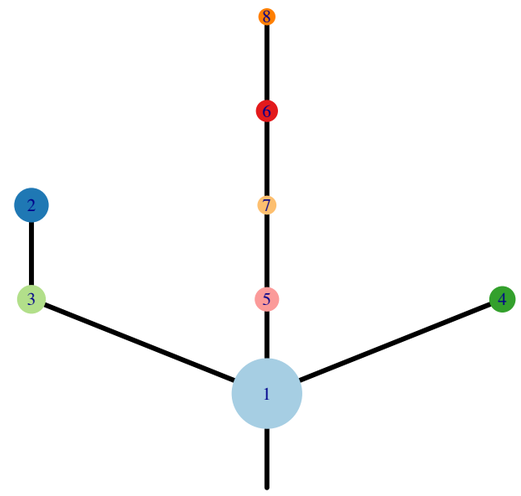


Fig.S12CL



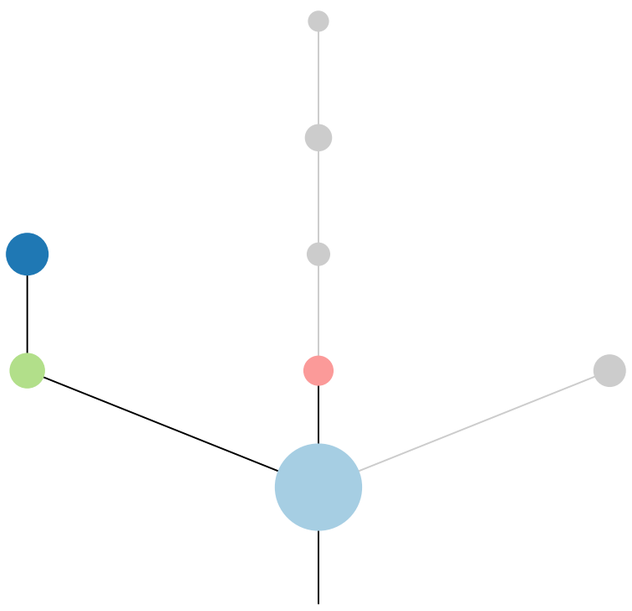
CRUK0100_C



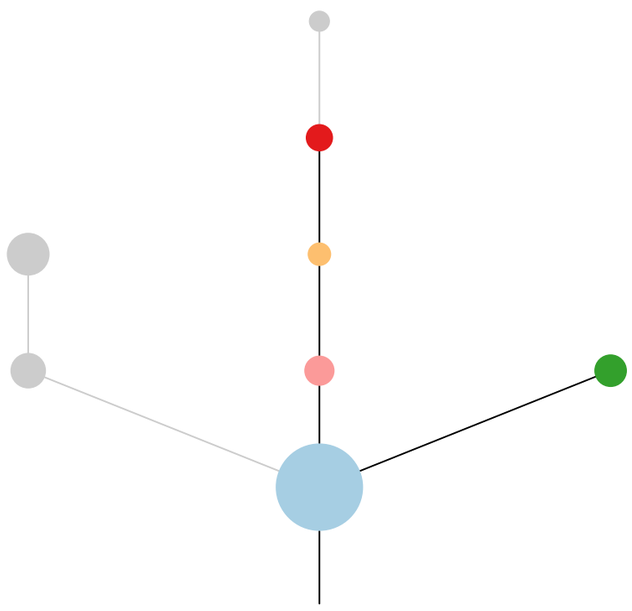
Gene	Cluster	Cytoband	Type
SPEN	1	1p36.13	SNV
COL5A2	1	2q32.2	SNV
PHOX2B	1	4p13	SNV
CDKN2A	1	9p21.3	Del
TP53	1	17p13.1	SNV
STK11	1	19p13.3	SNV
GOLGA5	6	14q32.12	Amp
ZNF521	6	18q11.2	Amp
SS18	6	18q11.2	Amp
AMER1	?		SNV

Histology:Other, Age:50, PackYears:25, Size:29
Stage:3a, Gender:Female, GD:Clonal GD, Recur:yes

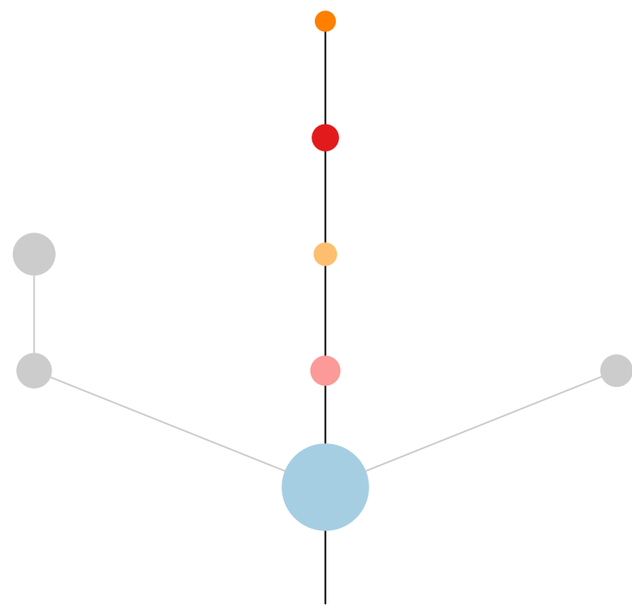
R1

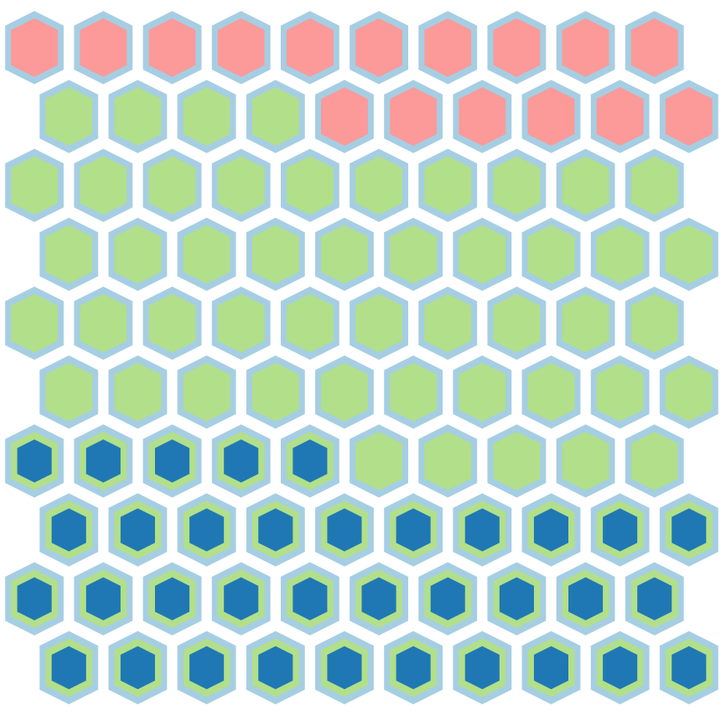


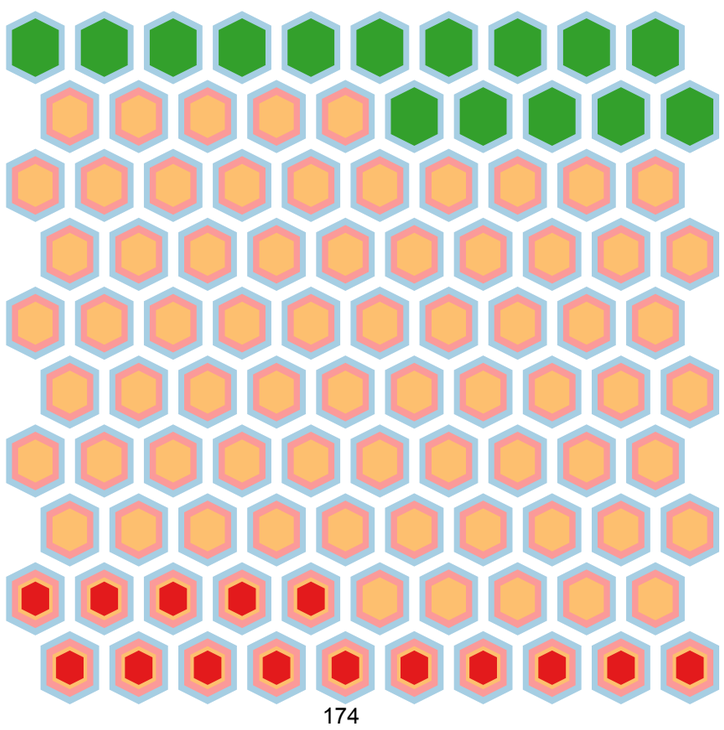
R2



R3







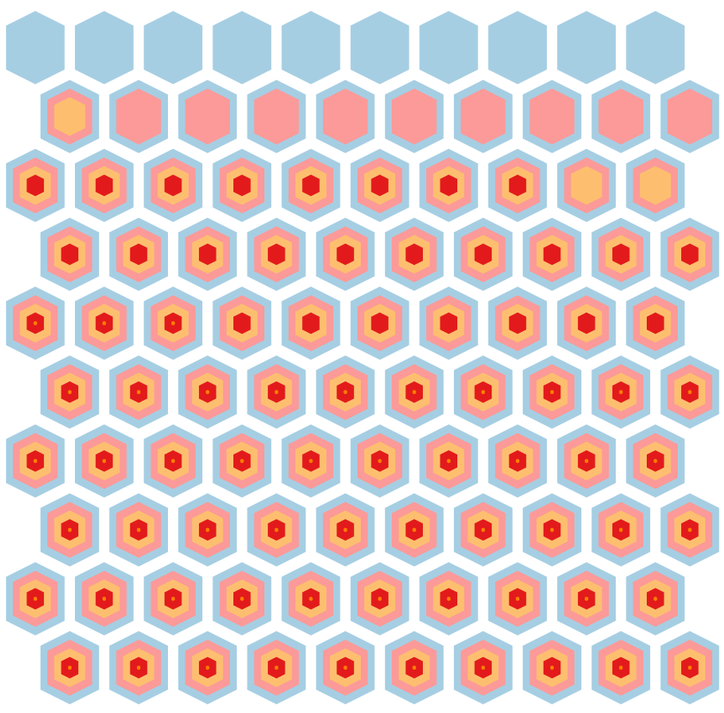


Figure S13

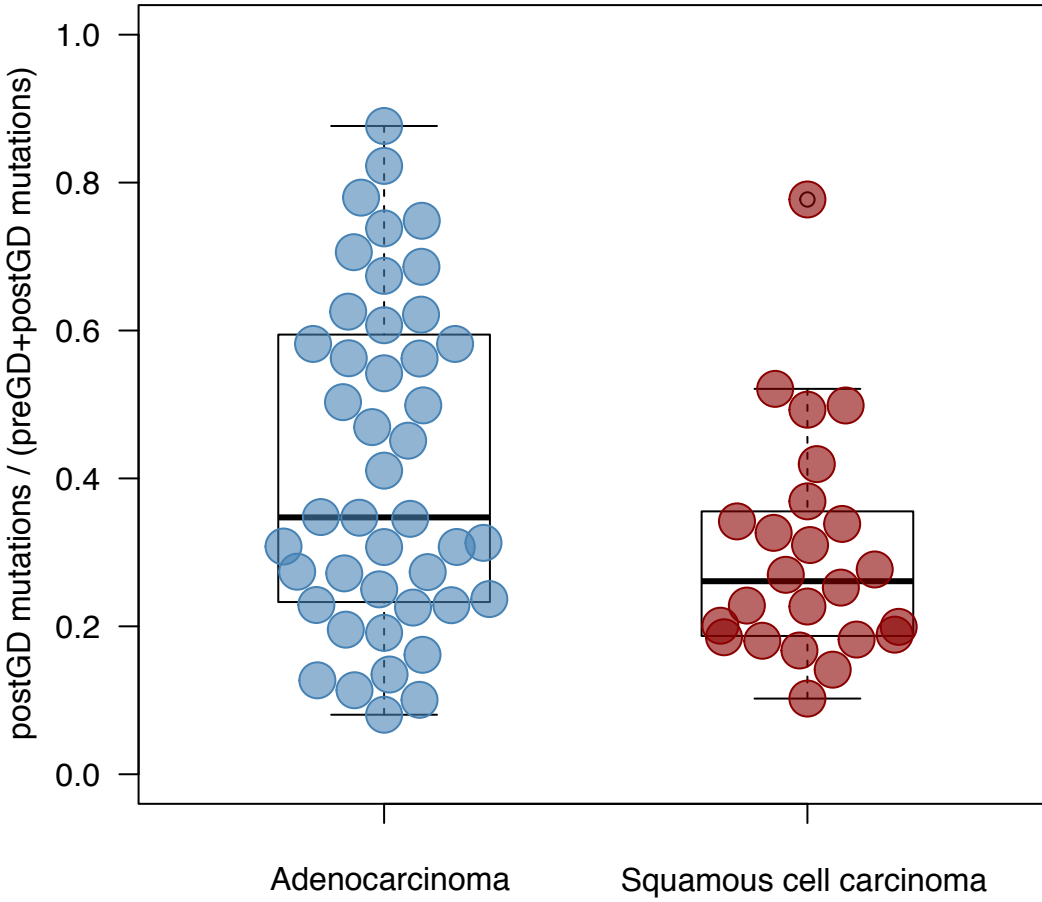


Figure S13. Timing of Genome Doubling
Boxplots indicate estimated timing of genome doubling in each adenocarcinoma and squamous cell carcinoma. On average, adenocarcinoma tumors exhibit significantly earlier clonal genome doubling events compared to squamous cell carcinomas. Timing is estimated based on the fraction of mutations occurring post genome doubling compared to pre- and post genome doubling.

Figure S14

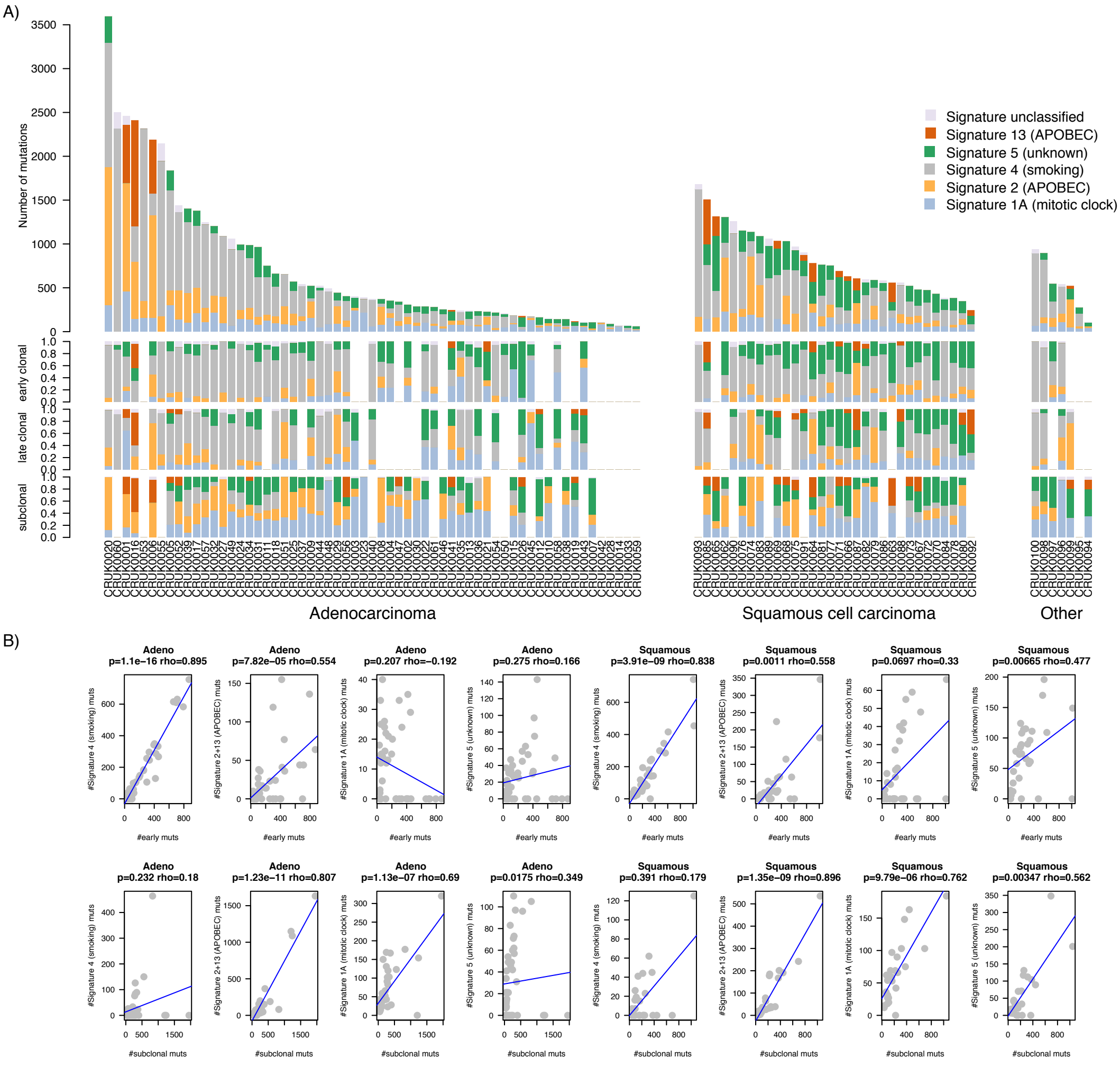


Figure S14. Mutational signatures in TRACERx samples

A) Mutational signatures identified in TRACERx samples, split according to histology and timing of mutations. B) Correlation of prevalence of mutational signatures and different mutation categories in adenocarcinoma and squamous cell cancers. Significance from Spearman's rank correlation is indicated.

Figure S15

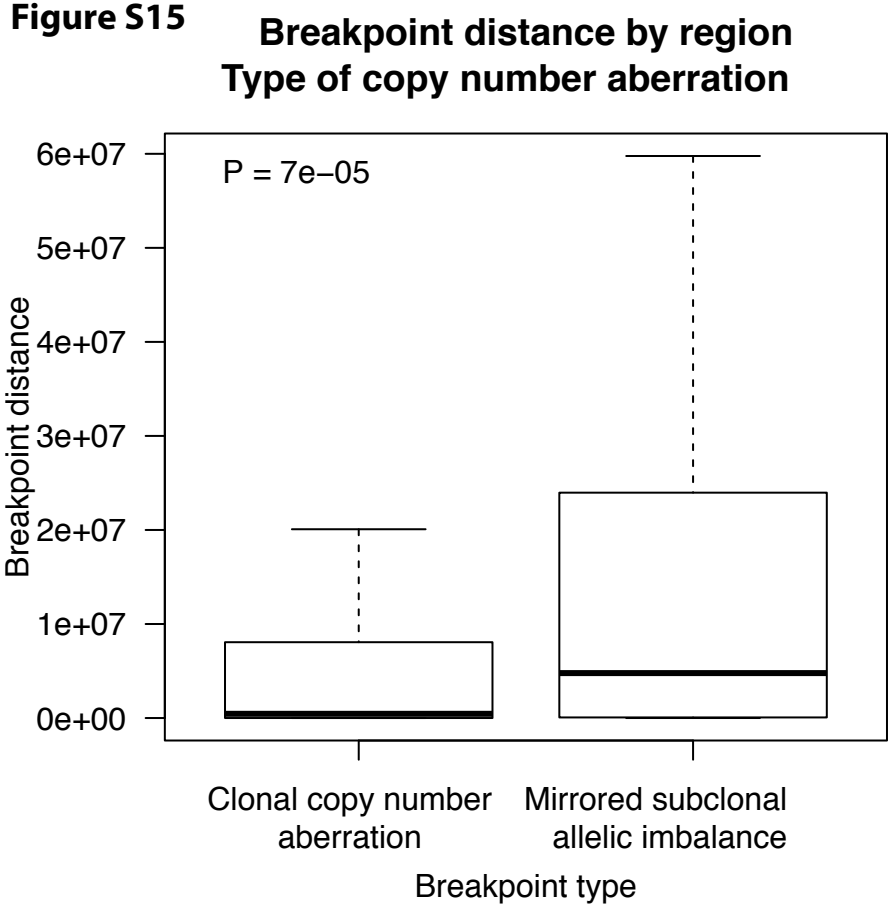


Figure S15. Breakpoint analysis comparing mirrored subclonal allelic imbalance events to clonal copy number aberrations

Figure comparing the minimum distance of breakpoints bordering clonal copy number gains/losses across tumor regions or mirrored subclonal allelic imbalance events. Increased breakpoint distance suggests distinct events in mirrored subclonal allelic imbalance compared to clonal copy number aberrations. Breakpoints bordering the chromosomal start and end positions and the centromere have been removed before the comparison. P-value from t-test.

Figure S16

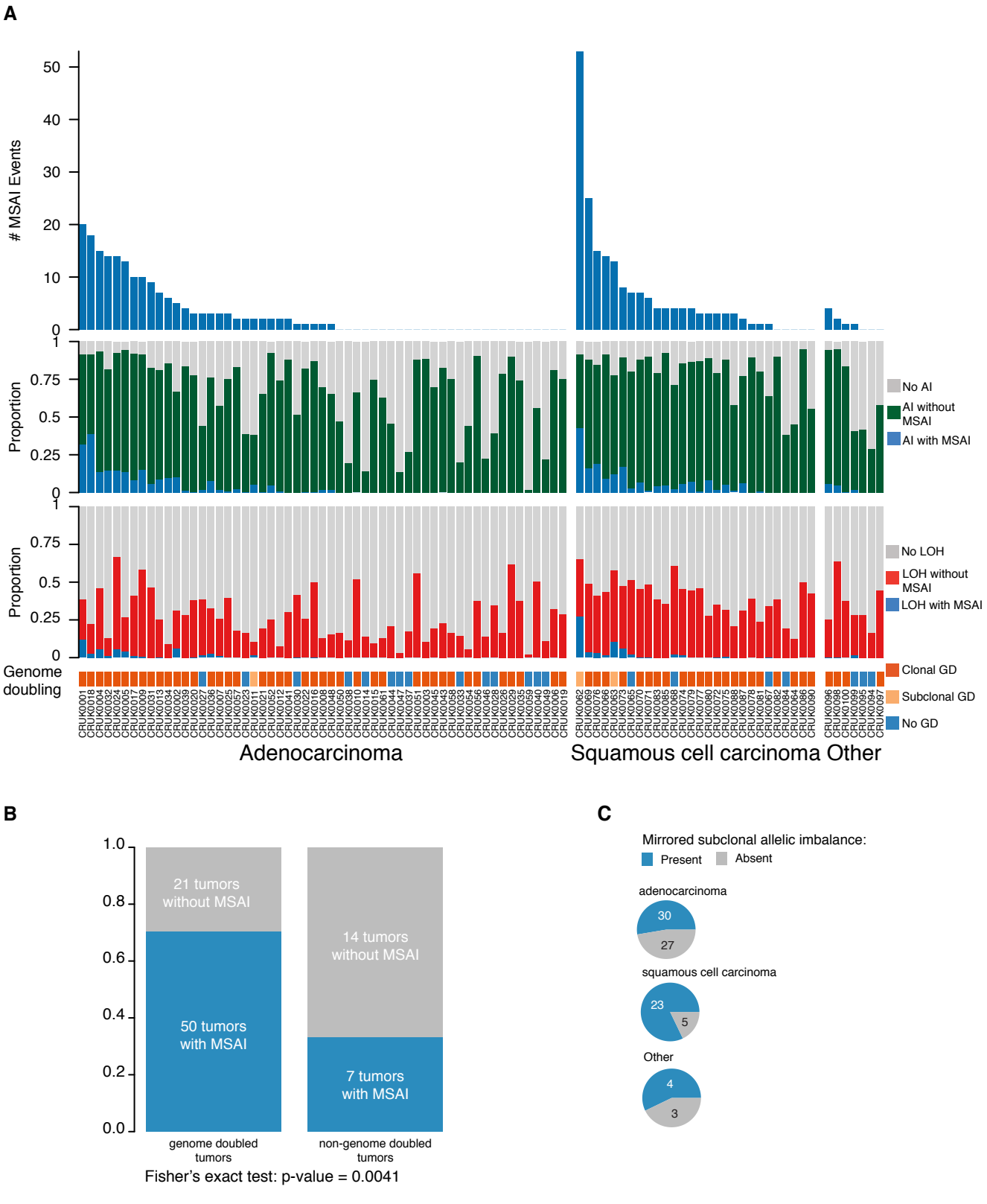


Figure S16. Summary of mirrored subclonal allelic imbalance events across TRACERx
A) Plot illustrating the extent of allelic imbalance across the TRACERx cohort. The top panel shows the number of mirrored subclonal allelic imbalance events detected in each tumor. The panels below highlight the proportion of each tumor displaying allelic imbalance and loss of heterozygosity. The proportion displaying mirrored subclonal allelic imbalance is shown (blue). AI: Allelic imbalance. MSAI: Mirrored subclonal allelic imbalance. LOH: Loss of heterozygosity. GD: Genome doubled. B) Plot showing the relationship between mirrored subclonal allelic imbalance and genome doubling, with a significant enrichment for mirrored subclonal allelic imbalance observed in genome-doubled tumors (P=0.0041, Fisher's Exact Test). C) Pie chart showing proportion of samples affected by mirrored subclonal allelic imbalance within adenocarcinoma, squamous cell cancer, and other NSCLCs.

Figure S17

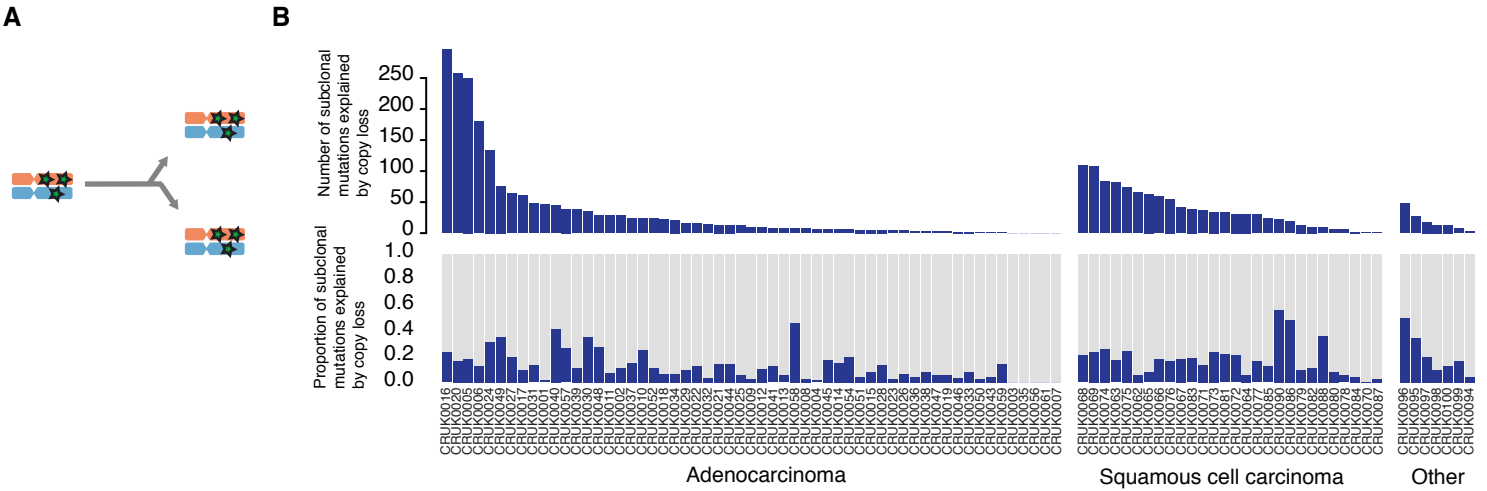


Figure S17. Copy number-driven mutational heterogeneity

A) Schematic illustrating how copy number events can result in loss of previously clonal mutations.

B) Barplot showing extent to which subclonal mutations may result from copy number losses.

Figure S18

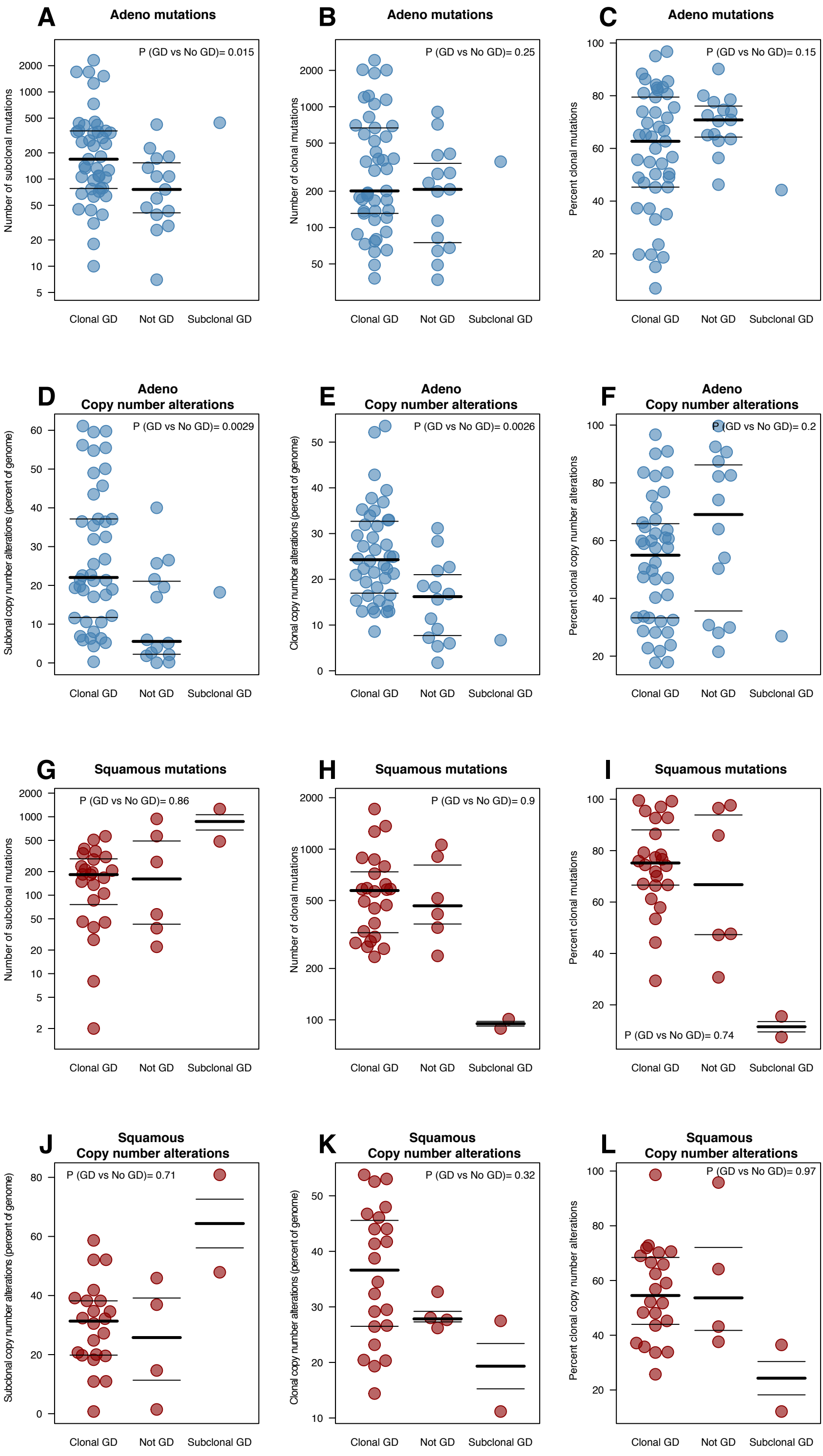


Figure S18. Genome doubling versus intratumor heterogeneity
Clonal, subclonal & fraction clonal mutations (A-C, adenocarcinoma, G-I, squamous cell carcinoma) and copy number alterations (D-E, adenocarcinoma, J-L, squamous cell carcinoma) by genome doubled status.

Figure S19

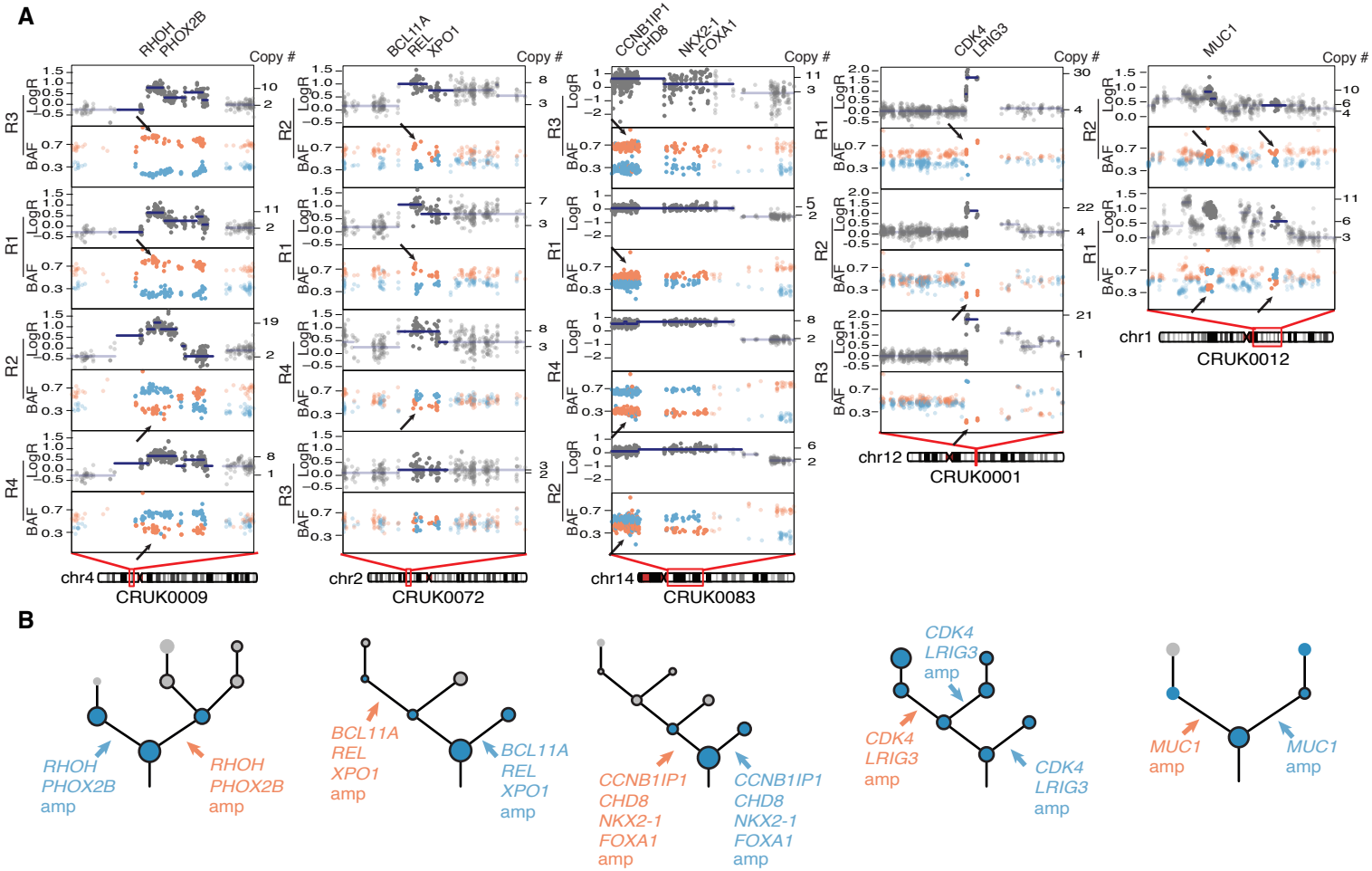


Figure S19. Parallel evolution through high-level copy number amplification

A) Parallel evolution of copy number amplification events observed in multiple tumors, detected through the observation of mirrored subclonal allelic imbalance, indicated by arrows. B) Phylogenetic trees indicating parallel evolution of driver amplifications detected through the observation of mirrored subclonal allelic imbalance (arrows). Subclones colored dark blue harbor a cancer driver event, black circle indicates subclone appears clonal in at least one tumor region.

Figure S20

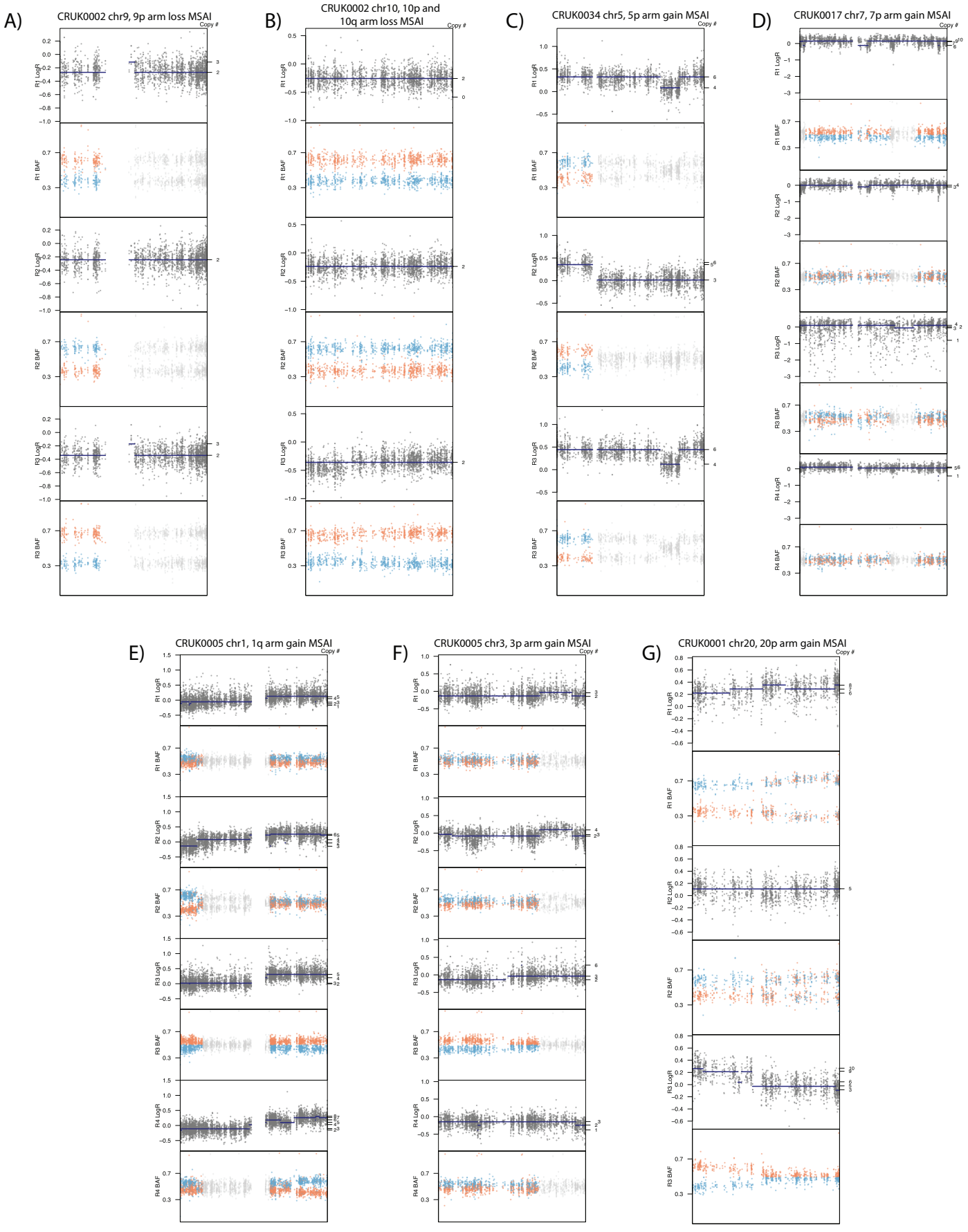


Figure S20. Phased SNPs shown for arm-level mirrored subclonal allelic imbalance events in adenocarcinoma
Plots A-G illustrate LogR and B allele frequency of heterozygous SNPs across all tumor regions from whole chromosomes on which there are mirrored subclonal allelic imbalance arm level gains in adenocarcinoma tumors. Copy number is indicated on the right hand side of each plot. Only heterozygous SNP's in areas involved in mirrored subclonal allelic imbalance events, including the arm event itself, are colored orange or blue (see Experimental Procedures in Supplementary Appendix).

Figure S21

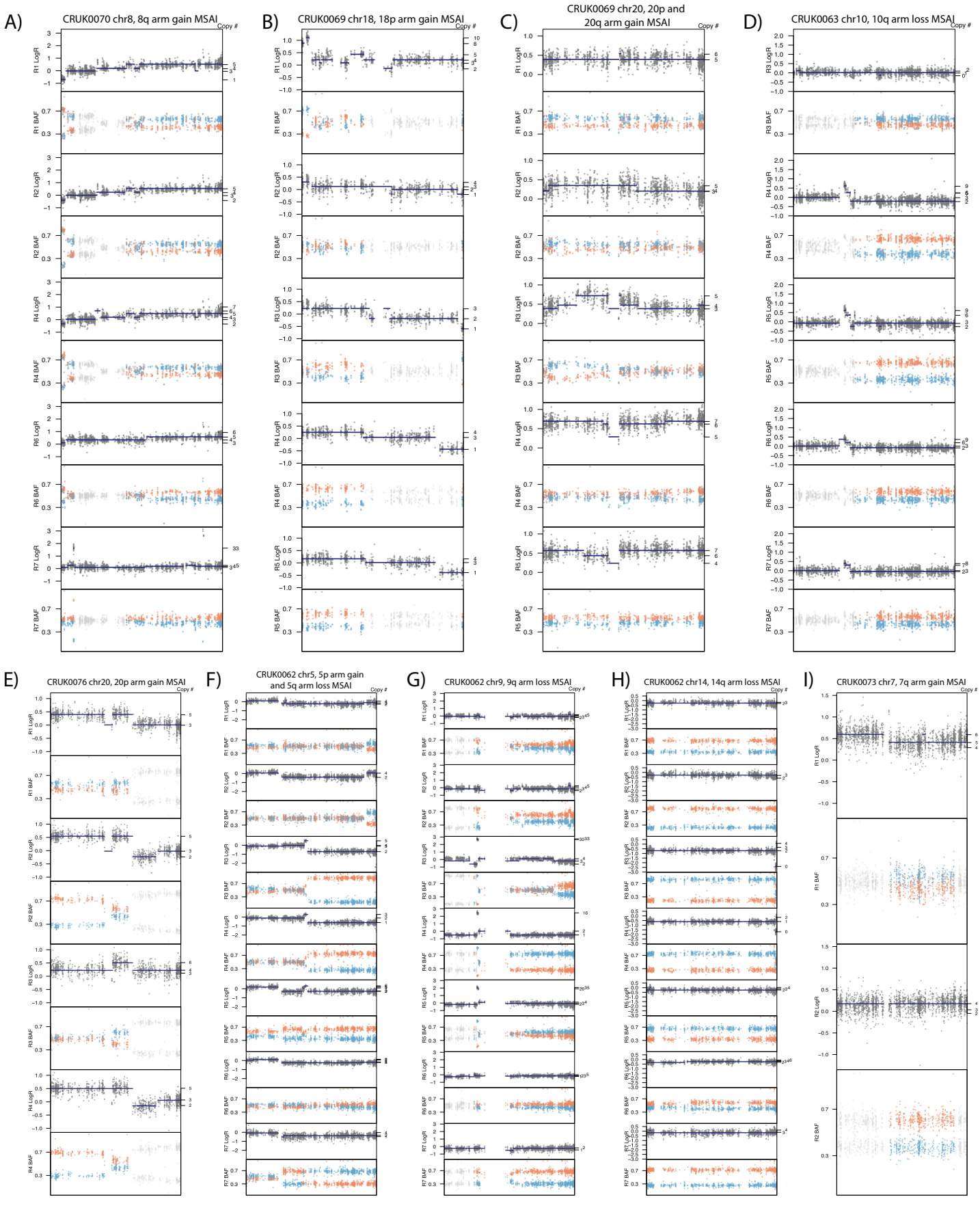


Figure S21. Phased SNPs shown for arm-level mirrored subclonal allelic imbalance events in squamous cell carcinoma

Plots A-I illustrate LogR and B allele frequency of heterozygous SNPs across all tumor regions from whole chromosomes on which there are mirrored subclonal allelic imbalance arm level gains in squamous cell cancers. Copy number is indicated on the right hand side of each plot. Only heterozygous SNP's in areas involved in mirrored subclonal allelic imbalance events, including the arm event itself, are colored orange or blue (see Experimental Procedures in Supplementary Appendix).

Figure S22

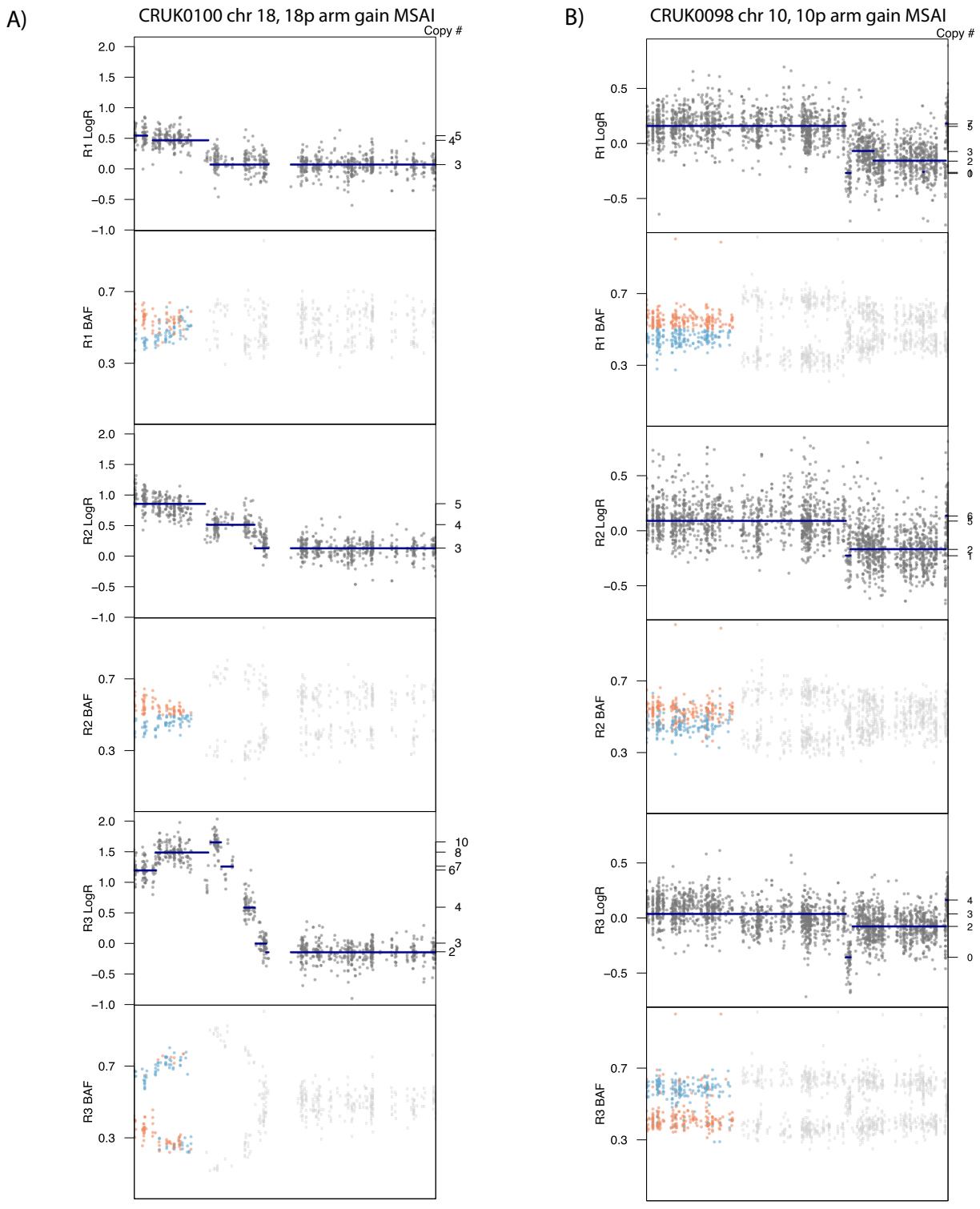


Figure S22. Phased SNPs shown for arm-level mirrored subclonal allelic imbalance events in other NSCLCs
Plots A-B illustrate LogR and B allele frequency of heterozygous SNPs across all tumor regions from whole chromosomes on which there are mirrored subclonal allelic imbalance arm level gains in other NSCLC tumors. Copy number is indicated on the right hand side of each plot. Only heterozygous SNP's in areas involved in mirrored subclonal allelic imbalance events, including the arm event itself, are colored orange or blue (see Experimental Procedures in Supplementary Appendix).

Figure S23

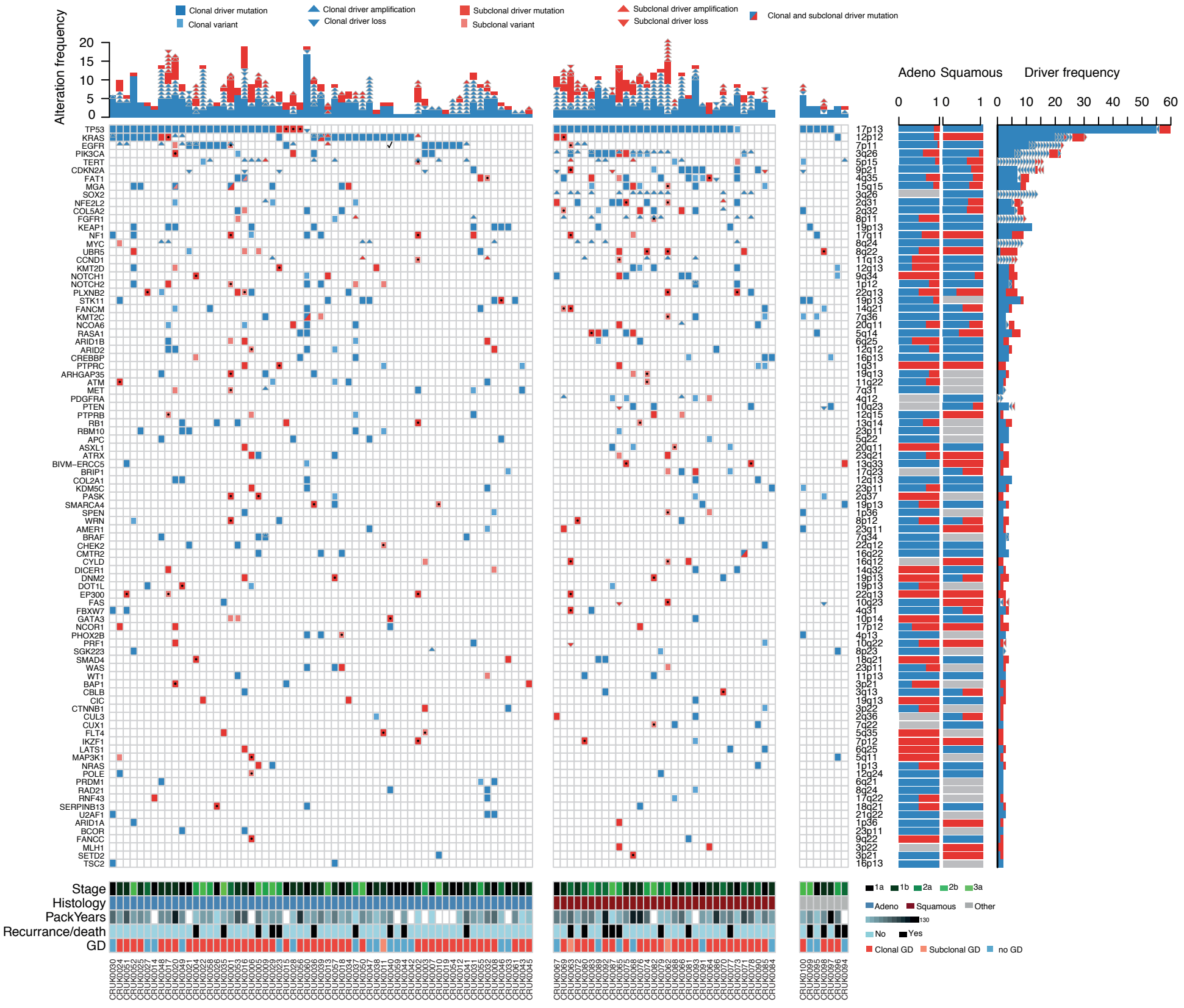


Figure S23. Heterogeneity of driver mutations in NSCLC

Driver alterations detected for each tumor. Only genes containing >=2 driver alterations across the cohort are included. The color represents the clonal status (blue = clonal, red = subclonal) and the shade/size indicates driver. The top panel displays the number of driver alterations identified across individual TRACERx tumors, the barplots to the right show the proportion of the variants found as clonal/subclonal in adenocarcinoma and squamous cell carcinoma for each gene, along with total count of driver alterations. GD: genome doubled. Clonal driver alterations: NSCLC, median 5, range 1-7; adenocarcinoma, median 4; range 1-17; squamous cell carcinoma, median 6, range 1-13. Subclonal driver alterations, NSCLC, median 1, range 0-12; adenocarcinoma, median 1 range 0-10; squamous cell carcinoma, median 2, range 0-12.

Figure S24

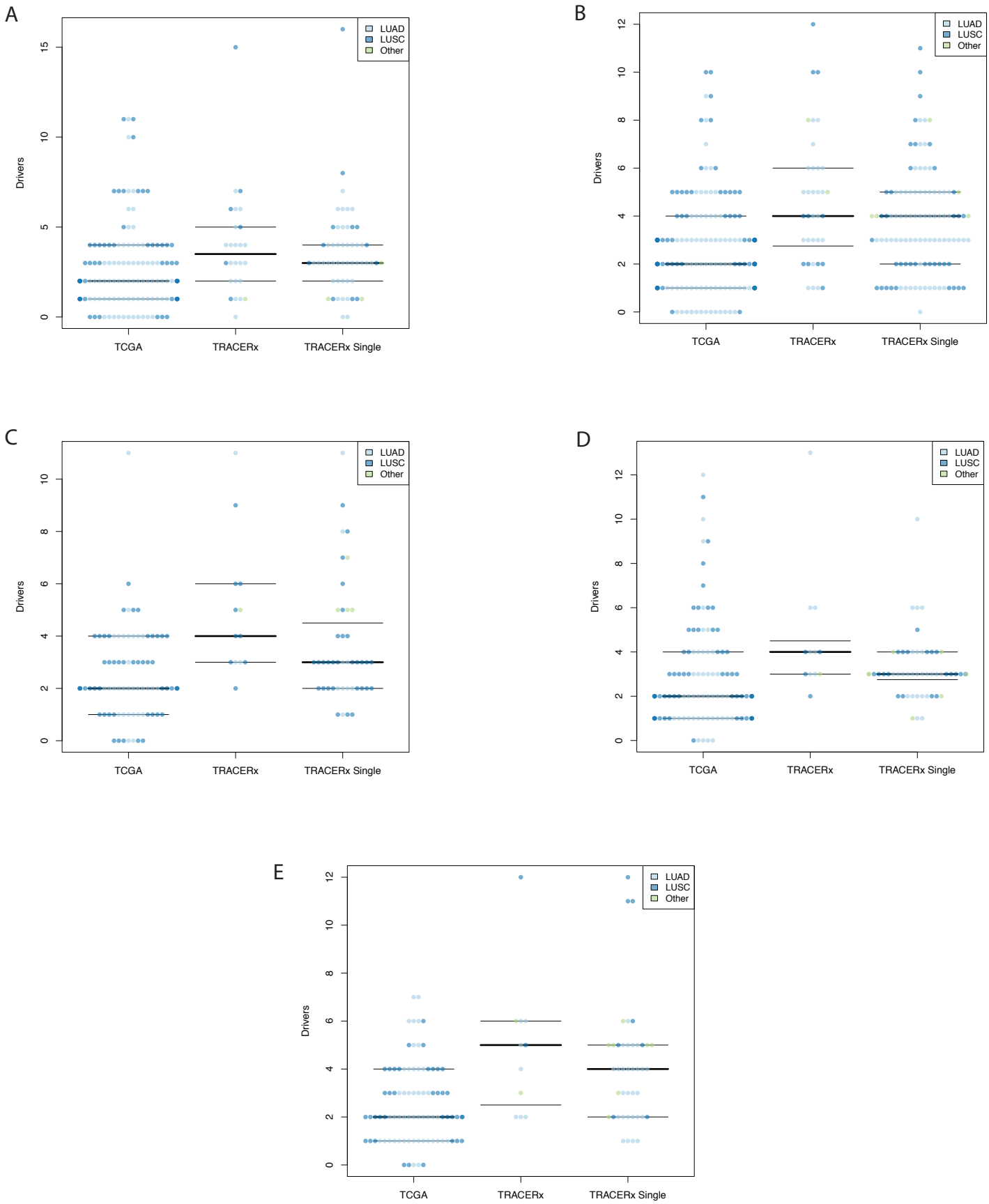


Figure S24.
Driver mutations called in TCGA adenocarcinoma and squamous cell carcinoma, TRACERx multi-region and TRACERx analyzed as single samples in A) Stage 1A tumors, B) Stage 1B tumors, C) Stage 2A tumors, D) Stage 2B tumors and E) Stage 3A tumors. LUAD: Lung adenocarcinoma. LUSC: Lung squamous cell carcinoma.

Figure S25

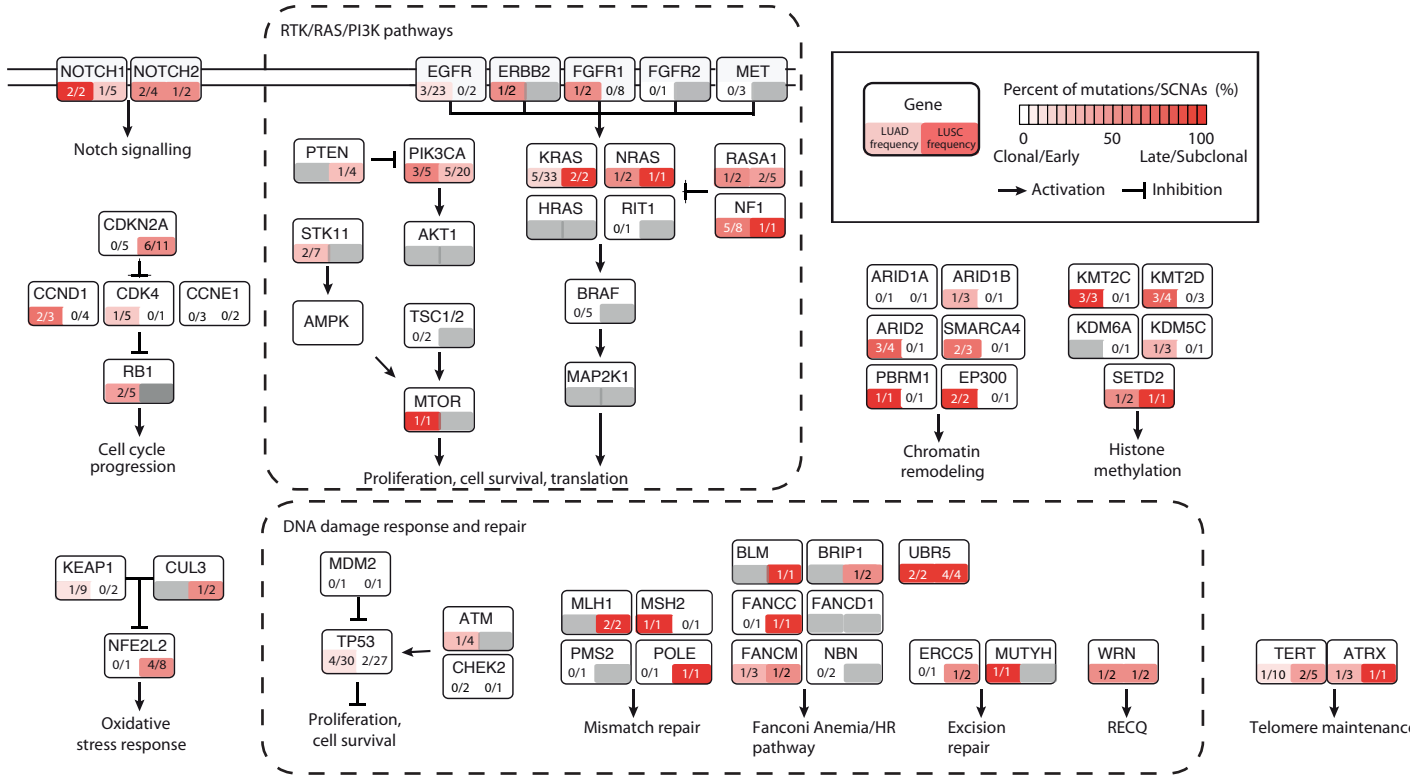


Figure S25. Timing of driver alterations in TRACERx by histological subtype

Driver alterations (mutations and copy number aberrations) found in key pathway components. For each gene, the number of alterations that are late/subclonal is indicated compared to total alteration frequency, with adenocarcinoma (LUAD, left) and squamous cell cancer (LUSC, right). The color of the box indicates proportion of alterations that are late/subclonal, with a deeper red indicating higher proportion late/subclonal. Grey indicates absence of mutations.

Figure S26

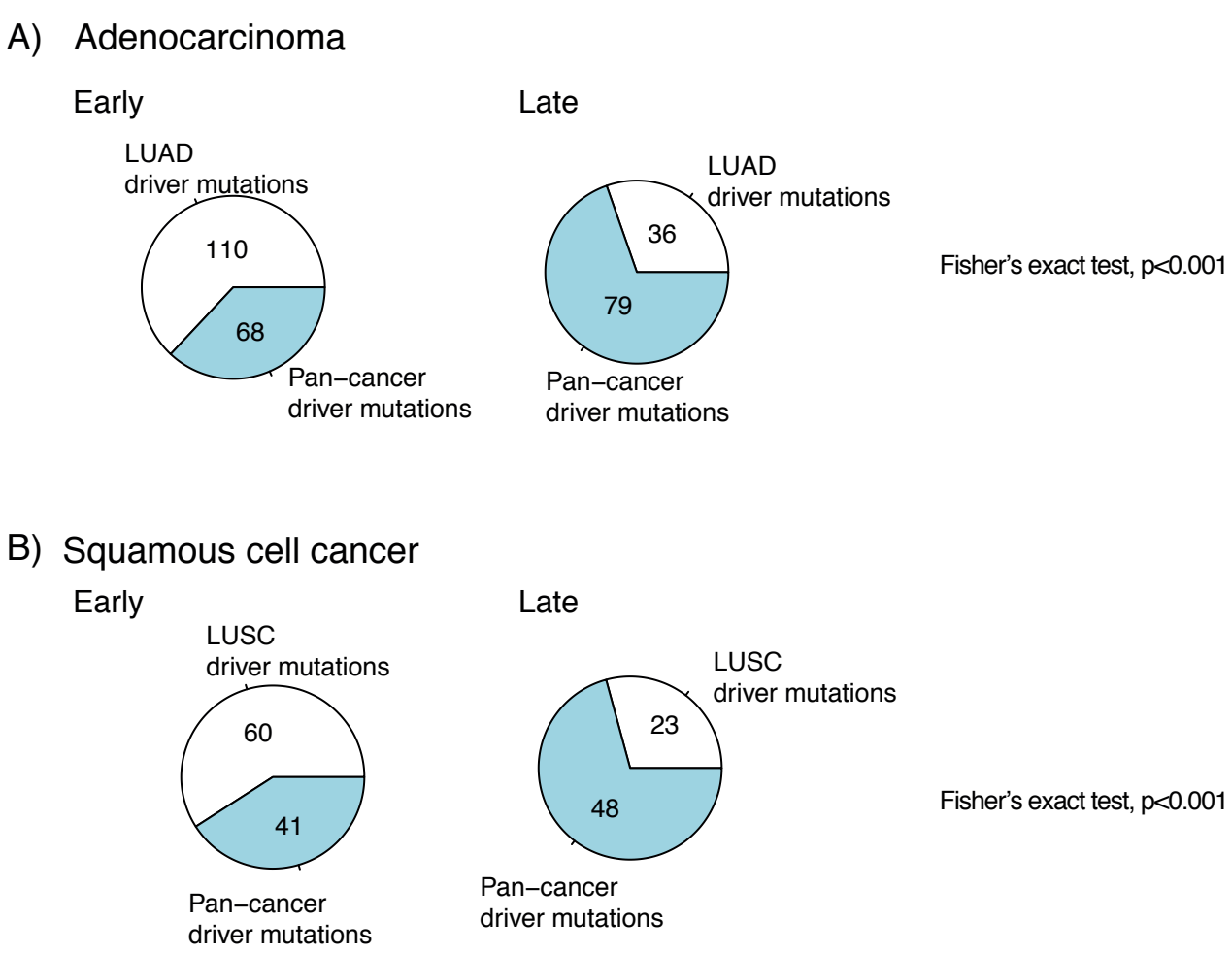


Figure S26. Constraints in cancer evolution
Pie charts show differences in the proportion of early and late mutations residing in cancer genes identified through subtype specific analysis versus other cancer genes. LUAD: Lung adenocarcinoma.
LUSC: Lung squamous cell carcinoma.

Figure S27

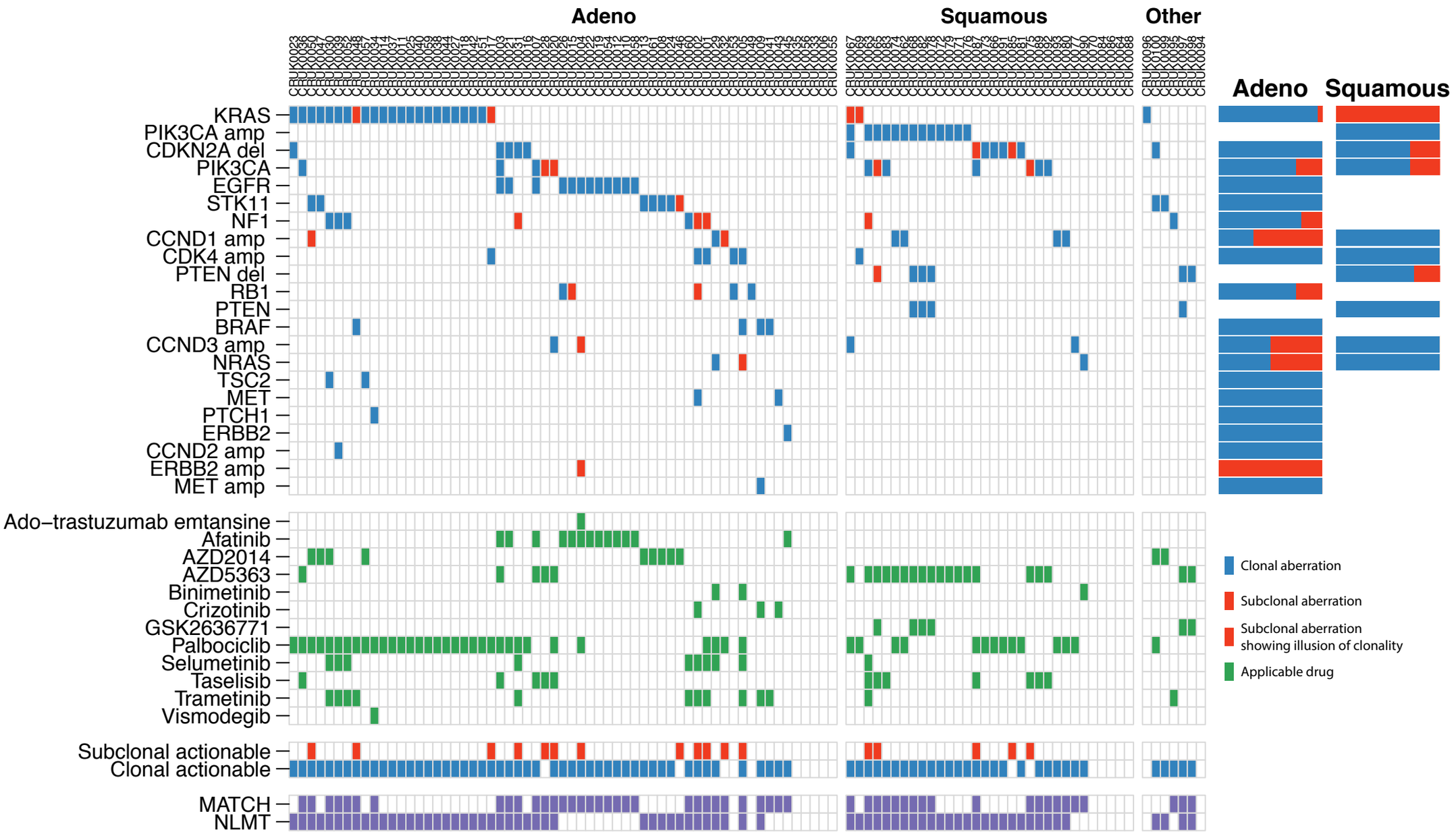
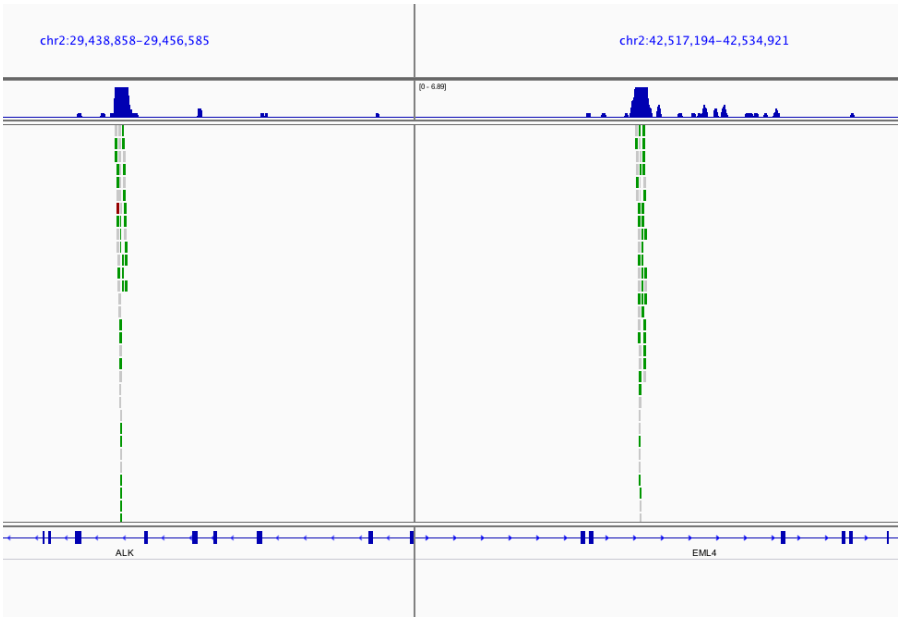


Figure S27. Figure showing patients with genomic alterations eligible for either the MATCH or NLMT genomically guided clinical study. Blue indicates clonal alteration, red indicates subclonal. Green indicates eligibility for specific targeted drugs.

Figure S28

A.



B.

C.

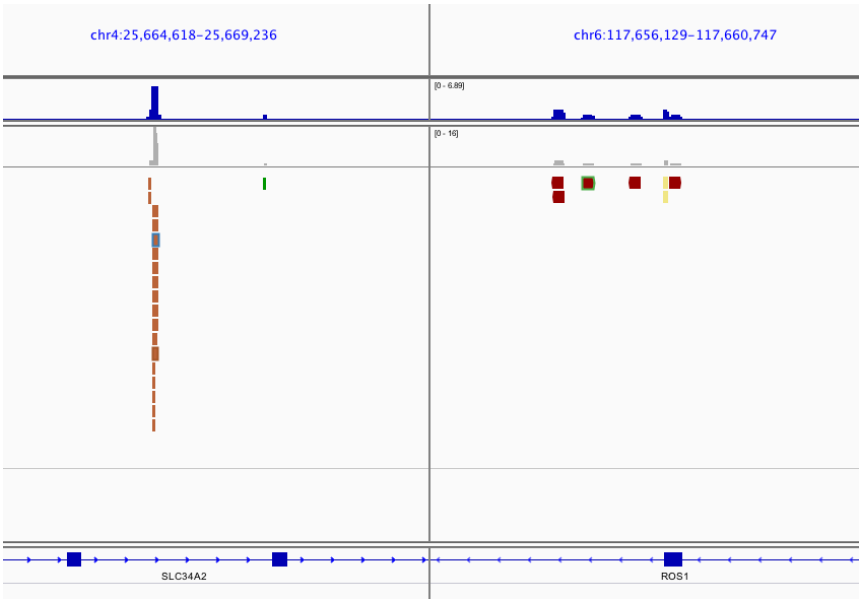


Figure S28. Gene fusion events detected from NGS data in postive control DNA in A) *ALK:EML4*, B) *RET:CCDC6* and C) *SLC34A2:ROS1*

Supplementary tables
Table S1

Table S1. Summary of clinical and pathological characteristics of the TRACERx 100 cohort.

Number of patients (n)		100
Median age at diagnosis and range (years)		68 (34-85)
Gender		
	Male	62
	Female	38
ECOG PS		
	0	51
	1	49
Status		
	Alive	84
	Dead	16
	Relapse (10 of which dead)	20
Ethnicity		
	White-British	89
	White-Irish	4
	White-Other	4
	Carribean	3
Smoking status		
	Current smoker	7
	Recent ex-smoker	33
	Ex-smoker	48
	Never smoked	12
Histological subtype		
	Adenocarcinoma	61
	Squamous cell carcinoma	32
	Adenosquamous	3
	Pleomorphic carcinoma	1
	Large cell carcinoma	1
	Sarcomatoid carcinoma of pleomorphic type	1
	Large Cell Neuroendocrine Carcinoma	1
Tumor stage (based on 7th TNM classification)		
	IA	26
	IB	36
	IIA	13
	IIB	11
	IIIA	13
	IIIB	1
Median tumour size and range (mm)		33.5 (10-120)
Resection margin		
	R0	95
	R1	5
Vascular invason		
	Yes	31
	No	69
Pleural invasion		
	Yes	27
	No	73
Adjuvant treatment		
	Adjuvant treatment	28
	No adjuvant treatment	72

Definitions:
Never smoked (smoked >100 cigarettes in lifetime)
Ex-smoker (smoked 100 or more cigarettes in lifetime and quit>1 year ago)
Recent ex-smoked (smoked 100 or more cigarettes in lifetime and quit <1 year ago)

Table S4

Table S4. Validation of mutation calling.

Validation Method	Mutation	Unique Mutations Tested	Total Tested	Clonal Validated	Clonal FP	Clonal FN	Clonal Somatic FP	Subclonal Validated	Subclonal FP	Subclonal FN	Subclonal Somatic FP
Multiplexed PCR	SNV	792	2856	1360	2 0.15%	0 0.00%	0 0.00%	1491	3 0.20%	0 0.00%	0 0.00%
AmpliSeq	SNV	985	4823	834	1 0.12%	0 0.00%	0 0.00%	3949	13 0.33%	23 0.58%	3 0.08%
AmpliSeq	INDEL	89	392	248	0 0.00%	1 0.4	0 0.00%	133	0 0.00%	10 6.99%	0 0.00%

Table S5

#Table S5. Clinical correlates with ITH based on Spearman-rank correlation.
#r: correlation estimate, p=P-value, p_fdr = P-values corrected for False Discovery Rate (q-values)

	Age_r	PackYears_r	Size_r	KI67_r	Stage_r	Age_p	PackYears_p	Size_p	KI67_p	Stage_p	Age_p_fdr	PackYears_p_fdr	Size_p_fdr	KI67_fdr	Stage_p_fdr
All_Clonal_SNV	-0.0623	0.3330	0.0881	0.4671	0.0414	0.5379	0.0017	0.3832	0.0000	0.6828	1.0000	0.1332	1.0000	0.0001	1.0000
All_Subclonal_SNV	-0.0350	0.2301	0.2558	0.3473	0.1856	0.7299	0.0331	0.0102	0.0005	0.0644	1.0000	1.0000	0.7251	0.0368	1.0000
All_FractionSubclonal_SNV	-0.0154	-0.0918	0.0808	-0.0651	0.0611	0.8794	0.4004	0.4242	0.5242	0.5457	1.0000	1.0000	1.0000	1.0000	1.0000
All_Clonal_SCNA	0.0299	0.0966	0.0173	0.2828	0.0126	0.7773	0.3940	0.8698	0.0069	0.9049	1.0000	1.0000	1.0000	0.5047	1.0000
All_Subclonal_SCNA	-0.1666	0.1749	0.3728	0.5065	0.3764	0.1125	0.1207	0.0003	0.0000	0.0002	1.0000	1.0000	0.0210	0.0000	0.0182
All_FractionSubclonal_SCNA	-0.1682	0.0901	0.3507	0.2774	0.3634	0.1091	0.4267	0.0006	0.0081	0.0004	1.0000	1.0000	0.0481	0.5846	0.0300
Adeno_Clonal_SNV	-0.1947	0.5197	0.1076	0.5007	0.1102	0.1327	0.0001	0.4090	0.0001	0.3980	1.0000	0.0093	1.0000	0.0046	1.0000
Adeno_Subclonal_SNV	-0.0883	0.2862	0.2414	0.5862	0.2460	0.4984	0.0439	0.0609	0.0000	0.0560	1.0000	1.0000	1.0000	0.0001	1.0000
Adeno_FractionSubclonal_SNV	0.0518	-0.1977	0.1211	0.1230	0.1075	0.6920	0.1688	0.3526	0.3535	0.4097	1.0000	1.0000	1.0000	1.0000	1.0000
Adeno_Clonal_SCNA	-0.0765	0.0478	-0.0306	0.1780	0.1011	0.5715	0.7495	0.8213	0.1935	0.4543	1.0000	1.0000	1.0000	1.0000	1.0000
Adeno_Subclonal_SCNA	-0.1825	0.1632	0.3605	0.5600	0.4618	0.1742	0.2732	0.0059	0.0000	0.0003	1.0000	1.0000	0.4343	0.0008	0.0245
Adeno_FractionSubclonal_SCNA	-0.1363	0.0955	0.3859	0.4269	0.4039	0.3122	0.5232	0.0030	0.0012	0.0018	1.0000	1.0000	0.2272	0.0898	0.1396
Squamous_Clonal_SNV	0.2735	-0.0699	-0.1992	-0.0050	-0.1333	0.1298	0.7137	0.2743	0.9785	0.4671	1.0000	1.0000	1.0000	1.0000	1.0000
Squamous_Subclonal_SNV	0.0229	0.2227	0.2655	0.1301	0.1523	0.9008	0.2368	0.1419	0.4778	0.4053	1.0000	1.0000	1.0000	1.0000	1.0000
Squamous_FractionSubclonal_SNV	-0.1318	0.1213	0.2406	0.1785	0.0962	0.4721	0.5232	0.1847	0.3283	0.6004	1.0000	1.0000	1.0000	1.0000	1.0000
Squamous_Clonal_SCNA	0.1793	-0.0366	0.0584	0.1817	0.0031	0.3612	0.8560	0.7677	0.3547	0.9873	1.0000	1.0000	1.0000	1.0000	1.0000
Squamous_Subclonal_SCNA	-0.0916	0.2375	0.2417	0.2748	0.0931	0.6430	0.2329	0.2154	0.1570	0.6376	1.0000	1.0000	1.0000	1.0000	1.0000
Squamous_FractionSubclonal_SCNA	-0.1522	0.1832	0.1347	0.0677	0.0737	0.4395	0.3604	0.4944	0.7320	0.7095	1.0000	1.0000	1.0000	1.0000	1.0000

Table S6

Multivariable analysis for copy number heterogeneity – recurrence-free survival (26 events among 100 patients)

Factor	Adjusted hazard ratio (95% CI)	P-value
Age	0.99 (0.95-1.03)	0.60
Pack years	0.99 (0.97-1.01)	0.24
Adenocarcinoma (vs not adenocarcinoma)	0.60 (0.24-1.51)	0.28
Adjuvant treatment (vs no adjuvant)	0.21 (0.08-0.59)	0.003
Stage (vs IA)		<0.0001 (trend)
IB	0.83 (0.16-4.24)	
IIA	3.11 (0.59-16.49)	
IIB	7.68 (1.52-38.91)	
IIIA	14.05 (3.00-65.83)	
Copy number heterogeneity (≥48 vs <48%)	3.70 (1.29-10.65)	0.01
Copy number heterogeneity (as continuous), for every increase of 10 percentage points	1.25 (1.00-1.56)	0.048

*each hazard ratio is adjusted for all the other factors in the table

Table S8

Table S8. dN/dS ratio across cohort.

Cancer type	timing	gene.subset	mutation.typ	MLE	CI_lower	CI_upper
NSCLC	All	Exome-wide	wMIS	1.04070458	1.01545605	1.0665809
NSCLC	All	Exome-wide	wNON	0.91800645	0.87233096	0.96607352
NSCLC	Early	Exome-wide	wMIS	1.00270804	0.95705375	1.05054019
NSCLC	Early	Exome-wide	wNON	0.82983701	0.75306628	0.91443406
NSCLC	Clonal	Exome-wide	wMIS	1.03622825	1.00379016	1.0697146
NSCLC	Clonal	Exome-wide	wNON	0.89292949	0.83649788	0.95316808
NSCLC	Late	Exome-wide	wMIS	1.05274521	1.01636675	1.09042576
NSCLC	Late	Exome-wide	wNON	0.94184063	0.87433876	1.01455386
NSCLC	Subclonal	Exome-wide	wMIS	1.04678657	1.00708913	1.08804881
NSCLC	Subclonal	Exome-wide	wNON	0.95971983	0.88429068	1.04158302
LUAD	All	Exome-wide	wMIS	1.04070458	1.01545605	1.0665809
LUAD	All	Exome-wide	wNON	0.91800645	0.87233096	0.96607352
LUAD	Early	Exome-wide	wMIS	1.00270804	0.95705375	1.05054019
LUAD	Early	Exome-wide	wNON	0.82983701	0.75306628	0.91443406
LUAD	Clonal	Exome-wide	wMIS	1.03622825	1.00379016	1.0697146
LUAD	Clonal	Exome-wide	wNON	0.89292949	0.83649788	0.95316808
LUAD	Late	Exome-wide	wMIS	1.05274521	1.01636675	1.09042576
LUAD	Late	Exome-wide	wNON	0.94184063	0.87433876	1.01455386
LUAD	Subclonal	Exome-wide	wMIS	1.04678657	1.00708913	1.08804881
LUAD	Subclonal	Exome-wide	wNON	0.95971983	0.88429068	1.04158302
LUSC	All	Exome-wide	wMIS	1.0404732	0.99802326	1.08472872
LUSC	All	Exome-wide	wNON	0.87772641	0.80256994	0.95992088
LUSC	Early	Exome-wide	wMIS	0.98286858	0.91584081	1.05480192
LUSC	Early	Exome-wide	wNON	0.83733075	0.72278845	0.97002489
LUSC	Clonal	Exome-wide	wMIS	1.02103985	0.97006655	1.07469159
LUSC	Clonal	Exome-wide	wNON	0.86650107	0.77795957	0.96511969
LUSC	Late	Exome-wide	wMIS	1.07869072	1.0127148	1.1489648
LUSC	Late	Exome-wide	wNON	0.89518014	0.77805491	1.02993692
LUSC	Subclonal	Exome-wide	wMIS	1.07916381	1.00460911	1.1592514
LUSC	Subclonal	Exome-wide	wNON	0.89829192	0.7647789	1.05511328
NSCLC	All	NSCLC cancer-genes	wMIS	2.37941844	2.09382071	2.68995405
NSCLC	All	NSCLC cancer-genes	wNON	5.28221092	3.95049275	6.88498035
NSCLC	Early	NSCLC cancer-genes	wMIS	5.60801899	4.79463381	6.50859613
NSCLC	Early	NSCLC cancer-genes	wNON	12.7388013	8.99674207	17.402257
NSCLC	Clonal	NSCLC cancer-genes	wMIS	3.20523033	2.82820421	3.61445452
NSCLC	Clonal	NSCLC cancer-genes	wNON	6.8829668	5.16529874	8.94652295
NSCLC	Late	NSCLC cancer-genes	wMIS	1.32349342	1.0907278	1.58736564
NSCLC	Late	NSCLC cancer-genes	wNON	2.85116607	1.80574751	4.23930362
NSCLC	Subclonal	NSCLC cancer-genes	wMIS	1.39920333	1.13290518	1.70446245
NSCLC	Subclonal	NSCLC cancer-genes	wNON	3.34652803	2.08306683	5.04085651
LUAD	All	LUAD cancer-genes	wMIS	2.99492963	2.4925525	3.56073482
LUAD	All	LUAD cancer-genes	wNON	5.90900666	3.85012519	8.59880563
LUAD	Early	LUAD cancer-genes	wMIS	4.83359457	3.86381694	5.95433114
LUAD	Early	LUAD cancer-genes	wNON	10.6349628	6.49217924	16.2512156
LUAD	Clonal	LUAD cancer-genes	wMIS	3.12938101	2.62892669	3.69188628
LUAD	Clonal	LUAD cancer-genes	wNON	5.91391685	3.9112894	8.51871389
LUAD	Late	LUAD cancer-genes	wMIS	2.08751355	1.56003885	2.72327663
LUAD	Late	LUAD cancer-genes	wNON	4.17556721	2.08187001	7.33685541
LUAD	Subclonal	LUAD cancer-genes	wMIS	2.4518015	1.75575221	3.31207581
LUAD	Subclonal	LUAD cancer-genes	wNON	5.47225841	2.571341	10.0127484
LUSC	All	LUSC cancer-genes	wMIS	3.181016	2.53396098	3.93001281
LUSC	All	LUSC cancer-genes	wNON	7.56691494	4.44001779	11.9043584
LUSC	Early	LUSC cancer-genes	wMIS	7.25687451	5.56805583	9.2578061
LUSC	Early	LUSC cancer-genes	wNON	17.2296757	9.40945588	28.4482894
LUSC	Clonal	LUSC cancer-genes	wMIS	6.6123967	5.27519385	8.16016897
LUSC	Clonal	LUSC cancer-genes	wNON	16.1539366	9.6374278	25.0597312
LUSC	Late	LUSC cancer-genes	wMIS	1.54135792	1.08841257	2.10529349
LUSC	Late	LUSC cancer-genes	wNON	3.83427707	1.66071656	7.38559001
LUSC	Subclonal	LUSC cancer-genes	wMIS	1.44549851	1.00500544	1.99967223
LUSC	Subclonal	LUSC cancer-genes	wNON	3.15735984	1.22596399	6.50992994
NSCLC	All	exome-wide (without cancer genes)	wMIS	1.03732189	1.01615769	1.05877915
NSCLC	All	exome-wide (without cancer genes)	wNON	0.86655664	0.80079622	0.93582941
NSCLC	Early	exome-wide (without cancer genes)	wMIS	0.97491229	0.94653874	1.00384757
NSCLC	Early	exome-wide (without cancer genes)	wNON	0.81656035	0.73257197	0.90675102
NSCLC	Clonal	exome-wide (without cancer genes)	wMIS	1.01316536	0.99195153	1.03467938
NSCLC	Clonal	exome-wide (without cancer genes)	wNON	0.85058293	0.78583244	0.91880113
NSCLC	Late	exome-wide (without cancer genes)	wMIS	1.07521841	1.04735065	1.10356748
NSCLC	Late	exome-wide (without cancer genes)	wNON	0.88689475	0.798327	0.98181335
NSCLC	Subclonal	exome-wide (without cancer genes)	wMIS	1.08585703	1.05366387	1.11869964
NSCLC	Subclonal	exome-wide (without cancer genes)	wNON	0.89724576	0.79440467	1.00862653
LUAD	All	exome-wide (without cancer genes)	wMIS	1.03596069	1.02340827	1.04861575
LUAD	All	exome-wide (without cancer genes)	wNON	0.90454233	0.86520645	0.94505662
LUAD	Early	exome-wide (without cancer genes)	wMIS	0.9966982	0.97761766	1.01602547
LUAD	Early	exome-wide (without cancer genes)	wNON	0.80428412	0.74864007	0.86262218
LUAD	Clonal	exome-wide (without cancer genes)	wMIS	1.02997093	1.01659884	1.04345975
LUAD	Clonal	exome-wide (without cancer genes)	wNON	0.87615607	0.83547515	0.91813637
LUAD	Late	exome-wide (without cancer genes)	wMIS	1.04749408	1.0321046	1.06303526
LUAD	Late	exome-wide (without cancer genes)	wNON	0.93154673	0.88199139	0.98292494
LUAD	Subclonal	exome-wide (without cancer genes)	wMIS	1.0444876	1.02743631	1.06172683
LUAD	Subclonal	exome-wide (without cancer genes)	wNON	0.9520923	0.89626714	1.01019321
LUSC	All	exome-wide (without cancer genes)	wMIS	1.03732189	1.01615769	1.05877915
LUSC	All	exome-wide (without cancer genes)	wNON	0.86655664	0.80079622	0.93582941
LUSC	Early	exome-wide (without cancer genes)	wMIS	0.97491229	0.94653874	1.00384757
LUSC	Early	exome-wide (without cancer genes)	wNON	0.81656035	0.73257197	0.90675102
LUSC	Clonal	exome-wide (without cancer genes)	wMIS	1.01316536	0.99195153	1.03467938
LUSC	Clonal	exome-wide (without cancer genes)	wNON	0.85058293	0.78583244	0.91880113
LUSC	Late	exome-wide (without cancer genes)	wMIS	1.07521841	1.04735065	1.10356748
LUSC	Late	exome-wide (without cancer genes)	wNON	0.88689475	0.798327	0.98181335
LUSC	Subclonal	exome-wide (without cancer genes)	wMIS	1.08585703	1.05366387	1.11869964
LUSC	Subclonal	exome-wide (without cancer genes)	wNON	0.89724576	0.79440467	1.00862653

Table S11

#Table S11. Combined rules for genomic stratification based on the NLMT and MATCH clinical trials.

Arm	Drug 1	Drug 2	Target	Pathway	Genes	Contingent_wt	Aberrations	Histology	Mutation_specific_protein	Mutation_type_specific
NLMT	AZD4547		FGFR	FGFR	FGFR2	NA	Mutation	NSCLC	NA	NA
NLMT	AZD4547		FGFR	FGFR	FGFR3	NA	Mutation	NSCLC	NA	NA
NLMT	AZD2014		mTORC1, mTORC2	mTOR	TSC1	NA	Mutation	NSCLC	NA	NA
NLMT	AZD2014		mTORC1, mTORC2	mTOR	TSC2	NA	Mutation	NSCLC	NA	NA
NLMT	AZD2014		mTORC1, mTORC2	mTOR	STK11	NA	Mutation	NSCLC	NA	NA
NLMT	Palbociclib		CDK4, CDK6	p16	CDKN2A	RB1	Loss	NSCLC	NA	NA
NLMT	Palbociclib		CDK4, CDK7	p16	CDK4	RB1	Amplification	NSCLC	NA	NA
NLMT	Palbociclib		CDK4, CDK8	p16	CCND1	RB1	Amplification	NSCLC	NA	NA
NLMT	Palbociclib		CDK4, CDK9	p16	KRAS	RB1	Mutation	Adeno	NA	NA
NLMT	Crizotinib		MET, ROS1, ALK	TKI	MET	NA	Amplification	NSCLC	NA	NA
NLMT	Crizotinib		MET, ROS1, ALK	TKI	ROS1	NA	Fusion	NSCLC	NA	NA
NLMT	Selumetinib	Docetaxel	MEK1	MAP kinase	NF1	NA	Mutation	Adeno	NA	NA
NLMT	Selumetinib	Docetaxel	MEK1	MAP kinase	NF1	NA	Mutation	Squamous	NA	NA
NLMT	Selumetinib	Docetaxel	MEK1	MAP kinase	NRAS	NA	Mutation	Adeno	NA	NA
NLMT	AZD5363		AKT	PI3 kinase	PIK3CA	NA	Mutation	NSCLC	NA	NA
NLMT	AZD5363		AKT	PI3 kinase	PIK3CA	NA	Amplification	NSCLC	NA	NA
NLMT	AZD5363		AKT	PI3 kinase	PTEN	NA	Mutation	NSCLC	NA	NA
NLMT	AZD5363		AKT	PI3 kinase	PTEN	NA	Loss	NSCLC	NA	NA
NLMT	AZD5363		AKT	PI3 kinase	AKT1	NA	Mutation	NSCLC	NA	NA
NLMT	AZD9291		EGFR	PI3 kinase	EGFR	NA	Mutation	NSCLC	T790M	NA
MATCH	Afatinib		EGFR	MAP kinase	EGFR	NA	Mutation	NSCLC	NA	NA
MATCH	Afatinib		ERBB2	MAP kinase	ERBB2	NA	Mutation	NSCLC	NA	NA
MATCH	Crizotinib		MET	TKI	MET	NA	Amplification	NSCLC	NA	NA
MATCH	Crizotinib		MET	TKI	MET	NA	Mutation	NSCLC	NA	splice
MATCH	AZD9291		EGFR	MAP kinase	EGFR	NA	Mutation	NSCLC	T790M	NA
MATCH	Crizotinib		ALK	ALK	ALK	NA	Fusion	NSCLC	NA	NA
MATCH	Crizotinib		ROS1	ROS1	ROS1	NA	Fusion	NSCLC	NA	NA
MATCH	Dabrafenib	trametinib	BRAF	MAP kinase	BRAF	NA	Mutation	NSCLC	V600	NA
MATCH	Taselisib		PIK3CA	PI3K	PIK3CA	NA	Mutation	NSCLC	NA	NA
MATCH	GSK2636771		PTEN	PI3K	PTEN	NA	Mutation	NSCLC	NA	NA
MATCH	GSK2636771		PTEN	PI3K	PTEN	NA	Loss	NSCLC	NA	NA
MATCH	Ado-trastuzumab emtansine		ERBB2	MAP kinase	ERBB2	NA	Amplification	NSCLC	NA	NA
MATCH	Trametinib		BRAF	MAP kinase	BRAF	NA	Mutation	NSCLC	NA	NA
MATCH	Trametinib		NF1	MAP kinase	NF1	NA	Mutation	NSCLC	NA	NA
MATCH	Trametinib		GNAQ	GNAQ	GNAQ	NA	Mutation	NSCLC	NA	NA
MATCH	Trametinib		GNA11	GNA11	GNA11	NA	Mutation	NSCLC	NA	NA
MATCH	Vismodegib		SMO	SMO	SMO	NA	Mutation	NSCLC	NA	NA
MATCH	Vismodegib		PTCH1	PTCH1	PTCH1	NA	Mutation	NSCLC	NA	NA
MATCH	Defactinib		NF2	NF2	NF2	NA	Loss	NSCLC	NA	NA
MATCH	Sunitinib		KIT	KIT	KIT	NA	Mutation	NSCLC	NA	NA
MATCH	AZD4547		FGFR	FGFR	FGFR1	NA	Mutation	NSCLC	NA	NA
MATCH	AZD4547		FGFR	FGFR	FGFR2	NA	Mutation	NSCLC	NA	NA
MATCH	AZD4547		FGFR	FGFR	FGFR3	NA	Mutation	NSCLC	NA	NA
MATCH	Dasatinib		DDR2	DDR2	DDR2	NA	Mutation	NSCLC	NA	NA
MATCH	AZD5363		AKT1	AKT1	AKT1	NA	Mutation	NSCLC	NA	NA
MATCH	Binimetinib		NRAS	MAP kinase	NRAS	NA	Mutation	NSCLC	NA	NA
MATCH	Palbociclib		CCND1	CCND	CCND1	NA	Amplification	NSCLC	NA	NA
MATCH	Palbociclib		CCND2	CCND	CCND2	NA	Amplification	NSCLC	NA	NA
MATCH	Palbociclib		CCND3	CCND	CCND3	NA	Amplification	NSCLC	NA	NA

References

1. Kadota K, Nitadori J, Woo KM, et al. Comprehensive pathological analyses in lung squamous cell carcinoma: single cell invasion, nuclear diameter, and tumor budding are independent prognostic factors for worse outcomes. J Thorac Oncol 2014;9:1126-39.
2. Kadota K, Suzuki K, Kachala SS, et al. A grading system combining architectural features and mitotic count predicts recurrence in stage I lung adenocarcinoma. Mod Pathol 2012;25:1117-27.
3. von der Thusen JH, Tham YS, Pattenden H, et al. Prognostic significance of predominant histologic pattern and nuclear grade in resected adenocarcinoma of the lung: potential parameters for a grading system. J Thorac Oncol 2013;8:37-44.
4. Islam A, Rafiq S, Kirwan M, et al. Haematological recovery in dyskeratosis congenita patients treated with danazol. Br J Haematol 2013;162:854-6.
5. Gerlinger M, Horswell S, Larkin J, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nature genetics 2014;46:225-33.
6. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine 2012;366:883-92.
7. Pengelly RJ, Gibson J, Andreoletti G, Collins A, Mattocks CJ, Ennis S. A SNP profiling panel for sample tracking in whole-exome sequencing studies. Genome Med 2013;5:89.
8. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 2010;20:1297-303.
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
10. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
11. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research 2012;22:568-76.
12. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature biotechnology 2013;31:213-9.
13. Karolchik D, Hinrichs AS, Furey TS, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res 2004;32:D493-6.
14. Roth A, Khattra J, Yap D, et al. PyClone: statistical inference of clonal population structure in cancer. Nat Methods 2014;11:396-8.
15. Yates LR, Gerstung M, Knappskog S, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nature medicine 2015;21:751-9.
16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
17. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015;43:D805-11.
18. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495-501.
19. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012;489:519-25.
20. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543-50.
21. Campbell JD, Alexandrov A, Kim J, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nature genetics 2016;48:607-16.
22. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols 2009;4:1073-81.

23. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;Chapter 7:Unit7 20.
24. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575-6.
25. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-2.
26. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling 2013;6:pl1.
27. Cheng J, Vanneste E, Konings P, Voet T, Vermeesch JR, Moreau Y. Single-cell copy number variation detection. Genome biology 2011;12:R80.
28. Rimmer A, Phan H, Mathieson I, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nature genetics 2014;46:912-8.
29. Van Loo P, Nordgard SH, Lingjaerde OC, et al. Allele-specific copy number analysis of tumors. Proceedings of the National Academy of Sciences of the United States of America 2010;107:16910-5.
30. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med 2015;7:283ra54.
31. Malikic S, McPherson AW, Donmez N, Sahinalp CS. Clonality inference in multiple tumor samples using phylogeny. Bioinformatics 2015;31:1349-56.
32. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome biology 2016;17:31.
33. Greenman C, Wooster R, Futreal PA, Stratton MR, Easton DF. Statistical analysis of pathogenicity of somatic mutations in cancer. Genetics 2006;173:2187-98.
34. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science (New York, NY 2015;348:880-6.
35. Hodgkinson CL, Morrow CJ, Li Y, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nature medicine 2014;20:897-903.
36. Middleton G, Crack LR, Popat S, et al. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2015;26:2464-9.
37. Lappalainen I, Almeida-King J, Kumanduri V, et al. The European Genome-phenome Archive of human data consented for biomedical research. Nature genetics 2015;47:692-5.