

Article

Not peer-reviewed version

Radiogenomics Pilot Study: Association Between Radiomics and SNP-Based Microarray Copy Number Variation in Diagnosing Renal Oncocytoma and Chromophobe Renal Cell Carcinoma

[Abeer Alhussaini](#) *

Posted Date: 22 October 2024

doi: 10.20944/preprints202410.1596.v1

Keywords: chromophobe; oncocytoma; radiogenomics; computed tomography; renal masses



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Radiogenomics Pilot Study: Association Between Radiomics and SNP-Based Microarray Copy Number Variation in Diagnosing Renal Oncocytoma and Chromophobe Renal Cell Carcinoma

Abeer J. Alhussaini ^{1,2,*} , Abirami Veluchamy ³ , Neil Kernohan ⁴ , Ghulam Nabi ¹ , Colin N. A. Palmer ⁵  and J. Douglas Steele ¹ 

¹ Division of Imaging Sciences and Technology, School of Medicine, Ninewells Hospital, University of Dundee, Dundee, United Kingdom, DD1 9SY

² Department of Medical Imaging, Al-Amiri Hospital, Ministry of Health, Sulaibikhat, 13001, Kuwait City, Kuwait

³ Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, United Kingdom, DD1 9SY

⁴ Department of Pathology, Ninewells Hospital, Dundee, United Kingdom, DD9 1SY, UK

⁵ Division of Population Pharmacogenetics, Population Health and Genomics, Biomedical Research Centre, Ninewells Hospital and Medical School, University of Dundee, Dundee, United Kingdom, DD1 9SY

* Correspondence: a.j.a.h.m.alhussaini@dundee.ac.uk

Simple Summary: Renal oncocytomas (RO) and chromophobe renal cell carcinomas (ChRCC) are kidney tumours with similar imaging characteristics, making their differentiation challenging. Accurate distinction between these tumours is essential for guiding treatment strategies, with resection currently serving as the reference standard for diagnosis. However, advances in radiomics and genomics offer the potential to improve non-invasive diagnostic methods by identifying tumour-specific features through imaging and molecular data. This pilot study aims to explore the association between radiomics and single nucleotide polymorphism (SNP) array-based copy number variation (CNV) in diagnosing RO and ChRCC, with the goal of creating a radiogenomics map that integrates these data to better differentiate RO and ChRCC. By correlating imaging features with genomic alterations, we seek to develop a more precise and less invasive diagnostic approach for these renal tumours.

Abstract: Background: RO and ChRCC are kidney tumours with overlapping characteristics, making differentiation between them challenging. Objectives: The objective of this research is to create a radiogenomics map by correlating radiomic features to molecular phenotypes in ChRCC and RO, using resection as the gold standard. Methods: Fourteen patients (6 RO and 8 ChRCC) were included in the study. A total of 1,875 radiomic features were extracted from CT scans, alongside 632 cytobands containing 16,303 genes from the genomic data. Results: Feature selection algorithms applied to the radiomic features resulted in 13 key features. From the genomic data, 24 cytobands highly correlated with histology were selected and cross-correlated with the radiomic features. The analysis identified four radiomic features that were strongly associated with seven genomic features. Conclusion: These findings demonstrate the potential of integrating radiomic and genomic data to enhance the differential diagnosis of RO and ChRCC, paving the way for more precise and non-invasive diagnostic tools in clinical practice.

Keywords: chromophobe; oncocytoma; radiogenomics; computed tomography; renal masses

1. Introduction

Renal oncocytomas (RO) and chromophobe renal cell carcinomas (ChRCC) are two types of renal neoplasms that often present a significant diagnostic challenge due to their overlapping clinical and radiological characteristics [1,2]. Accurate differentiation between these tumours is essential for appropriate patient management and therapeutic decision-making. However, traditional diagnostic methods, including imaging and histopathology, sometimes fail to reliably distinguish between these entities [3–5]. This highlights the need for more precise and advanced diagnostic techniques.

Radiogenomics, an emerging field that combines radiomics and genomics, holds promise for enhancing the precision of tumour characterisation by correlating imaging features with genetic data [6,7]. Radiomics involves the extraction of a large number of quantitative features from medical images, transforming them into high-dimensional data that can reveal subtle imaging characteristics not discernible to the naked eye [8]. This data can provide valuable insights into tumour biology and behaviour.

Simultaneously, genomic technologies, particularly SNP-based microarray analysis, have advanced our understanding of the genetic landscape of various cancers. Single Nucleotide Polymorphisms are the most common type of genetic variation and their analysis can uncover critical information about tumour genetics. Specifically, chromosomal copy number variation analysis through SNP-based microarrays can provide a comprehensive genomic profile of tumours, revealing gains and losses in chromosomal regions that are associated with different tumour types [9–12].

Copy Number Variation represents losses of material when an individual has less than two copies and gains of material when an individual has more than the two expected copies. Additionally, CNVs can involve the heterozygous deletion of one allele, or duplication of a maternal or paternal chromosome or chromosomal region and concurrent loss of the other allele as shown in Figure A1. The genotyping of CNVs from SNP arrays is based on the analysis of the B allele frequency (BAF), which is a measure of heterozygosity, and the log R Ratio (LRR) value, which is a normalised measure of DNA content. LRR is the logged ratio of observed probe intensity to expected intensity, with any deviations from zero in this metric indicating evidence for copy number change. BAF represents the proportion of hybridised sample that carries the B allele as designated by the Infinium assay. In a normal sample, discrete BAFs of 0.0, 0.5, and 1.0 are expected for each locus, representing AA, AB, and BB genotypes. BAF values are typically used to normalise signal intensity, cleaning poor signals and representing either 0 or 1 for homozygous probes and 0.5 for heterozygous probes. The log R Ratio value is used to detect CNV regions and is generally averaged at 0 [13–15].

The goal of radiogenomics is to establish a robust connection between tumour imaging phenotypes and molecular markers, offering a non-invasive alternative to traditional genomic analysis. By using imaging signatures in place of genomic signatures, which typically require invasive tissue sampling, this research aims to provide a less invasive means of assessing genomic characteristics. Additionally, these relationships could help identify patient groups who may benefit from further genomic analysis. Whereas there is a plethora of research on radiogenomics [16–18], none of them has investigated the application of radiogenomics in the distinction of ChRCC and RO. Integrating radiomics with SNP-based microarray CNV analysis, this study seeks to develop a comprehensive diagnostic approach that improves the accuracy in differentiating these two clinically challenging renal neoplasms.

In this study, we explored the synergistic potential of radiomics and SNP-based microarray CNV analysis for the differential diagnosis of RO and ChRCC. By correlating detailed imaging features with chromosomal CNV data, we aim to advance diagnostic precision and our understanding of these renal tumours. This integrated approach could significantly enhance diagnostic accuracy, leading to better patient outcomes and more personalised treatment strategies.

2. Materials and Methods

2.1. Ethical Approval

This study received approval from the East of Scotland Research Ethical Service. Access to patients' medical healthcare data was granted under Caldicott Approval Number IGTAL9519 on August 25, 2021. Additionally, the Tissue Bank Committee [19] approved the application number TR000611 for this study on March 29, 2022.

2.2. Patients and Tissues

This research was a prospective study conducted using a database of 35 patients (10 with ChRCC and 25 with RO) from Ninewells Hospital, collected between 2011 and 2021. All cases were pathologically confirmed at the institution. Patients lacking approval from the tissue bank and patients with poor DNA yield were excluded from the study. The participants underwent pre-operative contrast-enhanced CT scan imaging. The imaging data, provided in DICOM format with a resolution of 512×512 pixels, was obtained from the institution's Picture Archiving and Communication System (PACS). Likewise, the patients had various types of tissue samples. These samples included formalin-fixed paraffin-embedded (FFPE) and fresh-frozen tissues, with some patients undergoing biopsies, resections, or both. However, the total number of patients accessible for the study was significantly reduced to 14 due to several obstacles i.e, 15 patients had poor DNA yield from the biopsy FFPE samples and issues with bead carryover during the DNA extraction process. An additional 6 patients were eliminated due to a lack of approval from the tissue bank. Figure A2 represents the exclusion and inclusion criteria for the study. For the clinical report information of the 35 patients refer to Tables A1-A5.

As a result of certain patients lacking sufficient tissue samples for analysis, the study's focus shifted to utilising fresh-frozen tissues and FFPE samples obtained from patients who underwent partial or radical nephrectomy, while excluding those who only underwent biopsy. For the next step, namely DNA extraction, we sectioned 72 tissue samples (3 sections per sample). We utilised 19 FFPE samples and 5 fresh-frozen samples, totalling 24 samples collected from 14 patients (6 RO and 8 ChRCC) for this research. The DNA quantification and purity is presented in Table A6. Refer to Appendix A.2 for sample preparation and SNP-based microarrays technical lab details. Figure A3 summaries the study's methodological process.

Out of these 6 RO patients:

- 4 with only 1 FFPE sample.
- 1 with 2 FFPE samples and 1 frozen tissue samples.
- 1 with both 2 FFPE.

For the 8 ChRCC patients:

- 4 with 1 FFPE sample.
- 1 with 2 FFPE samples.
- 1 with 2 FFPE and 1 frozen tissue.
- 1 with 2 FFPE and 2 frozen tissues.
- 1 with both 1 FFPE and 1 frozen tissue samples.

2.3. Statistical Analysis

A statistical analysis was conducted using the SciPy package in Python to evaluate the relationships between age, tumour size (3D), gender, and histopathology, with a significance level set at 0.05. Additionally, the associations between radiomic features, cytogenetic features, CN size, CN value, and histopathology were investigated. The Chi-square test (χ^2) and Student's T-test (t) were used to assess differences between groups, while Pearson's correlation coefficient (r) was used to measure the strength and direction of the linear relationship between variables.

2.4. Computed Tomography Scans

Data was captured using a Helical CT scanner from GE-Healthcare. The scanning parameters included a large body scan field of view (SFOV), a gantry rotation time of 0.7 seconds, a slice thickness of 1.25 mm, a pitch of 1375:1, and a detector coverage of 40 mm. The Noise Index (NI) was set to 30, with a Computed Tomography Dose Index Volume (CTDIvol) of 9.59 mGy. The X-ray tube-voltage was 120 kVp, and the X-ray tube-current ranged from 100 to 560 mA (auto-modulated) depending on patient size. The contrast agent used was intravenous Omnipaque 300 mls, administered at 80–100 mL per patient. A Bayer Centargo contrast pressure injector was used, with a flow rate of 3 ml/s for the

renal scan. The crucial pre-operative CT nephrographic stage, occurring 100 to 120 seconds after IV contrast injection, was utilised in this study. This stage, identified by studies [20,21] allows the clearest identification of renal lesions.

2.5. Tumour Volume Segmentation Technique

CT image slices for each patient were converted to 3D NIFTI (Neuroimaging Informatics Technology Initiative) format using Python version 3.9. These 3D images were then imported into 3D Slicer software, version 4.11.20210226, for segmentation. Manual segmentation was conducted on the 3D images, delineating the edges of the tumour slice-by-slice to obtain the volume of interest (VOI) as shown in Figure 1.

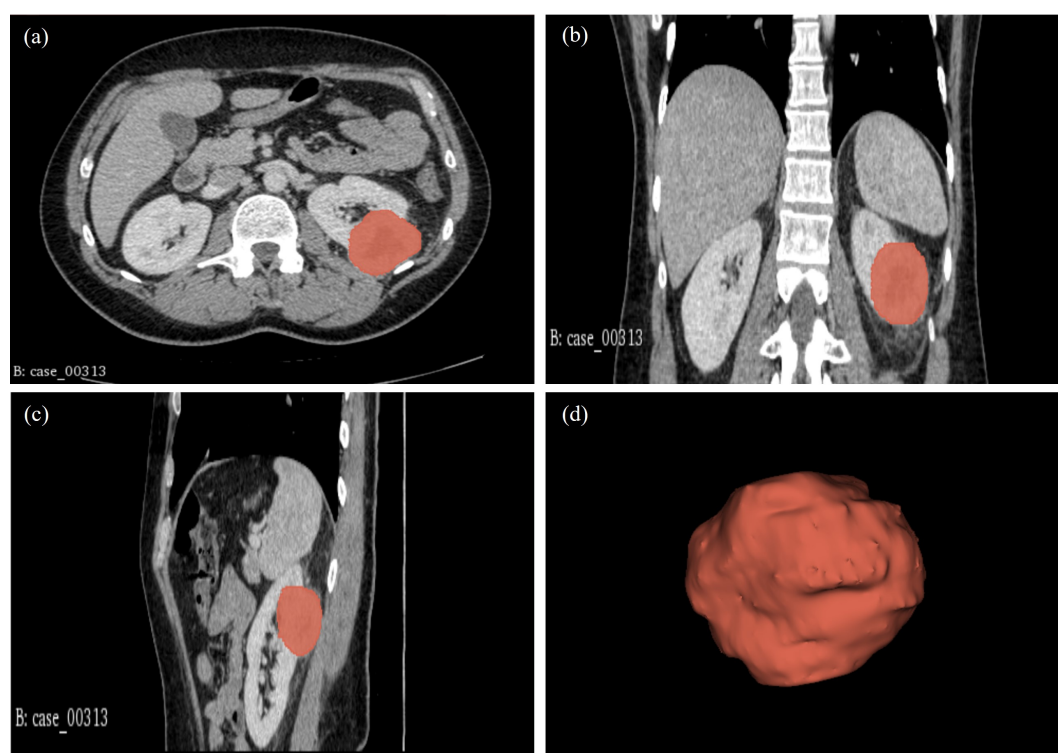


Figure 1. Manual segmentation of the 3D image slices using the 3D Slicer software: version 4.11.20210226 (a) CT scan axial plane; (b) Coronal plane; (c) Sagittal plane; and (d) 3D VOI.

The procedure was carried out twice by a blinded investigator (A.J.A.) with 14 years of experience in interpreting medical images, who was unaware of the tumour's final pathological grade. Another blinded investigator (A.J.), with 12 years of experience in medical imaging technology, conducted confirmatory segmentation. The inter-reader and intra-reader [22] agreement for the segmentations was determined using the Dice similarity coefficient (DSC). It effectively quantifies the overlap between two segmentations and provides a clear, interpretable measure of similarity, ranging from 0 (no overlap) to 1 (perfect overlap), making it ideal for evaluating the consistency and accuracy of segmentations in medical imaging. DSC robustness to small variations and wide acceptance in the field make it a reliable choice for assessing agreement in this study.

The Dice similarity coefficient (DSC) was calculated using the following formula [23]:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

where:

- $|A \cap B|$ is the size of the intersection between two sets A and B (i.e., the number of pixels in the case of image segmentation, that are common to both sets).
- $|A|$ and $|B|$ are the sizes of sets A and B , respectively (i.e., the number of pixels in each set).

In the context of image segmentation:

- A represents the set of pixels in the segmentation performed by one reader or at one time point.
- B represents the set of pixels in the segmentation performed by another reader or at another time point.

Subsequently, the segmentations were evaluated by an independent experienced urological surgical oncologist (G.N.), who considered radiology and histology reports. The gold standard for pathology diagnosis was assumed to be histopathology from partial or radical nephrectomy. The result of the segmentation was a binary mask of the tumour.

2.6. Radiomics Feature Computation

Texture descriptors of the features were computed using the PyRadiomics module in Python 3.6.1 [24]. The goal of PyRadiomics is to provide a standardised method for extracting radiomic features from medical images, minimising inter-observer variability [25]. The parameters used in PyRadiomics included a minimum region of interest (ROI) dimension of 2, a pad distance of 5, normalisation set to false, and a normaliser scale of 1. No outliers were removed, no re-sampling of pixel spacing was performed, and no pre-cropping of the image was done. SitkBSpline was used as the interpolator, with the bin-width set to 20.

On average, PyRadiomics generated approximately 1,500 features per image, allowing the extraction of seven feature classes from each 3D image. The extracted feature categories included: First-order (19 features), grey-level co-occurrence matrix (GLCM) (24 features), grey-level run-length matrix (GLRLM) (16 features), grey-level size-zone matrix (GLSZM) (16 features), grey-level dependence matrix (GLDM) (14 features), neighboring grey-tone difference matrix (NGTDM) (5 features), and 3D shape features (16 features). These features enable the computation of texture intensities and their distribution within the image [25].

In a previous study [22], combining original feature classes with filter features significantly enhanced model performance. Consequently, we extracted filter class features in addition to the original features. These filter classes included local binary pattern (LBP-3D), gradient, exponential, logarithm, square-root, square, Laplacian of Gaussian (LoG), and wavelet. Each filter was applied to every feature in the original feature classes. For example, since the first-order statistic feature class has 19 features, it also includes 19 LBP filter features. The filter class features were named by combining the name of the original feature with the name of the filter class [25].

2.7. Radiomics Feature Pre-Processing and Selection

The radiomics features were normalised using a standard scaler so that the mean of each feature was zero, with a standard deviation of one. The ground truth labels were annotated as 1 and 0 for ChRCC (positive class) and RO (negative class), respectively, in preparation for classification. Inter-feature correlation coefficients were computed, and when two features had a correlation coefficient greater than 0.8, one of the features was dropped. Thereafter, the least absolute shrinkage and selection operator (LASSO) model, recursive feature elimination (RFE), extreme gradient boosting (XGBoost), and random forest (RF) were used to select the essential features, as shown in Figure A4. Features selected by any two algorithms were included in the final feature set.

2.8. Tissue Data Scanning and Processing

Genomic DNA samples from 14 patients were extracted from formalin-fixed paraffin-embedded (FFPE), and frozen tissue samples. A total of 24 genomic DNA samples were genotyped using the Infinium CytoSNP-850K v1.2 BeadChip [26], according to the manufacturer's instructions. The

CytoSNP-850K v1.2 BeadChip is used to identify genetic and structural variants. The array contains approximately 850,000 single nucleotide polymorphism (SNP) markers spanning the entire genome, with an average probe spacing of 50-mer oligonucleotides, which covers cytogenomic-relevant genes from the International Collaboration for Clinical Genomics (ICCG) and the Cancer Cytogenomics Microarray Consortium (CCMC) for cancer research applications. It also provides enriched coverage for 3,262 dosage-sensitive genes, and the high 15× bead redundancy facilitates the identification of CNV calls with high confidence. Genotyped arrays were processed and scanned using iScan. The raw data were processed, quality assessed, and analysed initially using the Illumina Genome Studio genotyping module [27], based on the reference human genome (hg19/GRCh37). The clustering of intensities for all SNPs was performed using the raw data, the SNP manifest file (.bpm), and the standard cluster file (.egt). Genotyping calls (assigning a specific genetic version to each SNP) were conducted by the calling algorithm (GenCall) that is implemented in the GenTrain clustering algorithm. GenCall is more suitable for the identification of only common SNPs.

2.9. CNV Analysis

2.9.1. Performing CNV Analysis Using cnvPartition Algorithm

For CNV calling, Illumina provides its own algorithm named “CNV Partition”. This algorithm is to identify regions of the genome that are aberrant in copy numbers in samples based on intensity and allele frequency data. This tool is part of the Genome Studio platform 2.0 genotyping module, which can be freely downloaded from the Illumina support page [27].

The cnvPartition algorithm estimates copy number by comparing the observed log R ratio (LRR) and B allele frequency (BAF) for each locus, predicting the LRRs and BAFs for different copy number scenarios. This algorithm employs bivariate Gaussian distribution genotype models using LRRs and BAFs for each of 14 different copy number scenarios, as presented in Table A7 [15]. It provides a quick and straightforward overview of the data, boasting high specificity and a positive predictive rate for both deletions and duplications.

The processed data were analysed using cnvPartition 3.2.0 [28]. Additionally, three other algorithms for CNV analysis were employed: CNV Region Report, Homozygosity Detector, and LOH Score. B Allele Frequency and Log R Ratio plots are visualised using the Chromosome Browser within the Illumina Genome Viewer. Figure A5 illustrates the various standard parameter thresholds employed for CNV analysis, including a confidence threshold of 35 and a minimum probe count of 3. A stringent filtering criterion was implemented to exclude poor-quality samples; specifically, samples with a log R ratio standard deviation greater than 0.28 were removed from further analyses to reduce the incidence of false-positive CNVs. Normal regions will exhibit a CNV value of 2 and have an empty CNV confidence as represented in Table A8. Markers linked to identified CNVs will display values in the CNV confidence column and varying numbers in the CNV value column. Copy-neutral CNVs are indicated by a CNV value of 2, along with values present in the CNV confidence column [29]. cnvPartition is more user-friendly and best suited for routine analysis of common CNVs on Illumina platforms, offering quick and integrated CNV detection.

2.9.2. Performing CNV Analysis Using PennCNV Algorithm

The PennCNV tool was used to detect CNVs from SNP genotyping arrays. The PennCNV algorithm, which employs a Hidden Markov Model, was used to detect both CNVs and copy-neutral LOH (PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data Genome Research 17:1665-1674, 2007) [30]. The PennCNV algorithm was developed for genome-wide detection of CNVs using Illumina SNP data and is now available as a plug-in to Illumina’s Genome Studio analysis software. It utilises the log R ratio (total signal intensity) and B allele frequency (allelic intensity ratio). PennCNV is a more powerful

and flexible tool, ideal for advanced users working with complex datasets or needing detailed CNV detection. It excels at identifying rare and complex CNVs with greater sensitivity and adaptability.

2.10. Classify CNV

The detected CNVs were classified as pathogenic, likely pathogenic, benign, or of uncertain significance using the Classify CNV tool. This classification follows established guidelines from the International Standard Cytogenomic Array (ISCA) and the American College of Medical Genetics and Genomics (ACMG) guidelines [31,32]. The ClassifyCNV is a command-line tool that implements the 2019 ACMG guidelines to evaluate the pathogenicity of germline duplications and deletions [32]. The tool uses pre-parsed publicly available databases to calculate a pathogenicity score for each copy-number variant in accordance with the ACMG guidelines. It obtained a set of 17,683 duplications and 20,805 deletions from the nstd102 study in ClinVar (dbVar) [33]. The ClassifyCNV tool accepts a BED file as input, including chromosome, CNV start position, CNV end position, and CNV type (DEL or DUP). A study done by Gurbich and Ilinsky [32] illustrates how the ClassifyCNV algorithm works to classify the following CNVs based on pathogenicity scores: Pathogenic, Likely Pathogenic, Variant of Uncertain Significance (VUS) and Benign.

2.11. CNV Analysis Report and Data Extraction Using R Package

The Database of Genomic Variants (DGV) [34], the University of California, Santa Cruz (UCSC) Genome Browser [35], Copy Number Variation Explorer (CNVXPLOER) [36] and Copy Number Variation Clinical Viewer (CNV-ClinViewer) [37] were used to extract Cytogenetic location, CNV size and genomic coordinates. All statistical comparisons were performed by Chi-square testing using the R package with a significance level of 0.05.

2.12. Chromosomal Cytogenetic Band Selection

Different samples and variance types were segregated to obtain a list of common regions between samples and the associated genes with those regions. This was done by creating three (.bed) files per individual representing deletion, duplication and loss of heterozygosity from the original CNV data. These (.bed) files were ran through Bedsect [38], an online implementation of bed-tools intersect to find overlaps between the files i.e. the regions that vary in one than one samples by types of variation. Subsequently, the list of regions was analysed using UCSC's Table Browser [39], based on the defined regions from Genome Reference Consortium Human Build 37 (GRCh37), the National Center for Biotechnology Information Reference Sequence (NCBI RefSeq), and the complete RefSeq All dataset. The CNV values of the cytogenetic bands selected were correlated with histopathology to obtain the bands having the highest influence on histopathology. All bands with a correlation above 0.1 were retained for further analysis.

2.13. Model Construction

The cytogenetic band and the radiomics features which were found to be highly correlated to histopathology were combined and used to train a machine learning algorithm to predict the tumour subtype. A Random Forest (RF) was implemented using cross validation technique. Accuracy, Sensitivity, Specificity, AUC, MCC and F1 were used to evaluate the model performance.

3. Results

3.1. Statistical Analysis

Statistical analysis was conducted on patients' age, tumour size and gender. From the analysis, it was found that there was no significant difference between tumour size ($p = 0.791$), gender ($p = 1$) or age ($p = 0.653$) and histopathology. Detailed results are shown in Table 1 below. The inter-reader and intra-reader [22] agreement for the segmentations was assessed using the Dice similarity coefficient, resulting in scores of 0.93 and 0.89, respectively, indicating strong agreement for tumour segmentation.

Table 1. Statistical demographic characteristics of the patients’ data.

Patients Characteristics				
Variable		RO	ChRCC	p-Value
n = 14	Age (Mean ± SD)	63.5±8.67	61.40±7.13	0.653
	tumour Size	3.60±1.47	3.80±1.09	0.791
	Gender			1
	Male	6 (42.86%)	7 (50.0%)	
	Female	0 (0%)	1 (7.14%)	

* Statistical significant difference is considered at 0.05 significance level.

3.2. Radiomics Feature Extraction and Selection

A total of 1,875 features were extracted from the CT scan images. LASSO, RFE, XGBoost, and RF each selected the top 14 features from the original feature set as shown in Figures A6–A9. Thirteen features were found to be shared by at least two feature selection algorithms as shown in Figure A4 and Table A9 and were therefore included in the final feature set. Figure 2 and Table A9 highlights the correlation heat map of these 13 features with the histology target. Additionally, Figure A10 presents a heat map showing the correlation between the features and histopathology target per patient. Figure A11 displays a radar plot showing the radiomics feature profiles for each patient, highlighting patterns that correlate with specific histopathological findings. Figure 3 represents OLS regression analysis of radiomic features from 78 patients [22], showing the coefficients, standard errors, t-values, and p-values for 13 selected features. Several features, including ‘Wavelet LLH GLSZM Small Area Low Gray Level Emphasis’ (p=0.032), ‘Logarithm GLDM Large Dependence High Gray Level Emphasis’ (p=0.039) and ‘Log Sigma 2mm 3D GLSZM Small Area Low Gray Level Emphasis’ were found to be statistically significant $p \leq 0.05$, suggesting their potential predictive value in the analysis.

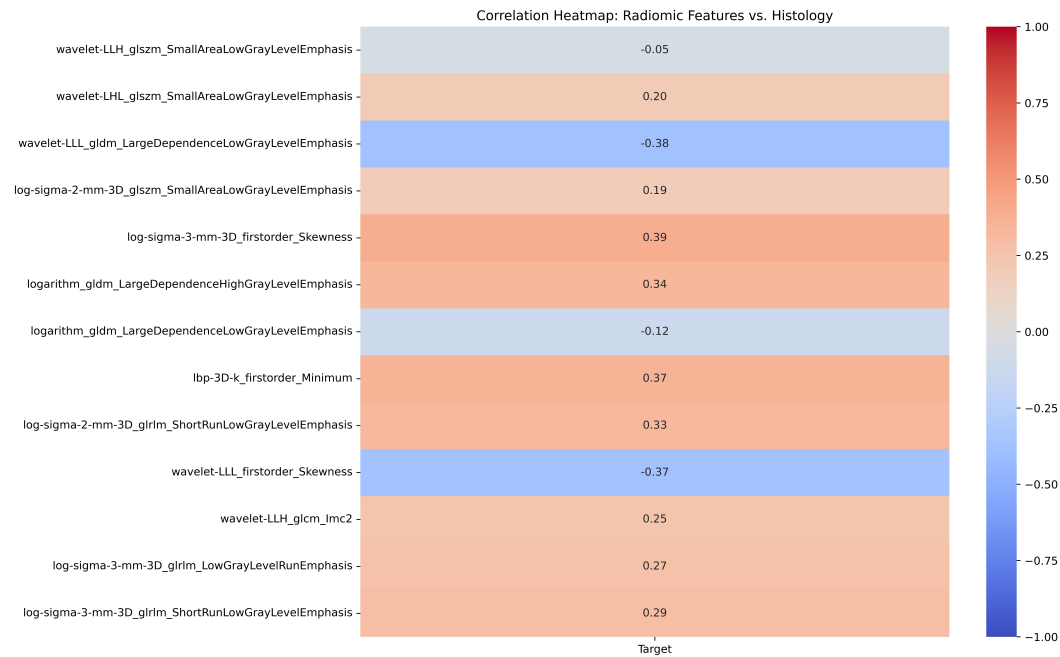


Figure 2. Representation of the correlation between the 13 radiomic features and the histological target.

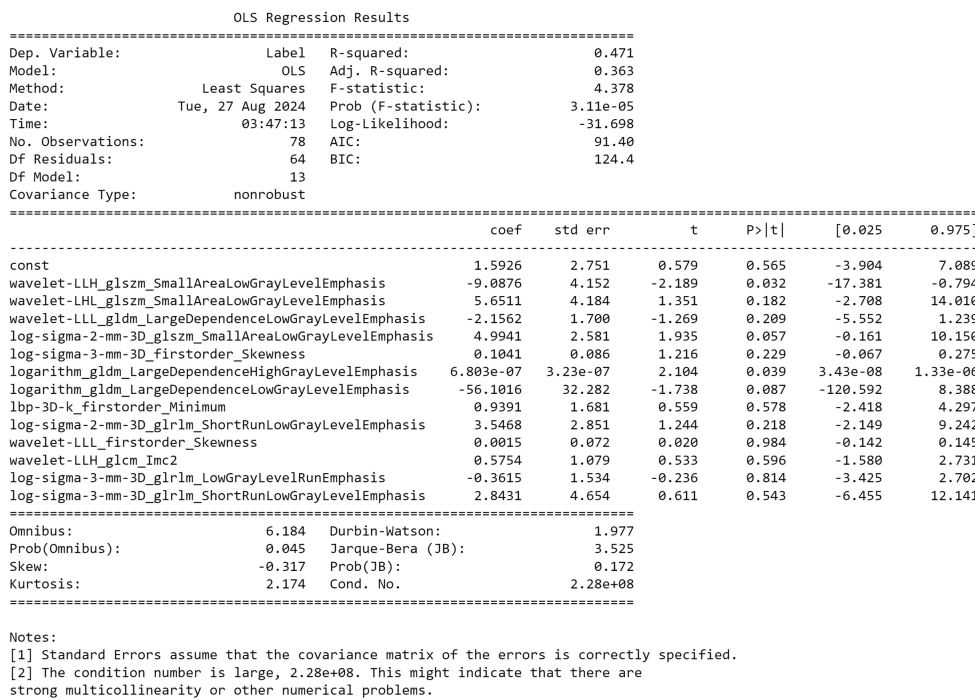


Figure 3. Representation of the OLS regression analysis of radiomic features from 78 patients.

3.3. Genomics Feature Extraction and Selection

ChXp21.2 (DMD) and ChXp11.23 (DYNLT3) were found to be the most correlated cytobands to histopathology with a correlation of 0.61. Moreover 22 cytobands were found to be highly correlated with histopathology as highlighted in the heat map in Figure 4 and Table A10.

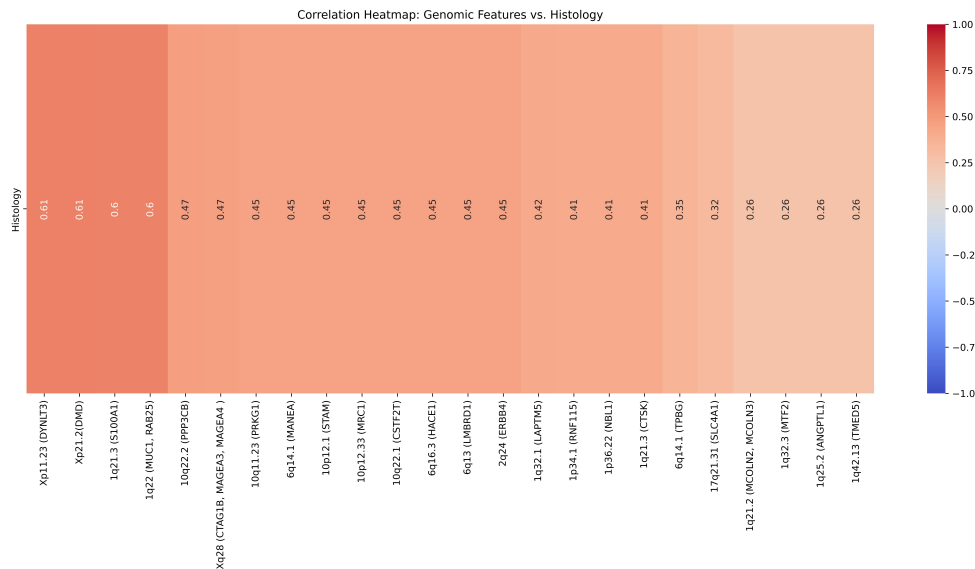


Figure 4. Representation of the correlation between the 24 genomic features and the histological target.

3.4. CNV Analysis

The CNV partition and PennCNV were used to detect CNVs from the data. The results were compared. The CNV region report generates three separate CNV reports: it lists each CNV and LOH region, estimates allele-specific copy numbers for each probe entry, and creates PLINK CNV input files.

- Standard Report: Lists each copy number variation and loss of heterozygosity (LOH) region for each sample.
- Allele-Specific Copy Number Report: Reports copy number-informed genotypes, such as A- and ABB.
- PLINK CNV Input Report: Creates input files for PLINK CNV Analysis Software.

As mentioned in the method section, CNV value and CNV confidence values are presented in the standard report file. For each sample, type of CNVs as loss (L) or gain (G) or LOH on different cytogenic location were determined based on the CNV values refer to Table A8. Based on the quality control criteria and sample types, we have selected only CNV data for 14 patients. Table A12 presents the number of amplified, deleted, and LOH segments for each patient, along with the percentage of the genome affected by copy number amplification, loss, and LOH. The percentage of genome copy number variation was calculated using the following equation: $CNV\% \text{ (Duplication (DUP), Deletion (DEL), or Loss of Heterozygosity (LOH))} = (\text{number of segments of a specific type for the patient} / \text{total number of segments of that type across all patients}) \times 100$. Additionally, the table provides the mean size of the segments affected by amplification, deletion, and loss of heterozygosity. Figure 5 illustrates the percentage of the CNV value per chromosome across histopathology. Table A13 presents the statistical evidence of chromosomal differences between ChrCC and RO.

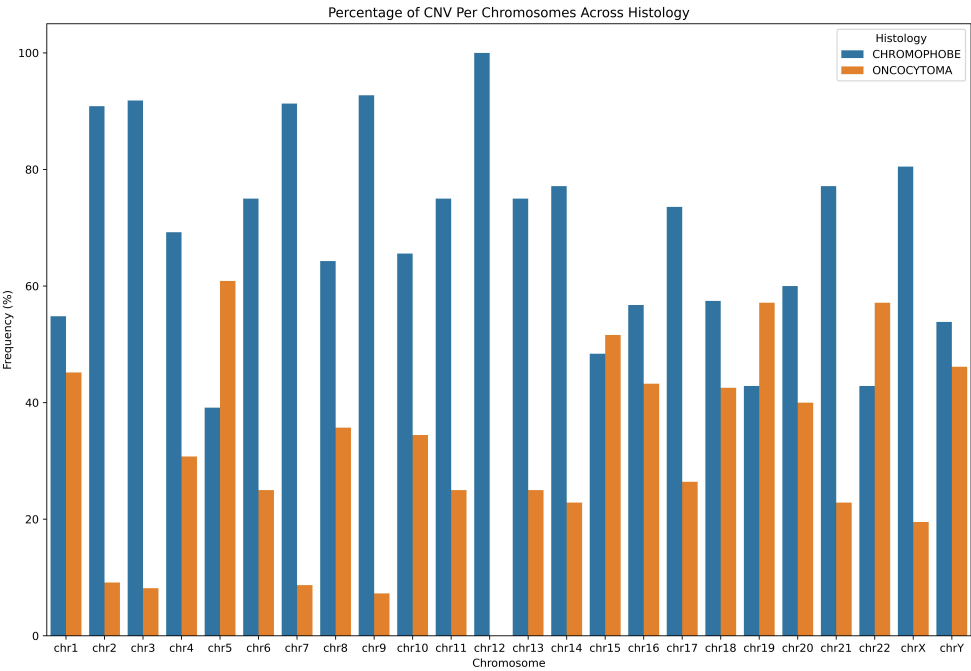


Figure 5. Percentage of CNV per chromosomes across histology.

3.5. Visualisation of Results in Illumina Genome Viewer

Illumina Genome Viewer was used to visualise CNV analysis results. Figure 6 illustrates the B-allele frequency and the normalised intensity, as represented by the Log R ratio, displayed in the context of copy number on the top right panel. Cytogenetic location and gene annotations of a specific sample are displayed on the bottom right panel. In Figure 6, the left panel shows the graphic display of the detected CNV regions across all samples selected for the analysis. Figure 7 represents the CNV regions across all chromosomes.

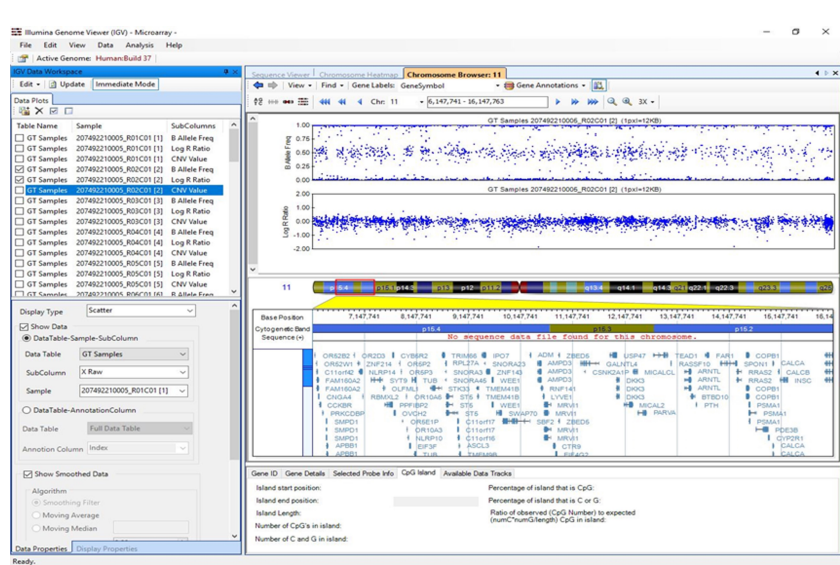


Figure 6. Representation of the results visualisation of the CNV analysis using Illumina genome viewer.

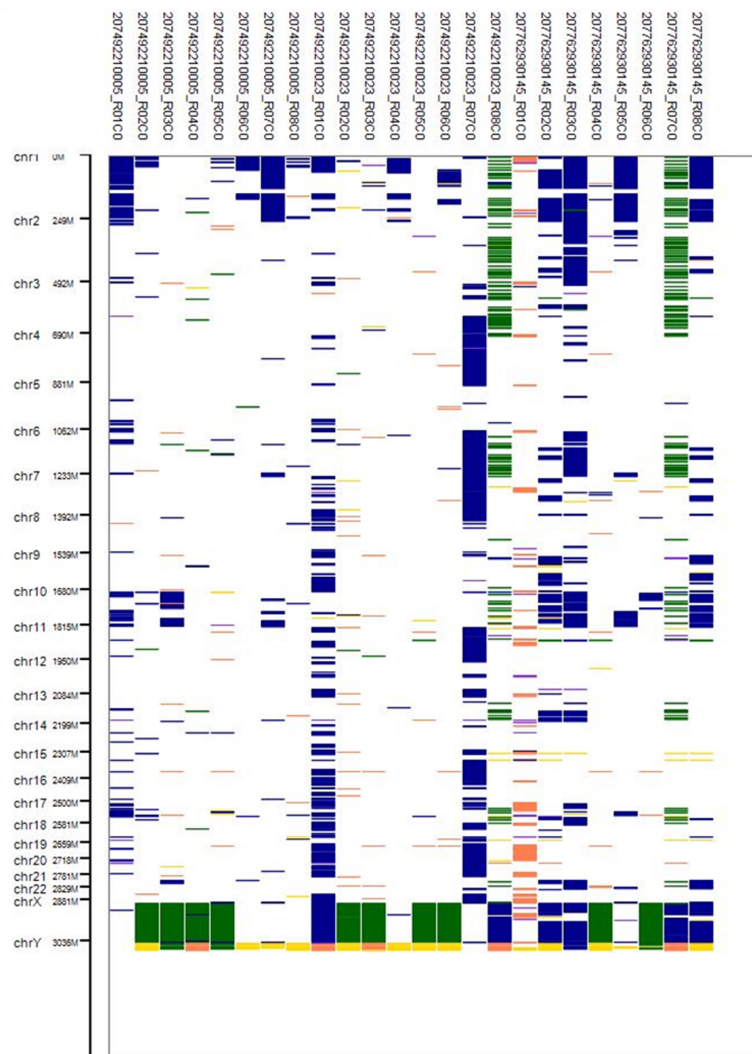


Figure 7. Representation of the CNV regions across all chromosomes in the 24 tissue samples. Dark green for CNV LOH, dark blue and blue violet for gain/duplication, gold and coral for CNV deletion/loss.

3.6. Classification of CNVs

The numeric pathogenicity score [32], calculated by ClassifyCNV, is converted to pathogenicity classification using the following cut-offs:

- Benign Variant: Scores less than or equal to -0.99.
- Likely Benign Variant: Scores between -0.90 and -0.98.
- Variant of Uncertain Significance: Scores between -0.89 and 0.89.
- Likely Pathogenic Variant: Scores between 0.90 and 0.98.
- Pathogenic Variant: Scores greater than or equal to 0.99.

This classification [32] helps in understanding the potential impact of genetic variants on health, guiding further medical investigation or action. It also includes dosage-sensitive genes contained within the CNV and a list of all protein-coding genes in the CNV. The classification was based on the size of the CNV, gene content, the inheritance pattern, and information in the medical literature and public databases. Table A14 and Figure A12 represents CNV classifications and types identified in ChrCC and RO subtypes.

3.7. Radiogenomics Analysis

Figure 8 displays the correlation heat map between the 13 radiomic imaging metrics and the 24 molecular genomic features that were highly correlated with the histology target. Table A11 identifies the radiomic and genomic features with a correlation greater than 0.4.

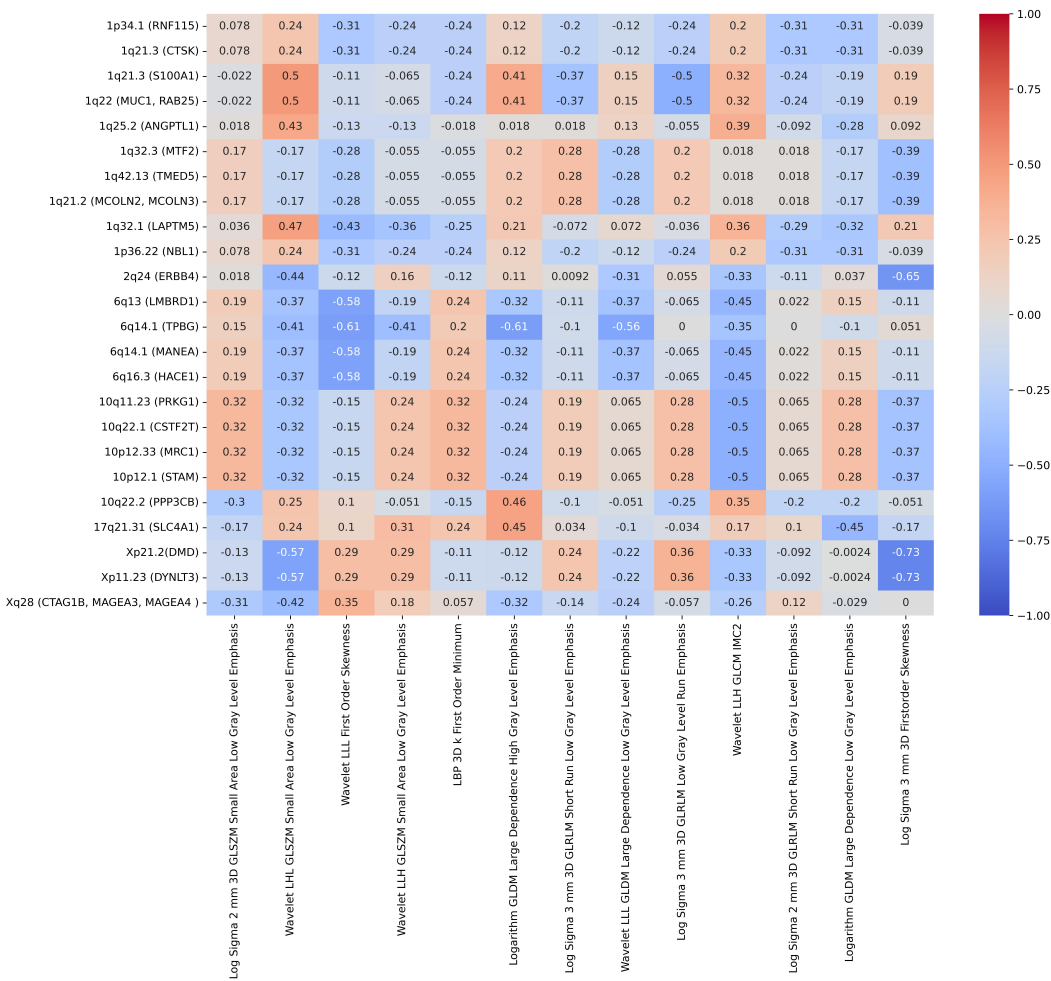


Figure 8. Analysis of the correlation between 13 radiomic and 24 genomic features.

3.8. Model Construction

The Random Forest (RF) algorithm was constructed using 201 estimators (trees) and different cut-offs for Pearson’s correlation coefficient (r), with the results presented in Table A15. The receiver operating characteristic (ROC) curve, illustrating the area under the curve (AUC-ROC) for the radiogenomics model with a correlation (r) greater than 0.55, is depicted in Figure 9.

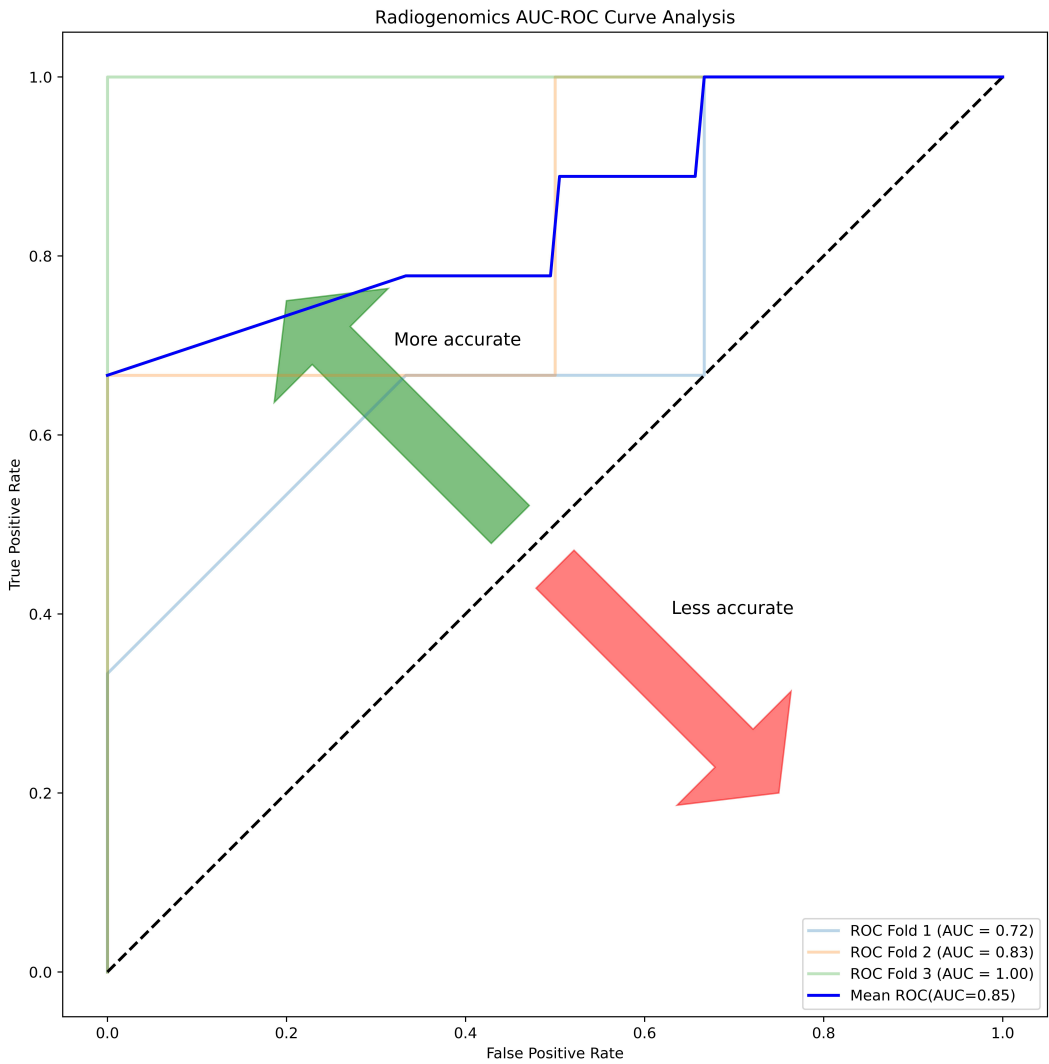


Figure 9. The AUC-ROC for the radiogenomics model with a Pearson’s correlation coefficient (r) greater than 0.55 was obtained using the following features: ‘Log Sigma 3 mm 3D Firstorder Skewness’, ‘Logarithm GLDM Large Dependence High Gray Level Emphasis’, ‘Wavelet LLL Firstorder Skewness’, and ‘Wavelet LHL GLSZM Small Area Low Gray Level Emphasis’.

4. Discussion

Renal tumours are highly heterogeneous, encompassing at least 16 distinct subtypes, with four of these being the most prevalent. Clear cell renal cell carcinoma, which arises from the proximal tubular epithelial cells, is the most prevalent subtype of RCC, representing 70–80% of cases. The next most common subtypes are papillary RCC (10–15%), chromophobe RCC (5%), and collecting duct RCC, which accounts for less than 1% of cases. Renal oncocytomas (ROs) account for approximately 3–7% of all adult renal neoplasms. These tumours are most commonly detected in individuals in their seventh decade of life. In contrast, the incidence of chromophobe renal cell carcinomas (ChRCCs) peaks in the sixth decade. Men are affected by ROs about twice as often as women, whereas ChRCCs generally affect men and women equally [40].

Distinguishing between RO and ChRCC is challenging due to their overlapping clinical features. Both tumours originate from the intercalated cells of the collecting ducts, which accounts for their histomorphologic, immunophenotypic, ultrastructural, and molecular similarities [41]. Typically, these tumours present as asymptomatic renal masses, often discovered incidentally during imaging performed for unrelated conditions. When symptoms do manifest, they may include weight loss, anorexia, flank pain, palpable masses, haematuria, and non-specific constitutional symptoms [42]. However, these symptoms are more commonly associated with ChRCC than RO [40].

The distinction between ChRCC and RO is crucial due to the significant differences in their prognosis and treatment. RO, being a benign tumour with no risk of metastasis, should be identified in advance to avoid unnecessary treatments. In contrast, ChRCC is a malignant tumour with the potential to spread, requiring more intensive treatment, such as surgical resection and constant monitoring. Misdiagnosing these conditions can lead to overtreatment in the case of RO or inadequate treatment in the case of ChRCC, emphasising the importance of accurate differentiation for effective therapeutic decision-making and enhanced patient outcomes.

On CT scans, ChRCC can appear as a well-defined, smooth, homogeneous mass with enhancement patterns similar to those of RO, making differentiation between the two particularly challenging [3,43,44]. Despite advancements in imaging techniques like CT and MRI, both tumours often present with overlapping radiographic features, such as well-circumscribed, homogeneous masses that enhance with contrast, leading to difficulties in distinguishing them [4,40,45]. Moreover, the similar growth rates of these tumours further complicate their clinical differentiation [46,47].

Histopathologically, the similarities between RO and ChRCC are particularly evident, especially with the eosinophilic variant of ChRCC, which closely resembles RO. Both tumours can feature large, polygonal cells with eosinophilic cytoplasm, making accurate diagnosis challenging [40]. This overlap necessitates further histochemical and immunohistochemical analyses to distinguish between these two entities accurately, given the malignant potential of ChRCC and the benign nature of RO [48]. The complexity of differentiating these tumours underscores the importance of detailed diagnostic investigations to ensure appropriate treatment strategies.

The human genome exhibits a wide range of genetic structural variations, from large-scale chromosomal abnormalities to single nucleotide polymorphisms (SNPs) [49]. Although SNPs were once considered the primary source of phenotypic variation, recent studies have highlighted the crucial role of copy number variants (CNVs) in contributing to genetic diversity. CNVs, which involve changes in DNA segments longer than 50 base pairs, can affect larger genomic regions, including partial or complete genes, making them a significant component of human genetic diversity [49,50].

Numerous cytogenetic abnormalities are known to exist in cancer cells and genome-wide studies; and are therefore used to identify chromosomal aberrations [51]. Accurate detection and clinical annotation of CNVs is essential because they can disrupt genes and regulatory elements and cause benignities or illnesses [32]. They have been connected to a number of hereditary illnesses, including autoimmune diseases, neurodevelopmental problems, and autism spectrum disorders [32,52]. As renal oncocytoma and chromophobe renal cell carcinoma share many characteristics, conventional approaches for distinguishing between the two tumours, such as imaging and histopathological techniques, frequently lack precision.

4.1. Comparison with Related Methodological Literature

In this study a correlation of radiomic texture features extracted from computed tomography images and SNP-based microarray copy number variation cytogenomic features was performed. From the findings of the study, it is evident that there is a correlation between radiomic and genomic features, an outcome which has never been found by other studies in the distinction of chromophobe and oncocytoma renal masses. Nonetheless, the research found that the radiogenomics model for features with correlation > 0.55 gave an accuracy, sensitivity, Specificity, and AUC of 81.25, 75.00, 87.50 and 85.00 respectively.

The research had 16303 genes in total, out of which 97 genes were either significantly overlapping between the two tumour subtypes or were majorly present in either of the tumours. These genes were found in 61 different cytobands and were therefore correlated with tumour subtype. On correlation 24 cytobands containing 28 affected genes were found to be highly correlated to histopathology. These cytobands were found in chromosomes 1, 2, 6, 10, 17 and X. This is comparable to previous studies [41,53–59] which found these chromosomes to be associated with either chromophobe or oncocytoma. Specifically, 1p34.1 (RNF115), 1q21.3 (CTSK), 1q21.3 (S100A1), 1q22 (MUC1, RAB25), 1q25.2 (ANGPTL1), 1q32.3 (MTF2), 1q42.13 (TMED5), 1q21.2 (MCOLN2, MCOLN3), 1q32.1 (LAPTM5) and 1p36.22 (NBL1) were all found in chromosome 1.

RNF115 has emerged as a significant gene in our study, as well as in the broader study of renal tumours, particularly in differentiating between chromophobe renal cell carcinoma (ChRCC) and renal oncocytoma (RO). Research indicates that RNF115 is consistently expressed in all cases of renal oncocytoma and in oncocytic neoplasms favouring oncocytoma, but it is barely detectable in ChRCC [33,60–63]. The study by Iakymenko et al. [64] investigated the expression of the CTSK gene and its product, Cathepsin K, in RO and ChRCC. The findings revealed that Cathepsin K was positively expressed in both tumour types, with stronger staining observed in renal oncocytoma compared to the weaker, more membranous staining in ChRCC. This expression pattern suggests that while Cathepsin K is present in both types of tumours, the differences in staining intensity might serve as a definitive marker for differentiating between the two. In the study by Li et al. [65], the S100A1 gene was expressed in 7 out of 8 RO cases, but not in any of the ChRCC cases. This gene expression pattern further supports the use of S100A1 as a diagnostic tool for differentiating between RO and ChRCC. These findings suggest that S100A1 is a useful marker, providing a reliable method for the differential diagnosis of renal RO and ChRCC. The study by Yusenko [66] discussed the expression of the MUC1 gene in the context of differentiating ChRCC from RO. MUC1, also known as epithelial membrane antigen (EMA), showed higher expression levels in ChRCC compared to RO. The study highlighted that while MUC1 is expressed in both tumour types, its stronger and more consistent expression in ChRCC makes it a useful marker in the differential diagnosis between these two renal tumours [66,67]. The findings in our study align with previous research, suggesting that copy number variations (CNVs) in RNF115, CTSK, S100A1, MUC1, RAB25 [68,69], ANGPTL1 [68], MTF2 [70], TMED5 [70], MCOLN2 [68], MCOLN3 [68], LAPTM5 [68] and NBL1 [71,72] could be valuable biomarkers for distinguishing benign RO from malignant ChRCC, which is crucial for accurate diagnosis and treatment planning.

The present research found only a single cytoband in Chromosome 2 containing ERBB4 gene 2q24 (ERBB4). The study by Liu et al. [73] demonstrated that hemizygous deletions of the ERBB4 gene were found in 33% of ChRCC cases, but not in any RO cases, indicating that ERBB4 deletions could serve as a useful marker for distinguishing between these two tumour types. In our study we found copy number alteration of 2q24 (ERBB4) in 38% of the ChRCC and none in RO which is comparable to Liu et al. [73].

In a study aimed at distinguishing ChRCC from RO, the genes LMBRD1, TPBG, MANEA, and HACE1 were integral components of a 30-gene signature known as chromophobe and oncocytoma related gene signature (COGS). These genes were selected based on their differential expression patterns, which were identified through univariate gene expression and ROC curve analyses. The inclusion of these genes in the COGS signature contributed to the study's ability to achieve a classification accuracy of 97.8% in the discovery dataset and 100% in the validation dataset, effectively differentiating ChRCC from RO using machine learning models. The cytobands 6q13 (LMBRD1), 6q14.1 (TPBG), 6q14.1 (MANEA), and 6q16.3 (HACE1) were found in our study to differentiate RO from ChRCC [68].

According to the study done by Yusenko et al. [70], the genes PRKG1 and CSTF2T are located within a region on chromosome 10q11.23–q21.1, where overlapping alterations were observed in both ChRCC and RO. The study by Krill-Burger et al. [57] found that both ChRCC and RO exhibit significant genomic alterations, including copy number variations, in regions where the MRC1 and STAM genes are located. Specifically, deletions involving MRC1 and STAM were identified in ChRCC, with MRC1

being entirely deleted and STAM partially deleted. These deletions were significant in distinguishing ChRCC from other types, including RO. In our study 10q11.23 (PRKG1), 10q22.1 (CSTF2T), 10p12.33 (MRC1) and 10p12.1 (STAM) occurred in 37.5% in ChRCC and non in RO indicating the potential of the two cytoband in differentiating the two tumours.

In the study conducted by Satter et al. [68], PPP3CB was identified as one of the top 197 genes through differential gene-expression and receiver-operating characteristic (ROC) analysis. This gene demonstrated a significant area under the curve (AUC) of 0.9 or higher, underscoring its potential role in distinguishing between chromophobe renal cell carcinoma and renal oncocytoma. Similarly, our study identified a copy number variation in the cytoband 10q22.2, which includes PPP3CB, in 33.33% of renal oncocytoma cases, with no such variation observed in chromophobe renal cell carcinoma cases.

The SLC4A1 gene, which encodes for a solute carrier family 4 member 1, plays a significant role in differentiating between RO and ChRCC. According to the study conducted by Molnar et al. [74] SLC4A1 was expressed in 60% of RO but only in 11% of ChRCC. This difference in expression suggests that while SLC4A1 is more commonly associated with ROs, its lower expression in ChRCCs can still be present, albeit less frequently. The findings suggest that SLC4A1, could be used in the differential diagnosis between RO and ChRCC, especially when morphological features overlap [74,75]. In our study the cytoband 17q21.31 (SLC4A1) occurred in only in 16.67% of RO which does not provide sufficient proof on its ability to distinguish RO from ChRCC.

In the study by Satter et al. [68] the DMD and DYNLT3 genes are included in the list of 197 top genes identified for their potential to differentiate ChRCC from RO. These genes were selected based on their differential expression and their ability to contribute to a gene signature (COGS) aimed at distinguishing between these two types of renal tumours. In our study Xp21.2 (DMD) and Xp11.23 (DYNLT3) occurred in 87.5% of ChRCC and 33.33% in RO patients.

The CTAG1B gene, also known as NY-ESO-1, was found to be expressed in 6 out of 18 ChRCCs and 15 out of 17 RO, suggesting its potential utility in diagnosing these tumours [76]. The study by Demirović et al. [77] investigated the expression of MAGE-A3/4 and NY-ESO-1 in RO and ChRCC, finding significant differences in the expression of these cancer testis antigens between the two tumour types, which may have diagnostic implications. In our study Xq28 (CTAG1B, MAGEA4, MAGEA3) occurred in 75% of ChRCC and 33.33% RO cases.

In radiomics, a total of 1,875 features [25] were initially extracted, however after applying several feature reduction techniques, this number was reduced to 13 final features. These selected features belong to five radiomic feature classes and four filter classes. Among the final features, two were First Order features—Skewness and Minimum—each associated with three filter classes: 'Log Sigma 3 mm 3D', 'Wavelet LLL', and 'LBP 3D k'. The GLCM class contributed one feature, the 'Informational Measure of Correlation 2' (IMC2), which was combined with the 'Wavelet LLH' filter. Additionally, two GLDM features were selected: 'Large Dependence Low Gray Level Emphasis' (LDLGLE) and 'Large Dependence High Gray Level Emphasis' (LDHGLE), each combined with the 'Wavelet LLL' and Logarithm filters. The GLRLM class included two features: 'Low Gray Level Run Emphasis' (LGLRE) and 'Short Run Low Gray Level Emphasis' (SRLGLE), which were combined with the 'Log Sigma 3 mm 3D' and 'Log Sigma 2 mm 3D' filters. Finally, one GLSZM feature having three filter types: 'Log Sigma 2 mm 3D', 'Wavelet LHL', and 'Wavelet LLH' was selected.

The 'Log Sigma 3 mm 3D First Order Skewness' and 'Wavelet LLL First Order Skewness' are both radiomic features that measure the asymmetry of the intensity distribution within a 3D medical image [25,78], but they do so using different filtering techniques. The 'Log Sigma 3 mm 3D First Order Skewness' involves applying a '3D Gaussian' smoothing filter with a 'sigma of 3 mm', followed by a logarithmic transformation of the image intensities. This process enhances subtle textural details, particularly in lower intensity ranges, and the skewness metric quantifies the asymmetry in the distribution of these intensities. A positive Skewness indicates that the distribution leans towards lower intensity values, while a negative Skewness suggests a bias towards higher intensities. This

property is particularly useful in highlighting variations in tissue composition that may be indicative of specific pathologies. On the other hand, the 'Wavelet LLL First Order Skewness' is derived from a different type of filter—the wavelet transform. The 'Wavelet LLL' filter applies low-pass filtering across all three dimensions (horizontal, vertical, and diagonal), which smoothens the image and emphasises large-scale, low-frequency components [79]. After this transformation, the Skewness is calculated to assess the asymmetry of the intensity distribution in the filtered image. This feature is effective in capturing broader structural patterns within the tissue, which can be crucial for distinguishing between different types of tissues or abnormalities. In summary, while both features measure Skewness, the 'Log Sigma 3 mm 3D First Order Skewness' focuses on fine details and intensity variations, particularly in lower intensity ranges, and the 'Wavelet LLL First Order Skewness' emphasises larger structural patterns by smoothing the image across multiple dimensions. Both features provide complementary insights into the textural characteristics of tissues, aiding in the differentiation of complex medical conditions like Chromophobe Renal Cell Carcinoma and Renal Oncocytoma. Our findings are similar to what have been highlighted by previous research [80–83].

'LBP 3D k First Order Minimum' is a combination of Local Binary Patterns (LBP) in three dimensions with 'First Order statistical Minimum' value. LBP is a texture descriptor that captures the local spatial structure of images by analysing the relationship between a pixel and its surrounding neighbours [84]. When applied in 3D, it extends this analysis to volumetric data, making it highly effective for capturing complex texture patterns in medical images [85–89]. The 'First Order Minimum' aspect focuses on the lowest intensity value in the voxel intensity distribution, providing insight into the darkest or least intense areas within the segmented volume [90]. This combination is particularly useful in radiomics for identifying and characterising subtle variations in texture, which could be indicative of specific tissue properties or pathological conditions.

The radiomic feature 'Wavelet LLH GLCM IMC2' represents a combination of wavelet transformation and Gray-Level Co-occurrence Matrix (GLCM) analysis focused on the 'Informational Measure of Correlation 2' (IMC2) [80,82,83,91]. Wavelet transformation is a powerful tool that decomposes an image into different frequency components, allowing for the analysis of various levels of detail [79]. The 'LLH filter' specifically applies low-pass filtering in the first two dimensions (L and L) and high-pass filtering in the third dimension (H), capturing the horizontal details within the image. GLCM is a texture analysis method that evaluates the spatial relationship between pixel intensities, and IMC2 is a specific feature derived from GLCM, which quantifies the complexity of the texture by measuring the correlation between pixel pairs in the image. High values of IMC2 indicate a more complex and less predictable texture [92–94]. By combining these techniques, the 'Wavelet LLH GLCM IMC2' feature provides a sophisticated measure of texture that is sensitive to subtle patterns in the image, particularly those related to structural complexity and spatial relationships, making it valuable in distinguishing between ChrCC and RO.

The three radiomic features—'Wavelet LLL GLDM Large Dependence Low Gray Level Emphasis', 'Logarithm GLDM Large Dependence Low Gray Level Emphasis', and 'Logarithm GLDM Large Dependence High Gray Level Emphasis'—are advanced texture metrics used in radiomic analysis to capture subtle tissue characteristics in medical images [25]. The Gray Level Dependence Matrix (GLDM) features focus on the relationship between a voxel and its dependent neighbours, emphasising different aspects of texture. 'Wavelet LLL GLDM Large Dependence Low Gray Level Emphasis' is derived from applying a Wavelet transformation with a low-pass filter across all three axes (LLL), which highlights the broader, smooth patterns in the image. The 'Large Dependence Low Gray Level Emphasis' then emphasises regions in the image where large groups of low-intensity pixels are clustered together, capturing homogeneity in low-density areas [95]. 'Logarithm GLDM Large Dependence Low Gray Level Emphasis', is similar to the first but uses a logarithmic transformation instead of a wavelet filter. The logarithm filter can enhance subtle differences in pixel intensity, making this feature particularly useful for detecting fine, low-intensity patterns in the image that might be missed by other filters. 'Logarithm GLDM Large Dependence High Gray Level Emphasis', unlike

the previous two, emphasises areas with large clusters of high-intensity pixels. The logarithmic transformation again helps to enhance the contrast and detail within these high-intensity regions, making this feature useful for identifying dense or bright areas within the image that may correlate with certain pathological features [96,97]. Together, these features allow for a detailed analysis of the image's texture, capturing both low- and high-intensity patterns that can be crucial for distinguishing between different tissue types or identifying specific pathological changes.

The three radiomic features—'Log Sigma 3 mm 3D GLRLM Low Gray Level Run Emphasis', 'Log Sigma 2 mm 3D GLRLM Short Run Low Gray Level Emphasis', and 'Log Sigma 3 mm 3D GLRLM Short Run Low Gray Level Emphasis'—are texture measures derived from the Gray-Level Run Length Matrix (GLRLM) combined with specific logarithmic filters applied to 3D images [25]. 'Log Sigma 3 mm 3D GLRLM Low Gray Level Run Emphasis' focuses on the emphasis of runs of low gray-level values, highlighting regions with low-intensity pixels that are clustered together [98]. The 'Log Sigma 3 mm 3D' filter applied to this feature enhances finer details within the image at a specific spatial scale, making it useful for identifying subtle low-intensity structures within the volume. 'Log Sigma 2 mm 3D GLRLM Short Run Low Gray Level Emphasis' measures the emphasis on shorter runs of low-intensity pixels, which indicates a texture where these pixels appear in smaller, more isolated clusters. The 'Log Sigma 2 mm 3D' filter is used here to capture finer, more localised texture patterns, emphasising the presence of smaller-scale low-intensity areas in the image. 'Log Sigma 3 mm 3D GLRLM Short Run Low Gray Level Emphasis' similar to the second feature, also emphasises short runs of low gray-level pixels but with a 'Log Sigma 3 mm 3D filter'. This filter size captures slightly larger texture patterns compared to the 2 mm filter, allowing the feature to identify small but slightly broader low-intensity areas, which could be indicative of certain pathological changes or tissue characteristics. Together, these features provide a nuanced analysis of the texture in medical images, particularly focusing on low-intensity regions, which can be critical for detecting and characterising specific tissue properties or abnormalities [99,100].

The three radiomic features—'Log Sigma 2 mm 3D GLSZM Small Area Low Gray Level Emphasis', 'Wavelet LHL GLSZM Small Area Low Gray Level Emphasis', and 'Wavelet LLH GLSZM Small Area Low Gray Level Emphasis'—are derived from the Gray-Level Size Zone Matrix (GLSZM), a texture analysis method that quantifies the size of homogeneous zones of gray levels in an image, combined with specific filters that enhance different aspects of the image texture [25,98]. 'Log Sigma 2 mm 3D GLSZM Small Area Low Gray Level Emphasis' emphasises small areas within the image that consist of low gray-level zones, highlighting regions where small clusters of low-intensity pixels are prevalent. The 'Log Sigma 2 mm 3D' filter enhances the detection of fine texture details at a specific spatial scale, making this feature useful for identifying subtle patterns of low-intensity areas in the image. 'Wavelet LHL GLSZM Small Area Low Gray Level Emphasis' [25] is a feature in which the 'Wavelet LHL' filter is applied capturing the horizontal high-frequency details along with low-pass filtering in the other directions. This combination focuses on small, low-intensity zones in the image, particularly those with fine horizontal structures, allowing for detailed texture analysis in specific directions. 'Wavelet LLH GLSZM Small Area Low Gray Level Emphasis' is similar to the second feature, it applies the 'Wavelet LLH' filter, which emphasises high-frequency details in the vertical direction while applying low-pass filtering horizontally. This feature targets small areas of low-intensity zones, especially those aligned with vertical structures, providing a focused analysis of these specific patterns within the image. These features collectively contribute to a detailed texture analysis by focusing on small, low-intensity areas within the image, enhanced by various filters that capture specific directional details. This facilitates the recognition of fine texture details that could be pivotal in distinguishing various tissue types or pinpointing specific pathological changes in medical imaging [101].

It's worth noting that the radiomic features extracted from the 14 patients did not achieved statistical significance. However, a previous study by Alhussaini et al. [22], involving a larger cohort of 78 patients found that at least four features either attained or approached statistical significance. This

finding highlights the vital role that sample size plays in enhancing the statistical power of analyses, demonstrating how a larger sample can reveal significant trends that smaller samples may not capture.

In conclusion, our research identified significant correlations between specific radiomics features and genomic markers, highlighting the potential of radiogenomics in non-invasive tumour characterisation. Notably, 'Log Sigma 3 mm 3D Firstorder Skewness' showed strong correlations with ChXp21.2 (DMD) (-0.73), ChXp11.23 (DYNLT3) (-0.73), and Ch2q24 (ERBB4) (-0.65). Additionally, 'Logarithm GLDM Large Dependence High Gray Level Emphasis' was linked with Ch6q14.1 (TPBG) (-0.61), while 'Wavelet LLL Firstorder Skewness' correlated with Ch6q14.1 (TPBG) (-0.61), Ch6q13 (LMBRD1) (-0.58), Ch6q14.1 (MANEA) (-0.58), and Ch6q16.3 (HACE1) (-0.58). Finally, 'Wavelet LHL GLSZM Small Area Low Gray Level Emphasis' was associated with ChXp21.2 (DMD) (-0.57) and ChXp11.23 (DYNLT3) (-0.57). These findings underscore the potential of radiomics features as surrogates for genomic data, offering promising avenues for enhancing non-invasive diagnostic and prognostic tools in clinical practice.

4.2. Limitations and Future Work

In the context of our study, one of the limitations is the relatively small number of subjects, a constraint often encountered in pilot studies. However, our patients reflect routine NHS practice. While this smaller sample size is a common characteristic of preliminary research, it does limit the generalisability of our findings. Despite this, we have meticulously detailed all the radiomics and genomics methodologies employed, ensuring that the study is transparent and reproducible.

The limited sample size underscores the need for independent replication of our findings with a larger dataset to validate the results. We calculated the required sample size using the power function:

$$N_i = \frac{p_1(1 - p_1) + p_2(1 - p_2)}{E^2} \cdot Z^2 \quad (2)$$

Where:

- N_i is the sample size of each independent sample.
- p_1 is the proportion of the first sample.
- p_2 is the proportion of the second independent sample.
- Z is the Z-score of the confidence interval.
- E is the margin of error.

Using this formula, with $p_1 = 0.286$, $p_2 = 0.714$, $E = 0.05$, and $Z = 1.96$, we estimate that a sample size of approximately 1254 subjects is necessary to achieve adequate statistical power refer to Appendix A. We recognise that with a larger and more diverse cohort, the accuracy of future research is likely to offer better and more consistent results due to a more varied population.

In future research, we recommend the use of machine learning with nested cross-validation on the identified features to enhance the robustness and accuracy of such models. Nested cross-validation [102–104] is advantageous because it helps prevent over-fitting by incorporating an additional layer of validation. This method is particularly beneficial in tuning hyperparameters while simultaneously assessing model performance, leading to more reliable and generalisable results when applied to new data.

Acknowledging these limitations, we believe that our study provides valuable insights and a strong foundation for future research. Nonetheless, independent replication with a more extensive dataset will be crucial to confirm the robustness and generalisability of our findings.

4.3. Strengths

- **Non-Invasive Analysis:** It is considered to be an alternative to biopsies; by leveraging imaging data, which can be obtained non-invasively, reducing the need for tissue biopsies. This is particularly beneficial for patients with tumours in hard-to-reach locations or those who cannot undergo invasive procedures.

- Comprehensive Tumour Profiling: Unlike traditional biopsies, which sample only a small portion of a tumour, radiogenomics analyses the entire tumour through imaging. This provides a more comprehensive view of tumour heterogeneity, capturing variations across different regions of the tumour.
- Molecular Insights from Imaging: Radiogenomics establishes correlations between imaging features and molecular markers, allowing for the prediction of genetic and molecular characteristics based on imaging data. This can lead to better understanding and characterisation of tumours.
- Tailored Treatment Strategies: By linking imaging features with specific genetic mutations, radiogenomics can help in personalising treatment plans. This ensures that therapies are more closely aligned with the molecular profile of the tumour, potentially improving patient outcomes.
- Potential for Early Detection and Prognosis: Radiogenomic research can identify imaging biomarkers that correlate with molecular signatures, which may be used for early detection of diseases or to predict disease outcomes such as response to treatment or risk of recurrence.
- Widespread Imaging Availability: Imaging technologies like CT, MRI, and PET scans are widely available in clinical settings, making radiogenomics more accessible and scalable compared to genetic testing, which may require specialised laboratories and significant costs.
- Cost-Effectiveness: In low-resource settings, where molecular testing may be cost-prohibitive, radiogenomics offers a more affordable alternative for tumour characterisation and risk stratification.
- Real-Time Tracking: Allows for continuous monitoring of tumour changes over time through serial imaging, enabling the assessment of treatment response and disease progression without repeated invasive procedures.
- Utilisation of Routine Clinical Data: Research can utilise existing imaging data routinely collected in clinical practice, making it possible to conduct large-scale studies without the need for new data collection efforts.
- Facilitating Research and Clinical Trials: Can aid in the discovery and validation of new biomarkers, enhancing the design and effectiveness of clinical trials. It also enables the stratification of patients based on imaging-genomic correlations, improving trial outcomes.
- Advancing Precision Medicine: Enhances the precision of medical interventions by integrating imaging and genomic data, leading to more accurate diagnoses, better-targeted therapies, and improved patient management.

4.4. Summary

Our radiogenomic study reveals that CT scans can effectively capture changes linked to intrinsic molecular characteristics in rare renal tumour subtypes. This is supported by predictive imaging-based models that showed performance comparable to validated gene signatures for the ChRCC and RO molecular subtypes. These findings suggest that imaging features could serve as accessible, non-invasive surrogates for molecular characteristics, providing a cost-effective alternative to genetic assessment. This is particularly beneficial in low-income settings where molecular testing may be financially prohibitive, thereby promoting more equitable access to personalised cancer care and enhancing clinical decision-making. However, further research is essential to validate and refine these imaging-based signatures across diverse populations and clinical contexts. Future studies should focus on integrating standardised imaging protocols and advanced machine learning techniques to improve the accuracy of these non-invasive biomarkers. Validation by independent researchers is crucial to confirm the robustness and clinical applicability of these findings.

5. Conclusions

In conclusion, this pilot study offers important insights into the role of radiogenomics in distinguishing between RO and ChRCC. By examining the relationship between radiomic features and SNP-based microarray copy number variations, our findings suggest that imaging can serve as a viable non-invasive alternative to traditional molecular diagnostics. The observed correlations between specific imaging characteristics and genomic markers highlight the potential of radiogenomics to improve

diagnostic precision, especially in contexts where molecular testing is less accessible. These results provide a foundation for future research aimed at validating and enhancing these imaging-based biomarkers, ultimately paving the way for the integration of radiogenomics into clinical practice to better manage renal tumours and improve patient care.

Author Contributions: Conceptualisation, A.J.A., G.N. and J.D.S.; Data curation, A.J.A., G.N. and J.D.S.; Formal analysis, A.J.A.; Investigation, A.J.A.; Methodology, A.J.A.; Project administration, A.J.A. and J.D.S.; Resources, A.J.A. and J.D.S.; Software, A.J.A.; Supervision, Veluchamy, A., Kernohan N., G.N., Palmer, Colin N.A. and J.D.S.; Validation, A.J.A.; Visualisation, A.J.A.; Writing—original draft, A.J.A.; Writing—review and editing, A.J.A., Veluchamy, A. and J.D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study received approval from the East of Scotland Research Ethical Service. Access to patients' medical healthcare data was granted under Caldicott Approval Number IGTAL9519 on August 25, 2021. Additionally, the Tissue Bank Committee [19] approved the application number TR000611 for this study on March 29, 2022.

Informed Consent Statement: For the prospective protocol-based study; informed consent form was obtained from all participants.

Data Availability Statement: The data provided are available on request from the corresponding author. The codes used to reproduce the results can be found on GitHub, upon request, at the following link: https://github.com/abeer2005/Radiogenomics_ChRCC_RO (accessed on: 20 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Acknowledgments: We extend our gratitude to Drs. Sharon King, Sally Chalmers, Gemma Skinner, and Susan Bray for arranging patient histopathology for DNA extraction and for the assistance in the technical work for tissue preparations, as well as to Drs. Norman Pratt and Robert Gordon Hislop for their assistance in securing tissue samples. Special thanks to pathologist Dr. Neil Kernohan for his critical role in histopathology annotation. We appreciate the technical support for DNA extraction provided by Dr. Gwen Kennedy, Ms. Cheryl Wood, Mrs. Karen Wilson, and Dr. Abi Veluchamy. We would like to thank Mrs. Janette Bownass and Mr. Mike Kelly from the Radiology Department for their support in collecting the CT scan data. We also acknowledge Mr. Adel Jawli (A.J.) for serving as the second observer for tumour segmentation. Finally, we wish to acknowledge the support of the Kuwait Foundation for the Advancement of Sciences (KFAS), as well as the Division of Imaging Science and Technology and the Division of Population Health and Genomics at Ninewells Hospital and Medical School, University of Dundee, for their invaluable help in this project.

Appendix A

Appendix A.1

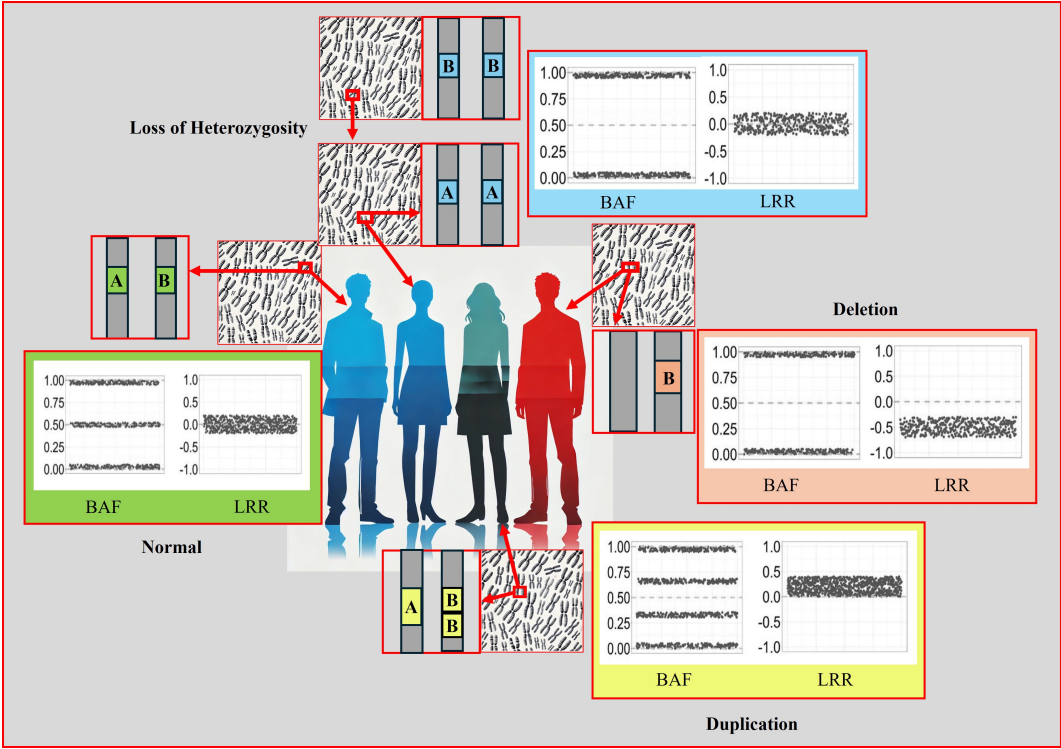


Figure A1. Illustration of different CNV types showing the corresponding changes in B-allele frequency (BAF) and log R ratio (LRR) values. In a normal diploid sample, BAF, which measures the intensity ratio (proportion) of the B allele compared to the total alleles at a specific genetic locus, typically shows values of 0.0, 0.5, and 1.0, representing the AA, AB, and BB genotypes respectively. In the case of a deletion of one allele, the BAF will shift towards 0.0 or 1.0, depending on whether the remaining allele is A or B, reflecting the presence of only one allele at that locus. Duplications lead to BAF values between the typical 0.0, 0.5, and 1.0, depending on the proportion of A and B alleles in the extra copy, often clustering around 0.33 for AAB and 0.66 for ABB. For regions without CNV, the Log R Ratio (LRR), which measures the total signal intensity compared to a reference genome to indicate copy number changes, should hover around 0, reflecting no deviation from the expected copy number. In a deletion, the LRR drops below 0, indicating a reduction in the overall copy number, while in a duplication, the LRR increases above 0, reflecting an extra copy of the affected region.

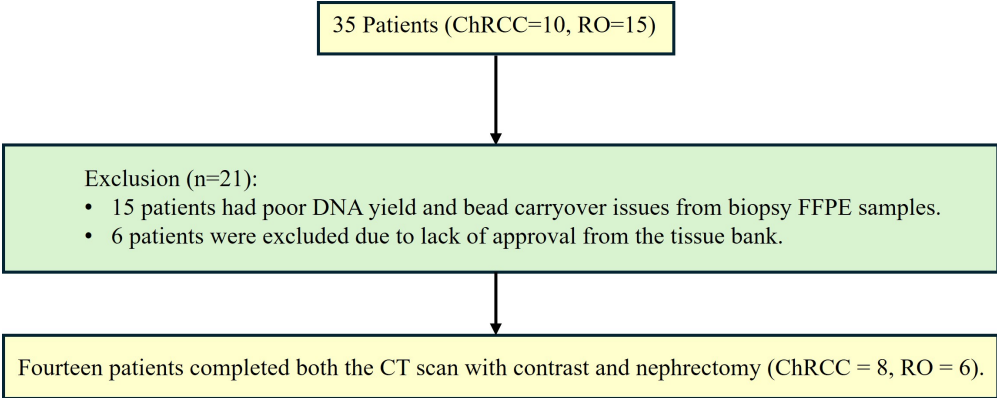


Figure A2. Exclusion and inclusion criteria of the study.

Table A1. Clinical report information of the 35 patients.

Patient	CT Scan and Histopathology Report
00294*	CT: Renal neoplastic. Histology: RN shows a tumour composed of cells with well-defined cell borders, abundant granular eosinophilic cytoplasm, and central nuclei with surrounding perinuclear halo. Occasional resinoid and binucleate forms are noted. Immunohistochemistry shows diffuse positive staining within the tumour cells for cytokeratin 7 and focal positivity for CD117. Immunohistochemistry for racemase is negative. The morphological and immunophenotypical features are consistent with chromophobe carcinoma.
00295*	CT: Well-defined enhancing lesion 35 HU non-contrast, 56 HU post-contrast. Histology: PN; chromophobe RCC. Invasion of prerenal fat.
00296*	CT: Small renal mass lesion. Histology: PN; macroscopically sections show a partial resection of kidney and perirenal fat with a well-circumscribed tumour. The tumour exhibits solid growth and nests of polygonal cells with distinct cell borders. Nuclei are pleomorphic, irregular, and wrinkled. Some cells have eosinophilic granular cytoplasm and other cells have clear cytoplasm. The appearance is suggestive of a chromophobe renal cell carcinoma. While the tumour is present very close to the capsule, it does not extend beyond the capsule. There is no evidence of vascular invasion, within the specimen. The appearance is suggestive of chromophobe RCC.
00297*	CT: Enhancing mass suggestive of malignancy. Histology: PN; 3.4 cm; chromophobe histology shows a well-circumscribed tumour composed of islands of cells separated by vessels of varying calibre. The tumour cells have well-defined cell borders, a moderate volume of granular eosinophilic cytoplasm, perinuclear halos, and in most cases central slightly irregular nuclei. In areas, nuclear pleomorphism is more prominent, atypical and multinucleate forms are present. The features are those of a chromophobe carcinoma of the kidney.
00300*	CT: Enhancing heterogeneous mass, partly exophytic unenhanced 36 HU, post-contrast 71 HU. Highly suspicious of malignancy. Histology: PN; sections show a cellular tumour composed of cells with prominent cell borders, and large and occasionally crinkled nuclei with perinuclear halos. The features are typical chromophobe RCC. Margin clear.
00301*	CT: Renal neoplastic. Histology: PN; microscopically show a lesion composed of nested tumour cells with abundant associated eosinophilic cytoplasm. The nuclei show moderate atypia with often prominent nucleoli and irregular nuclear outlines. Some cells show perinuclear clearing. In areas, the cytoplasm shows peripheral clearing giving a cell wall-type appearance. Much of the tumour is cystically dilated with numerous blood-filled pools and areas of haemosiderin deposition. Morphologically the appearance is most in keeping with those of a renal chromophobe type renal cell carcinoma. Unfortunately, confirmatory immunohistochemistry has not been helpful in this case. Excision margins are clear.
00302*	CT: Enhancing lesion. Histology: PN; microscopically histology shows a well circumscribed tumour composed of cells with well-defined cell borders, pale eosinophilic granular cytoplasm and central nuclei with prominent nucleoli and occasional raisinoid forms. These cells are arranged predominantly in solid nests with occasional more cystic areas. Immunohistochemistrey has been performed. This shows strong diffuse positive staining for Cytokeratin 7. There is also positivity for CD10. A Hale’s colloidal iron stain is positive. Whilst CD10 positivity is less common (around 26% according to some studies), the presence of Cytokeratin 7 positivity and Hale’s colloidal iron positivity, together with the morphological features, are consistent with those of a chromophobe carcinoma. The lesion adequately excised.

Table A1. Cont.

Patient	CT Scan and Histopathology Report
00303*	CT: Exophytic left renal cortical lesion concerning for RCC. Histology: PN; microscopically sections show a tumour composed of closely packed cells with clear or mildly eosinophilic cytoplasm, with other cells showing cytoplasmic clearing. Many of the nuclei are irregular and crinkled in shape and some have perinuclear halos. Immunohistochemistry: The tumour cells show positivity for cytokeratin 7 over most of the areas of the tumour in the section. Widespread membrane CD117 staining is also present. Vimentin staining is negative. The appearance is in keeping with a chromophobe carcinoma. The tumour appears well clear of the sinus excision margin. However, in areas at the outer aspect of the tumour where it bulges the capsule; there is evidence of spread beyond the capsule (pT3a).
00309*	CT: Small renal mass lesion. Histology: PN; a benign oncocytoma characterised by a well-circumscribed tumour composed of nests and tubules of cells with granular eosinophilic cytoplasm and central smooth nuclei. In the centre of the lesion, the tubules and nests are set within a paucicellular edematous stroma, corresponding to the scar seen macroscopically. This benign lesion has clear margins.
00311*	CT: Enhancing solid left renal mass. Histology: RN; microscopically histology shows a high relatively well-circumscribed tumour composed of variably sized nests of uniform cells with eosinophilic granular cytoplasm and central nuclei with inconspicuous nucleoli. A central oedematous area with infiltration of small nests into this is noted. The morphological features are highly characteristic of a benign oncocytoma. This is confirmed on immunohistochemistry with patchy positivity for Cytokeratin 7 and occasional cells and strong diffuse positivity for CD117 and PAX8. Notably, there is infiltration into the renal sinus fat and an area of vascular invasion is noted. However, this is a recognised phenomenon in benign oncocytomas and in large series following patient outcomes, did not affect their benign behaviour. Left kidney benign oncocytoma with infiltration of the renal sinus and renal vein. Margin clear.

*Represents the patients included for Radiogenomics study.

Table A2. Continuation of clinical report information for the 35 patients.

Patient	CT Scan and Histopathology Report
00313*	CT: RCC. Highly suspicious of malignancy. Histology: PN; sections show a cellular tumour composed of large nodules of bland cells with abundant eosinophilic cytoplasm and regular nuclei. In areas, smaller islands of similar cells are present with surrounding oedematous stroma. At the periphery of the larger nodules, there are groups of slightly different appearing cells with more hyperchromatic nuclei and less eosinophilic cytoplasm. Central myxoid degeneration/scarring is present. Immunohistochemistry shows that the more eosinophilic tumour cells are negative for vimentin and slightly positive for CD10. There is focal positivity for CK7 but diffuse positivity for CD117. The smaller more hyperchromatic tumour nuclei are positive for vimentin and cytokeratin 7. Despite some unusual morphological and immunohistochemical features, the overall appearance in keeping with the gross appearance of the tumour is best regarded as those of a benign oncocytoma. The tumour appears clear of the sinus excision margin by 2-3 mm.

Table A2. Cont.

Patient	CT Scan and Histopathology Report
00315*	CT: Small renal mass of 28 HU pre-contrast and 73 HU post-contrast. Suspicious of malignancy. Histology: Biopsy; two cores up to 8 mm and fragments. All taken. Microscopically; sections show two cores of renal parenchyma with one of them containing nested groups of closely packed round cells. These cells have abundant intensely eosinophilic cytoplasm, uniform small round central nuclei, mild pleomorphism, evenly distributed chromatin, and smooth nuclear membranes. There is no necrosis or mitotic activity. For immunohistochemistry these cells show strong positive staining with CD117 and patchy positive staining with CK7. The other core shows benign renal parenchyma. The features are in keeping with a benign oncocytoma. PN; macroscopically section displays renal parenchyma predominantly replaced by a neoplasm comprising nests and trabeculae of cells demonstrating intensely eosinophilic and granular cytoplasm, round nuclei, and central nucleoli. The lesion appears vascular and shows no evidence of necrosis. On immunohistochemistry these cells stain diffusely positive with CD117, show scattered positivity with CK7, and stain negative with EMA, Vimentin, and CD10. This shows negative to scattered weak positivity with CK20. Overall, the features are those of a benign oncocytoma which appears 0.4 mm away from the resection margin.
00317*	CT: Exophytic enhancing solid mass lesion appearance may represent RCC. Histology: PN; oncocytoma. Histology shows a relatively well-circumscribed tumour with a central oedematous area in which small nests and pseudo cystic structures of cells are present which have abundant granular eosinophilic cytoplasm and central nuclei with inconspicuous nucleoli. Immunohistochemistry is positive within the tumour cells for CD117 and PAX8 with only one or two cells staining focally or cytokeratin 7. The features are consistent with a benign oncocytoma. Notably, there is an area of infiltration into the perinephric fat attached to the main specimen; however, this is a recognised phenomenon in benign oncocytoma and does not affect its benign behaviour. The margin is clear of the tumour.
00327*	CT: RCC. Histology: PN; sections show this neoplasm is disrupted but comprises eosinophilic cells with smooth nuclei with no perinuclear halo's. These neoplastic cells are evenly spaced and immunohistochemistry demonstrates patchy positivity for CK7 and negative staining of CD10. Taking these morphological and IHC phenotypic features this neoplasm is consistent with an oncocytoma. No evidence of a malignant tumour.
00328	CT: Enhanced lesion. Histology: PN; sections show an encapsulated but well-delineated tumour composed of oncocytic epithelial cells which have abundant finely granular eosinophilic cytoplasm with large round nuclei, some of which have prominent nucleoli. Occasional perinuclear halo is identified along with binucleate cells; however, these are sparse and not seen across the whole tumour. The oncocytic cells also lack well-defined cell borders. Very focally there is marked nuclear pleomorphism, which is associated with regions that are regarded as being degenerative. There is extensive haemorrhage throughout the tumour and in some areas, small nests and single oncocytic cells are seen set in a background oedematous stroma. There is a marked haemorrhage within the tumour and around the capsule. Immunohistochemistry shows a diffuse expression of CD117 within the tumour with only scattered occasional cells, showing expression of cytokeratin 7. These features are most in keeping with an oncocytoma. There is no evidence of extension into the adjacent perirenal fat and the tumour is clear of the surgical resection margin by 3.8mm. However, in several areas, the capsule has been disrupted and the tumour lies on the surface therefore complete excision cannot be guaranteed.
00316	CT: Mass with stellate-like central hypodensity which may represent necrosis or scarring. The appearance is suggestive of but not pathognomonic of an oncocytoma. Histology: PN; microscopically, sections show a cellular tumour composed of cells arranged in solid sheets, nodules, or microcysts. The tumour cells have abundant eosinophilic cytoplasm and the nuclei contain prominent central nucleoli. There is no significant mitotic activity or necrosis. In some areas, the tumour also has an oedematous and focally haemorrhagic stroma. The tumour is well-circumscribed and is well-clear of the inked sinus excision margin. Immunohistochemistry shows that the tumour cells are diffusely positive for pan-cytokeratin and CD117 but only very focally positive for CK7. Staining for CD10, Vimentin, and renal cell carcinoma antigen is negative. The morphological and immunohistochemical features are entirely in keeping with renal oncocytoma.

*Represents the patients included for Radiogenomics study.

Table A3. Continuation of clinical report information for the 35 patients.

Patient	CT Scan and Histopathology Report
00321	CT: Renal neoplasm, likely RCC. Suspicious of malignancy. Histology: PN; microscopically, sections show a tumour composed of nodules and strands of tumour cells with abundant eosinophilic cytoplasm and relatively regular nuclei. The nodules are separated by an oedematous and occasionally haemorrhagic stroma with some haemosiderin deposition. The features are those of a benign oncocytoma. The tumour appears confined to the kidney and is well clear of renal sinus excision margin by approximately 5 mm.
00322	CT:Small solid lesion in keeping with small renal cell carcinoma. Histology: PN; microscopically, sections show a well circumscribed tumour partly composed of islands of regular cells with abundant eosinophilic cytoplasm in an oedematous stroma. The rest of the tumour appears cystic the cysts being lined by similar tumour cells. The appearance is those of an oncocytoma. The tumour appears clear of the sinus excision by approximately 2 mm.
00308	CT: An exophytic hyper-enhancing solid lesion, RCC cannot be excluded, further, follow-up is advised. Histology: RN; microscopically, histology shows a well circumscribed tumour with a central area of hyalinised scarring. The tumour cells are arranged in tubular structures and small nests formed of cells with central nuclei and abundant granular eosinophilic cytoplasm. Immunohistochemistry has been performed. This shows patchy focal positivity for cytokeratin 7, focal positivity for CD10 and positive staining for Vimentin predominantly concentrated in the cells around the central scar area. Immunohistochemistry for CD117 is positive. Hale’s colloidal iron shows no evidence of positivity within tumour cells. On close examination, one or two small areas are identified in which there is a slightly greater degree of nuclear abnormality with slightly irregular nuclear contours. Different diagnosis considered are chromophobe carcinoma and oncocytoma, however, on balance the central scar, morphological features of the vast majority of the tumour and the distinctive, pattern of immunohistochemistry with patchy cytokeratin 7 positivity and characteristic Vimentin positivity distribution, the features are regarded on balance as representing those of a benign oncocytoma.
00324	CT: Small renal mass demonstrate enhancement and are highly suspicious of small renal malignancies. Histology: PN; the lesion consists of packeted nests of an oncocytic neoplasm. The tumour cells have abundant oncocytic cytoplasm and central very smooth contoured round nuclei with minimal atypia. The features in both cases appear to represent those of benign oncocytoma. There is no evidence of malignancy.
00298	CT: Renal lesion Ca/Oncocytoma. Histology: Biopsy; multiple cores up to 12 mm plus fragments. Micro report: histology shows fragmented cores of tissue lesions composed of polygonal cells with abundant eosinophilic slightly granular cytoplasm, with a trabecular architecture. The nuclei are round and central, displaying little variation in size. No prominent cell membranes are identified. Immunohistochemistry shows strong expression of CK7 and EMA by the tumour cells. CD10, vimentin, and CD117 are not expressed. Hales colloidal iron staining shows a very weak focal suggestion of cytoplasmic staining but is considered largely unhelpful in further typing. In conclusion: the appearances are those of an eosinophilic cell renal tumour, the differential diagnosis of which includes the eosinophilic variant of clear cell renal cell carcinoma, chromophobe, and oncocytoma. While the morphological features are not entirely typical, the immunohistochemistry profile is most in keeping with a chromophobe neoplasm. However, a definitive diagnosis cannot be achieved on this small sample, and excision of the lesion or treatment by other means is advised. Discussion of this case at the urology MDT meeting is recommended. The patient under AS and for RFA.

Table A3. Cont.

Patient	CT Scan and Histopathology Report
00299	CT:Solid renal mass most likely to represent carcinoma. Histology: Biopsy; two cores up to 7 mm. Microscopically the specimen consists of cores of a tumour arranged in a trabecular manner, the cells of which contain abundant eosinophilic cytoplasm. The nuclei are relatively regular with no definite perinuclear halo. The impression on H&E staining was that of an oncocytoma but the immunohistochemical profile contradicts this. The tumour is negative for CD10 and vimentin but shows strong diffuse positivity for cytokeratin 7. This is much in keeping with the eosinophilic variant of chromophobe carcinoma. Patient underwent RFA.
00304	CT: Exophytic enhancing small renal mass. Histology: Biopsy; two cores up to 8 mm. All taken. Microscopically, sections show fragments of renal parenchyma and solid tumour composed of clusters and trabeculae of eosinophilic epithelial cells with minimal nuclear pleomorphism and no necrosis. Immunohistochemistry tumour cells show diffuse strong positivity for CD117 and are negative for CK7. Conclusion: taken together the morphological appearance and immunohistochemistry are in keeping with eosinophilic cell neoplasm favouring oncocytoma. Clinical correlation and MDT discussion are advised. Patient under AS.

Table A4. Continuation of clinical report information for the 35 patients.

Patient	CT Scan and Histopathology Report
00305	CT: It has a mean HU of 20 on non-contrast CT, 42 NP and 50 excretory. The heterogeneous left interpolar lesion demonstrates enhancement and although morphologically may represent an oncocytoma, it is suspicious of an RCC. Histology: Biopsy; Three Cores. Microscopically; sections show needle core biopsies of renal parenchyma which are largely replaced by a solid epithelial neoplasm arranged in solid nests and tubules. Tumour cells are remarkably monomorphic displaying abundant eosinophilic slightly granular cytoplasm with round regular nuclei and inconspicuous nucleoli. Nuclear pleomorphism, mitotic activity, and necrosis are not identified. Background perirenal soft tissue is also seen. Immunohistochemistry; tumour cells mildly express a little CD117 but appear almost negative for Cytokeratin 7 except for a few scattered cells. Vimentin and CD10 appear negative. Conclusion: taken together the morphological appearance and immunohistochemical profile favour an oncocytoma. Further clinical correlation and MDT discussion are required. Patient under AS and for RFA.
00306	CT: Interpolar renal mass with a central low attenuation, though this could be RCC, oncocytoma cannot be excluded. Histology: Biopsy; two cores up to 25 mm. Microscopically, the cores include some normal kidney but are mainly of a tumour made up of tubular and solid arrangements of round strikingly eosinophilic epithelial cells. There is no significant nuclear pleomorphism. Immunohistochemistry; the cells show patchy expression of CD10 and very focal expression of cytokeratin 7. CD117 is expressed in a membrane fashion. The appearance is entirely in keeping with oncocytoma. There is no histological evidence of malignancy. The patient under AS.
00310	CT: Heterogeneous enhancing exophytic mass, most likely to be a renal cell carcinoma. Histology: Biopsy; small fragments, Sections show a tumour consisting of nests of relatively regular cells with small nuclei and abundant eosinophilic cytoplasm. These nests are separated by a vascular oedematous stroma. Immunohistochemistry shows very slight patchy staining for CK7 but negative staining for CD10 and vimentin. The appearance is in keeping with benign oncocytoma.

Table A4. Cont.

Patient	CT Scan and Histopathology Report
00312	CT:Suspicious enhancing rounded solid (average HU 90). May represent RCC. Histology: Biopsy; microscopically histology shows core biopsies of a tumour with nests and trabecular arrangements of cells that have abundant eosinophilic granular cytoplasm and central round nuclei with occasional nucleoli. Immunohistochemistry has been performed. This is negative for CD10 and shows only focal positivity for cytokeratin 7. The features are of an oncocyctic/eosinophilic cell neoplasm, the morphology, and immunophenotype favouring origin from an oncocytoma. Patient under AS.
00318	CT: Exophytic enhancing small renal mass. Histology: Biopsy; multiple small cores, the largest 12 mm, microscopically the specimen consists of cores of a tumour composed of regular cells with small round nuclei and abundant eosinophilic granular cytoplasm. No preuclear haloes or clear cell change is seen. Immunohistochemistry shows that the tumour cells are virtually negative for cytokeratin 7 and negative for vimentin and RCC markers. However, staining for KIT is strongly positive. The morphological and immunohistochemical features are in keeping with benign oncocytoma. Patient under AS.
00319	CT: Enhancing exophytic mass. The heterogeneous left interpolar lesion demonstrates enhancement and although morphologically may represent an oncocytoma, it is suspicious of an RCC. Histology: Biopsy; three cores of tissue. Multiple fragments up to 5 mm. Microscopically, most of the specimen consists of connective tissue, but a small focus of neoplastic cells is present. These cells have abundant eosinophilic cytoplasm and small, regular nuclei, with no obvious prominent nucleoli. Immunohistochemical staining shows that the tumour cells show slight positivity for CD10 and very occasional cells are positive for CK7. The tumour cells are negative for vimentin, racemase and HMB45. The features in this biopsy are of a low grade renal epithelial neoplasm. It is often difficult to definitely subtype eosinophilic renal tumours in a small biopsy such as this. However, the morphological and immunohistochemical features are suggestive of an oncocytoma. Patient underwent RFA.
00320	CT: Renal mass. Histology: Biopsy; microscopically the sections show strands of tissue that have been derived from a tumour comprising packets, cords, and trabecula of cells set in the relatively abundant loose and pale connective tissue matrix. The tumour cells have bland cytological features but with some variation in nuclear size. Perinuclear halos are not a prominent feature; however, nucleoli can be identified in several of the nuclei. The cytoplasm shows strong granular eosinophilia. Immunohistochemistry: The sections stained for Cytokeratin 7 demonstrate occasional positive cells within the tumour. The majority of the tumour cells are negative. The tumour cells do however show positive staining for CD117 but there is no significant staining for CD10. The principal differential diagnosis, in this case, is between an oncocytoma and a chromophobe form of renal cell carcinoma. The morphology with the immunohistochemical pattern of staining particularly with respect to the very focal staining for cytokeratin 7 would favour an oncocytoma. The patient under AS.

Table A5. Continuation of clinical report information for the 35 patients.

Patient	CT Scan and Histopathology Report
00307	CT: Contrast-enhanced renal mass. Histology: Biopsy; microscopically, sections show nests of round cells with abundant finely granular eosinophilic cytoplasm and uniform small, round, and central nuclei with evenly dispersed chromatin. Immunohistochemistry: these cells stained positive for CD10 and CD117 and were focally positive for CK7. They stained negative with vimentin. The morphology appearances and immunohistochemical profile are in keeping with renal oncocytoma. The patient under AS.
00314	CT: Solid enhancing exophytic small renal mass. May represent RCC. Histology: Biopsy; three cores up to 30 mm. Micro Report: one of these cores is of the unremarkable renal parenchyma. The other two are samples of neoplasm made up of homogenous rounded epithelial cells with prominent eosinophilic cytoplasm. Nuclear pleomorphism is not notable and there is no obvious mitotic activity. Immunohistochemistry shows only rare cells expressing cytokeratin 7. The tumour cells show membranous expression of CD117 and are negative for vimentin. The appearance is highly suggestive of oncocytoma. Patient under AS.
00323	CT: Small renal mass, enhancing solid lesion. Histology: Biopsy; three cores up to 12mm. Micro report: these cores are partially replaced by a tumour with a very eosinophilic morphology. Immunohistochemistry shows diffuse staining for CD117 and only patchy staining for Cytokeratin 7. Taken together the morphology and immunohistochemistry profile would favour an oncocytoma. Patient under AS.
00325	CT: Solid renal mass, suspicious of RCC. Histology: Biopsy; two cores. X1 up to 10 mm plus fragments. Sections show a core biopsy of renal parenchyma bearing a focus on neoplastic tissue towards one end. The tumour cells are arranged in rosettes and solid islands. Tumour cells have abundant eosinophilic cytoplasm and small hyperchromatic nuclei. No mitotic activity or necrosis is identified. Results of special stains immunohistochemical, Tumour cells appear entirely negative for cytokeratin 7. Tumour cells moderately express KIT. Tumour cells appear negative for CD10. Tumour cells are negative for RCC. Conclusion: Taken together the morphological appearance and immunohistochemical profile are most in keeping with an oncocytoma. Further clinical correlation and discussion of this case at the urology MDT meeting are strongly recommended. Patient under AS.
00326	CT: Enhanced solid small renal mass, in the arterial and delayed phase. Histology: Biopsy; 2018; two cores of tissue up to 14 mm in length. Micro Report: The section shows cores of tissue that have been derived from a tumour that is formed of trabeculae and groups of bland-appearing cells. The cells show oncocytic features with uniform pink cytoplasm. The cells have round nuclei with minimal pleomorphism. Perinuclear halos are not apparent. The cytoplasmic boundaries are not sharply defined. A panel of immunohistochemistry demonstrates that the tumour exhibits diffuse strong positivity for broad-spectrum cytokeratin's recognised by MNF. Scattered positive tumour cells are identified in the sections stained for broad-spectrum cytokeratin's recognised by AE1/3 as well as for cytokeratin 7. The sections stained for cytokeratin 20 show a blush of staining in the tumour cells but the section stained for cytokeratin 14 is negative. The tumour cells are negative for vimentin. There is a patchy variable cytoplasmic expression of CD10. The tumour cells are substantially negative for RCC but do show variable positive staining for EMA. Overall, the appearance indicates cores of tissue derived from a tumour with prominent oncocytic features. The principal differential diagnosis rests between oncocytic features; morphology of the lesion together with the immunohistochemical profile favour oncocytomas rather than a chromophobe RCC. Biopsy; 2019; increased in size; malignant? one core up to 20 mm. Micro Report: morphologically this is a primary renal neoplasm with marked oncocytic features with cells arranged in nests and cords. There is abundant eosinophilic cytoplasm and central round nuclei with smooth contour. Atypical features are not identified and morphological features more suggestive of a chromophobe RCC are not present. Furthermore, immunohistochemistry: there is virtually negative staining for cytokeratin 7 and positive staining for CD117. The features therefore remain consistent with those of an oncocytoma. Further clinical and radiological correlation is advised. The patient under AS.

Table A6. DNA quantification and purity results for FFPE and fresh-frozen samples using NanoDrop Microvolume spectrophotometer. The 260/280 ratio provides a measure of DNA purity, while **ng/μl** indicates the DNA concentration.

Sample ID	260/280	ng/μl
T511 3A (1)	1.75	62.21
T52223 5H	1.84	215.9
T53222 2C	1.78	104.2
T54444 1C	1.83	135.5
T55447 1C	1.81	203.7
T56643 1C	1.82	516.3
T535 2Cre	1.78	104.2
T5721 1Bre	1.83	297.8
T57987 1B	1.83	297.8
T58876 2I	1.77	89.14
T59543 1E	1.78	175.2
T51032 1D	1.83	168.5
T51123 4B	1.85	517.9
T5134 2Cre	1.71	101.1
T512 3A (2)	1.58	37.82
T51308 2C	1.71	101.1
T51444 2E	1.71	66.63
T51555 1A	1.75	145.21
115161 14b	1.74	91.41
114939 3re	1.65	83.63
135177 34e	1.69	69.51
1114188 39	1.65	83.63
115199 07f	1.64	74.89
1154 14Bre	1.74	91.41

Note: This table provides the results of DNA quantification and purity assessment for various FFPE samples and fresh-frozen samples using a NanoDrop Microvolume spectrophotometer. The ****260/280 ratio**** is used as an indicator of DNA purity, with a ratio of around 1.8 typically considered indicative of pure DNA. Lower ratios may suggest the presence of protein contamination, while higher ratios might indicate RNA contamination. The ****ng/μl**** column represents the concentration of DNA in each sample, measured in nanograms per microlitre. Higher concentrations indicate larger amounts of DNA in the sample, which are essential for downstream molecular biology applications.

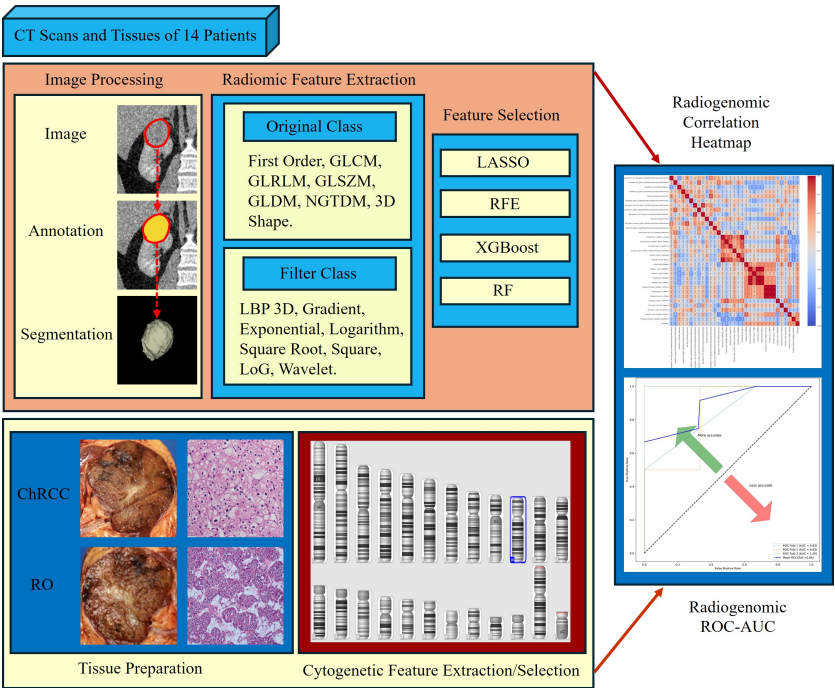


Figure A3. The diagram offers a representation of the study’s methodological process.

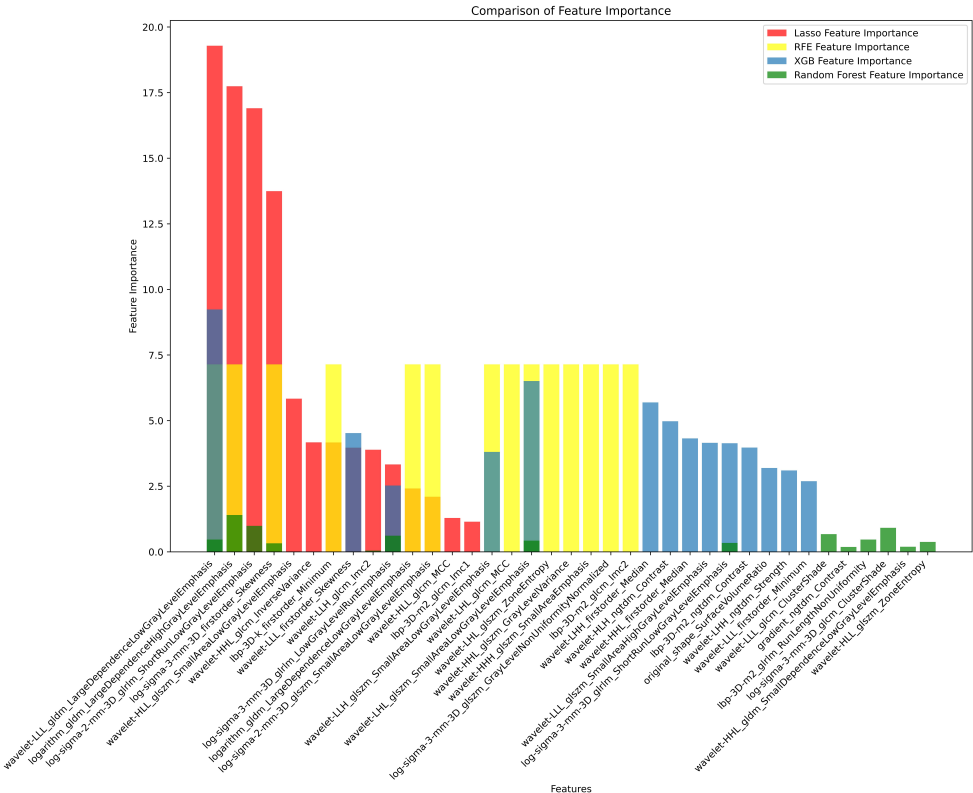


Figure A4. Comparison of feature importance between the four feature selection algorithms.

Table A7. The parameters for each of the fourteen genotypes analysed by cnvPartition are displayed. For homozygous deletions (DD), BAFs are modelled as a uniform distribution between zero and one. All other genotypes are modelled using Gaussian distributions with the specified parameters. The genotype AABB is excluded from modelling because it would represent two independent duplication events, which are rare in nature. (CN = copy number, DD = double deletion, SD = standard deviation) [15].

Genotype	CN	LRR-Mean	LRR-SD	BAF-Mean	BAF-SD
DD	0	-5	2	NA	NA
A	1	-0.45	0.18	0	0.3
B	1	-0.45	0.18	1	0.3
AA	2	0	0.18	0	0.3
AB	2	0	0.18	0.5	0.3
BB	2	0	0.18	1	0.3
AAA	3	0.3	0.18	0	0.3
AAB	3	0.3	0.18	1/3	0.3
ABB	3	0.3	0.18	2/3	0.3
BBB	3	0.3	0.18	1	0.3
AAAA	4	0.75	0.18	0	0.3
AAAB	4	0.75	0.18	0.25	0.3
ABBB	4	0.75	0.18	0.75	0.3
BBBB	4	0.75	0.18	1	0.3

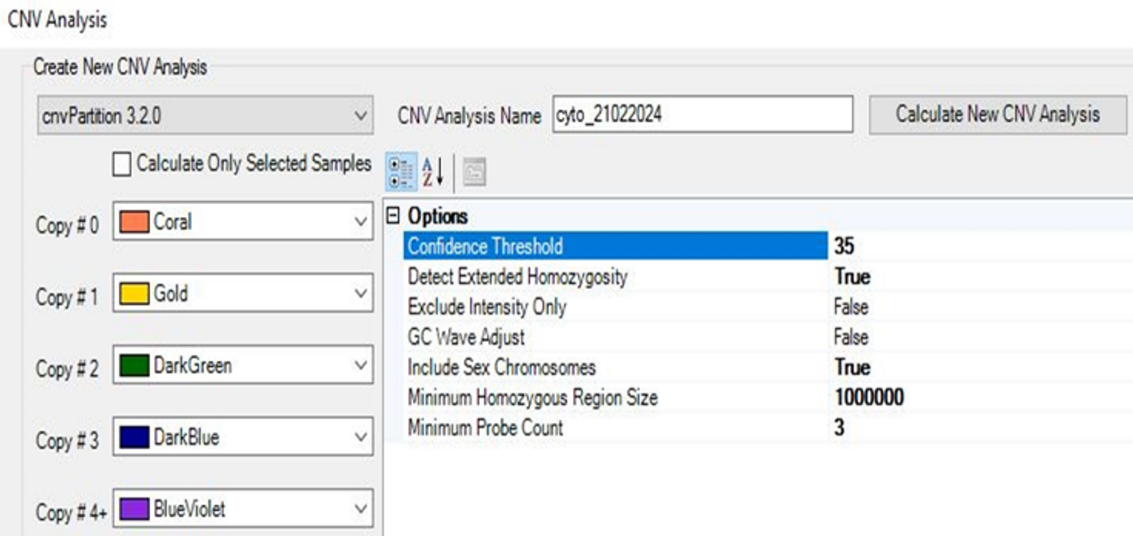


Figure A5. Visualisation of CNV analysis settings using the Illumina Genome Viewer within the Chromosome Browser. The figure represents the various standard parameter thresholds that were used for CNV analysis, including confidence threshold of 35 and minimum probe count of 3. The left panel indicates colour coding of copy numbers (CN) as determined during the analysis.

Table A8. The Illumina Genome Viewer, integrated within Genome Studio, is utilised to graphically visualise copy number regions and to display copy number analysis results in tabular format [29].

CNV-Type	CNV-Value	CNV-Confidence
Normal	2	Blank
Duplication	3 or 4	Contains Value
Deletion	1	Contains Value
Copy Neutral-LOH	2	Contains Value

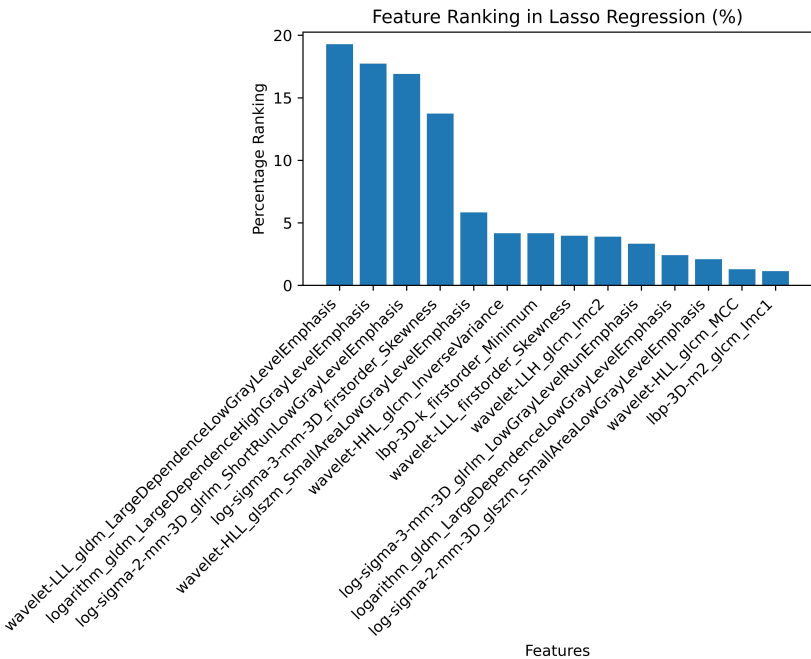


Figure A6. Representation of the feature ranking using Lasso Regression.

wavelet-LLH_glszm_SmallAreaLowGrayLevelEmphasis
wavelet-LHL_glcm_MCC
wavelet-LHL_glszm_SmallAreaLowGrayLevelEmphasis
wavelet-LHL_glszm_ZoneEntropy
wavelet-HHL_glszm_GrayLevelVariance
wavelet-HHH_glszm_SmallAreaEmphasis
wavelet-LLL_gldm_LargeDependenceLowGrayLevelEmphasis
log-sigma-2-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
log-sigma-3-mm-3D_firstorder_Skewness
log-sigma-3-mm-3D_glszm_GrayLevelNonUniformityNormalized
logarithm_gldm_LargeDependenceHighGrayLevelEmphasis
logarithm_gldm_LargeDependenceLowGrayLevelEmphasis
lbp-3D-m2_glcm_lmc2
lbp-3D-k_firstorder_Minimum

Figure A7. Representation of the best 14 features selected using RFE.

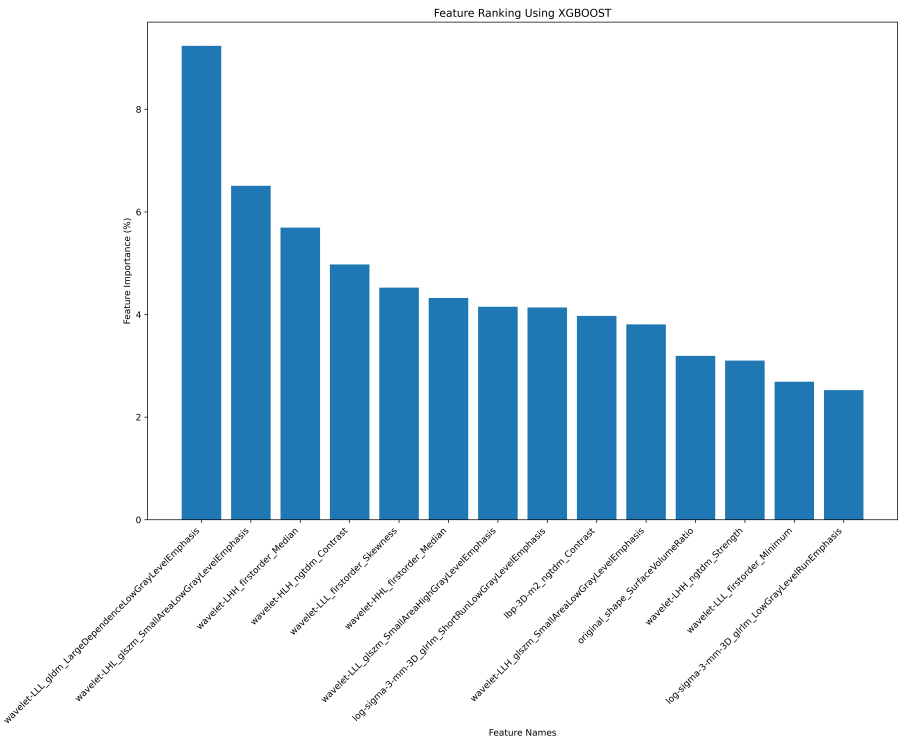


Figure A8. Representation of the feature ranking using XGBOOST.

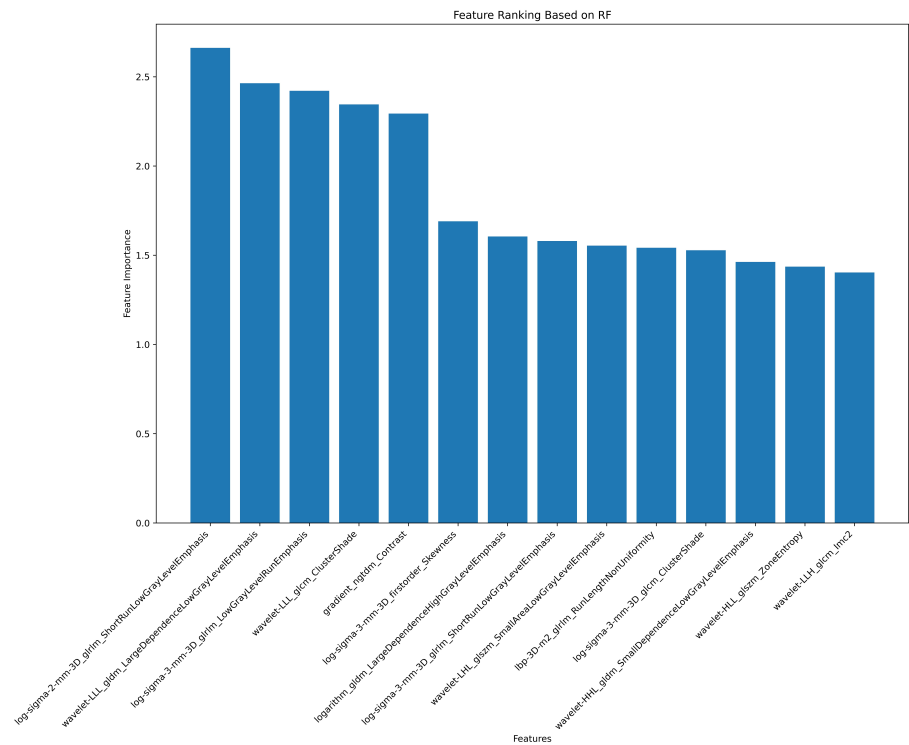


Figure A9. Representation of the feature ranking using RF.

Table A9. Correlation of radiomic features with histopathological differentiation of ChRCC and RO using various filters.

#	Filter Type	Feature Category	Radiomic Feature	Correlation (r)	p-value (t)
1	Log Sigma 3 mm 3D	First Order	Skewness	0.39	0.698
2	Wavelet LLL	First Order	Skewness	-0.37	0.527
3	LBP 3D k	First Order	Minimum	0.37	0.79
4	Wavelet LLH	GLCM	Informational Measure of Correlation '2' (IMC2)	0.25	0.988
5	Wavelet LLL	GLDM	Large Dependence Low Gray Level Emphasis (LDLGLE)	-0.38	0.85
6	Logarithm	GLDM	Large Dependence Low Gray Level Emphasis (LDLGLE)	-0.12	0.594
7	Logarithm	GLDM	Large Dependence High Gray Level Emphasis (LDHGLE)	0.34	0.07
8	Log Sigma 3 mm 3D	GLRLM	Low Gray Level Run Emphasis (LGLRE)	0.27	0.69
9	Log Sigma 2 mm 3D	GLRLM	Short Run Low Gray Level Emphasis (SRLGLE)	0.33	0.626
10	Log Sigma 3 mm 3D	GLRLM	Short Run Low Gray Level Emphasis (SRLGLE)	0.29	0.956
11	Log Sigma 2 mm 3D	GLSZM	Small Area Low Gray Level Emphasis (SALGLE)	0.19	0.079
12	Wavelet LHL	GLSZM	Small Area Low Gray Level Emphasis (SALGLE)	0.2	0.137
13	Wavelet LLH	GLSZM	Small Area Low Gray Level Emphasis (SALGLE)	-0.05	0.516

Note: In the table, Pearson’s correlation coefficient (*r*) was used to quantify the linear relationship between various radiomic features and the histopathological differentiation of ChRCC and RO. This coefficient is well-suited for measuring the strength and direction of linear associations, where values close to 1 or -1 indicate strong correlations, and values near 0 suggest little to no linear relationship. Its use in this analysis helps identify which features might be most relevant for distinguishing between ChRCC and RO, providing valuable insights for predictive models or diagnostic tools. The T-test (*t*) was employed to determine whether there are statistically significant differences in the means of radiomic features between the two groups, ChRCC and RO. This test assumes normality within each group and is effective for detecting even small differences in means.

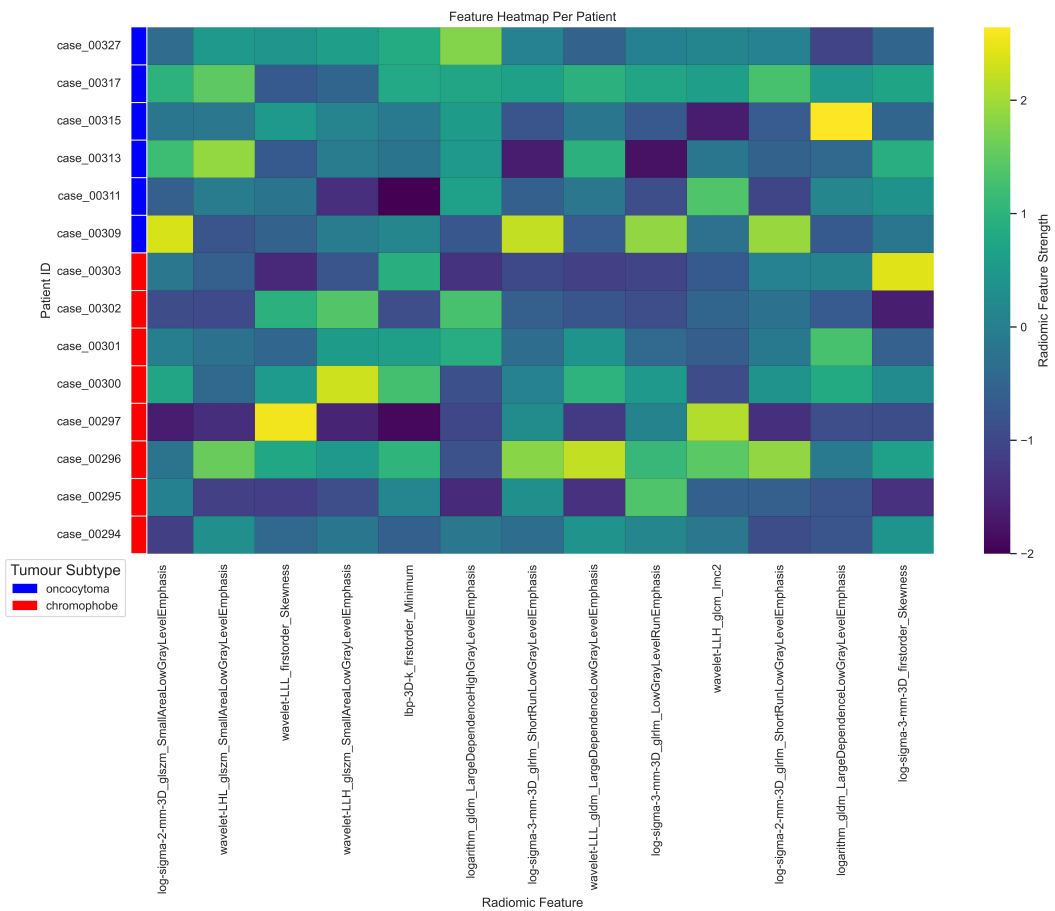


Figure A10. Representation of the heat map showing the correlation between the features and histopathology target of each patient. The colour gradient indicates the intensity of each feature, ranging from blue/purple (low) to yellow (high). The heat map highlights variations in feature expression between the two tumour subtypes, which may assist in differentiating between oncocyoma and chromophobe renal cell carcinoma based on radiomic signatures.

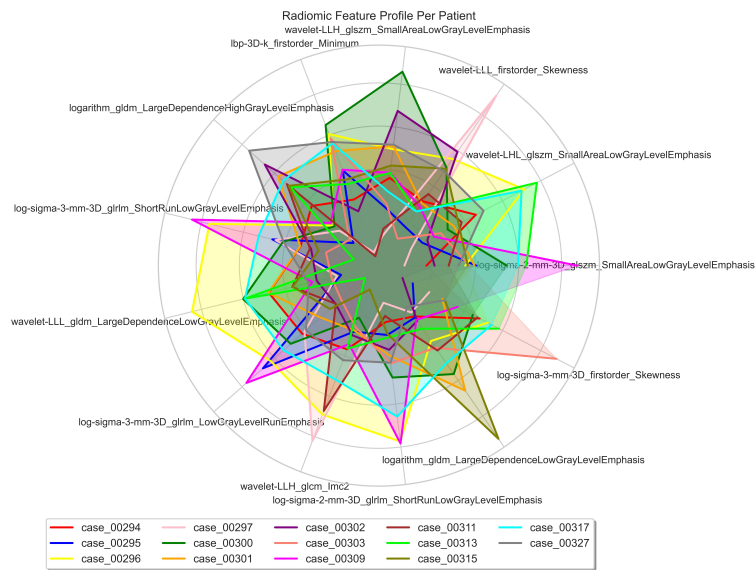


Figure A11. This radar plot illustrates the radiomic feature profiles for 14 patients, highlighting the variation in radiomic feature strength across different cases. Each line represents a unique patient, with radiomic features plotted around the perimeter. The plot enables comparison of individual feature patterns, revealing potential correlations between radiomic characteristics and specific histopathological findings. Different colors represent each patient, allowing for a clear visual comparison of feature expression across the cohort.

Table A10. 24 Genetic and statistical data with correlation values for ChrCC and RO analysis.

Cytoband	Gene	Correlation (<i>r</i>)	ChrCC%	RO%	p-value (χ^2)	p-value (<i>t</i>)
1p34.1	RNF115	0.41	12.5	50	0.347	0.14
1q21.3	CTSK	0.41	12.5	50	0.347	0.14
1q21.3	S100A1	0.6	12.5	50	0.347	0.02
1q22	MUC1, RAB25	0.6	0	50	0.109	0.2
1q25.2	ANGPTL1	0.26	25	50	0.687	0.37
1q32.3	MTF2	0.26	25	50	0.687	0.37
1q42.13	TMED5	0.26	25	50	0.687	0.37
1q21.2	MCOLN2, MCOLN3	0.26	25	50	0.687	0.37
1q32.1	LAPTM5	0.42	25	66.67	0.31	0.14
1p36.22	NBL1	0.41	12.5	50	0.347	0.14
2q24	ERBB4	0.45	37.5	0	0.301	0.11
6q13	LMBRD1	0.45	37.5	0	0.301	0.10
6q14.1	TPBG	0.35	25	0	0.58	0.21
6q14.1	MANEA	0.45	37.5	0	0.301	0.10
6q16.3	HACE1	0.45	37.5	0	0.301	0.10
10q11.23	PRKG1	0.45	37.5	0	0.301	0.10
10q22.1	CSTF2T	0.45	37.5	0	0.301	0.10
10p12.33	MRC1	0.45	37.5	0	0.301	0.10
10p12.1	STAM	0.45	37.5	0	0.301	0.10
10q22.2	PPP3CB	0.47	0	33.33	0.58	0.09
17q21.31	SLC4A1	0.32	0	16.67	0.88	0.26
Xp21.2	DMD	0.61	87.5	33.33	0.126	0.02
Xp11.23	DYNLT3	0.61	87.5	33.33	0.126	0.02
Xq28	CTAG1B, MAGEA4, MAGEA3	0.47	75	33.33	0.31	0.09

Note: In this table, the Pearson correlation coefficient (*r*) was employed to measure the strength and direction of the linear relationship between CNV values of specific genetic regions and the histopathological differentiation of ChrCC and RO. The Chi-square test (χ^2) was used to assess whether there is a statistically significant association between categorical variables—specifically, the presence or absence of certain CNVs (categorised as gain, loss, or neutral) and the histopathological classifications of ChrCC and RO—helping to determine if the distribution of these categorical CNV data differs between the two histological types. Additionally, the T-test (*t*) was applied to compare the means of CNV values between the two groups, ChrCC and RO. This test was appropriate for determining whether there were statistically significant differences in the average CNV values of specific genetic regions between these two histopathological types, under the assumption that the CNV data within each group were approximately normally distributed with similar variances.

Table A11. Representation of the radiomics and cytogenomics features that are highly correlated and mapped with each other and with the histopathology target.

Correlation	Radiogenomics Features n=34
-0.73	ChXp21.2 (DMD) and Log Sigma 3 mm 3D Firstorder Skewness
-0.73	ChXp11.23 (DYNLT3) and Log Sigma 3 mm 3D Firstorder Skewness
-0.65	Ch2q24 (ERBB4) and Log Sigma 3 mm 3D Firstorder Skewness
-0.61	Ch6q14.1 (TPBG) and Logarithm GLDM Large Dependence High Gray Level Emphasis
-0.61	Ch6q14.1 (TPBG) and Wavelet LLL Firstorder Skewness
-0.58	Ch6q13 (LMBRD1) and Wavelet LLL First Order Skewness
-0.58	Ch6q14.1 (MANEA) and Wavelet LLL First Order Skewness
-0.58	Ch6q16.3 (HACE1) and Wavelet LLL First Order Skewness
-0.57	ChXp21.2 (DMD) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
-0.57	ChXp11.23 (DYNLT3) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
-0.56	Ch6q14.1 (TPBG) and Wavelet LLL GLDM Large Dependence Low Gray Level Emphasis
0.5	Ch1q21.3 (S100A1) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
0.5	Ch1q22 (MUC1, RAB25) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
-0.5	Ch1q21.3 (S100A1) and Log Sigma 3 mm 3D GLRLM Low Gray Level Run Emphasis
-0.5	Ch1q22 (MUC1, RAB25) and Log Sigma 3 mm 3D GLRLM Low Gray Level Run Emphasis
-0.5	Ch10q11.23 (PRKG1) and Wavelet LLH GLCM Imc2
-0.5	Ch10q22.1 (CSTF2T) and Wavelet LLH GLCM Imc2
-0.5	Ch10p12.33 (MRC1) and Wavelet LLH GLCM Imc2
-0.5	Ch10p12.1 (STAM) and Wavelet LLH GLCM Imc2
0.47	Ch1q32.1 (LAPTM5) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
0.46	Ch10q22.2-q23.1 (PPP3CB) and Logarithm GLDM Large Dependence High Gray Level Emphasis
0.45	Ch17q11.1-q21.32 (SLC4A1) and Logarithm GLDM Large Dependence High Gray Level Emphasis
-0.45	Ch6q13 (LMBRD1) and Wavelet LLH GLCM Imc2
-0.45	Ch6q14.1 (MANEA) and Wavelet LLH GLCM Imc2
-0.45	Ch6q16.3 (HACE1) and Wavelet LLH GLCM Imc2
-0.45	Ch17q21.31 (SLC4A1) and Logarithm GLDM Large Dependence Low Gray Level Emphasis
-0.44	Ch2q24 (ERBB4) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
0.43	Ch1q25.2 (ANGPTL1) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
-0.42	ChXq28 (CTAG1B, MAGEA3, MAGEA4) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
-0.41	Ch6q14.1 (TPBG) and Wavelet LLH GLSZM Small Area Low Gray Level Emphasis
-0.41	Ch6q14.1 (TPBG) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis
0.41	Ch1q21.3 (S100A1) and Logarithm GLDM Large Dependence High Gray Level Emphasis
0.41	Ch1q22 (MUC1, RAB25) and Logarithm GLDM Large Dependence High Gray Level Emphasis
-0.41	Ch6q14.1 (TPBG) and Wavelet LHL GLSZM Small Area Low Gray Level Emphasis

Note: Pearson’s correlation coefficient (*r*) is used to quantify the strength and direction of the linear relationship between radiomic features and cytogenomic variations, as well as their association with histopathological outcomes. By using Pearson’s correlation, we were able to identify and measure the degree to which specific genetic alterations (represented by cytogenomic features) are associated with specific radiomic characteristics. This analysis helps to highlight potential radiogenomic biomarkers that could be relevant for differentiating between ChRCC and RO based on their histopathological classification, contributing valuable insights for both diagnosis and potential treatment strategies.

Table A12. Representation of the number and percentage of genome segments with duplication, deletion, and loss of heterozygosity for each patient. The table also shows the mean size of the affected segments in each category.

Subtype	Patient	DUP	DEL	LOH	Total	DUP (%)	DEL (%)	LOH (%)	DUP (Mb)	DEL (Mb)	LOH (Mb)
ChRCC	294	20	7	4	31	6.35	5.26	2.35	1.14	2.61	37.31
	295	66	5	2	73	20.95	3.76	1.8	14.91	0.33	2.57
	296	3	25	6	34	0.95	18.8	3.53	1.35	0.74	25.18
	297	5	6	10	21	1.59	4.51	5.88	1.72	3.14	15.26
	300	9	13	8	30	2.6	9.77	4.71	10.69	0.03	20.93
	301	44	15	2	61	13.97	11.28	1.78	12.67	1.91	1.82
	302	11	19	131	161	3.49	14.29	77.06	14.21	1.08	1.35
RO	303	59	0	0	59	18.73	0	0	19.26	0	0
	309	6	13	3	22	1.9	9.77	1.76	10.62	1.4	49.52
	311	14	2	0	16	4.4	1.5	0	20.09	6.67	0
	313	4	3	1	8	1.27	2.56	0.59	17.77	5.94	1.22
	315	5	14	3	22	1.59	10.53	1.76	0.037	1.37	50.01
	317	11	10	0	21	3.49	7.52	0	1.62	1.95	0
p-value	327	58	1	0	59	18.41	0.75	0	5.5	0.0065	0
	χ^2	0.35	0.43	0*	-	-	-	-	1.0	0.08	0.78

* Statistical significant difference is considered at 0.05 significance level. DUP: Duplication, DEL: Deletion and LOH: Loss of Heterozygosity.

Table A13. Count of chromosomes per patient, comparing ChrCC cases numbered 294 to 303 with RO cases numbered 309 to 327.

#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
294	9	1	1	0	0	1	0	0	0	2	1	0	0	3	0	0	4	0	0	0	0	1	4	4
295	8	11	6	2	1	10	4	1	4	4	1	1	3	1	0	0	6	0	0	0	2	0	5	3
296	3	1	0	1	1	1	3	2	0	2	2	1	1	1	1	2	0	0	0	1	1	0	6	4
297	2	0	3	0	0	1	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	6	5
300	0	1	0	0	0	2	0	2	1	8	0	0	2	0	1	0	1	0	0	2	1	0	6	3
301	6	4	3	0	0	2	3	0	6	7	2	2	2	1	1	0	4	3	0	1	2	0	8	4
302	25	33	28	0	1	21	1	1	3	10	3	0	8	1	1	0	7	1	0	2	2	0	8	5
303	2	2	4	3	3	2	3	6	2	0	3	2	3	1	1	5	4	4	4	2	1	1	1	0
309	6	0	0	1	2	1	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	4	4
311	3	1	0	1	0	2	0	0	0	3	0	0	1	0	0	1	0	0	0	1	0	0	1	2
313	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	3
315	5	0	2	0	0	1	0	1	0	1	1	0	0	0	1	0	0	0	1	0	1	1	2	5
317	7	0	0	0	0	1	0	2	0	1	0	0	1	0	0	1	1	2	0	0	0	1	0	4
327	11	3	1	0	4	5	0	2	1	8	2	0	3	2	2	2	5	2	2	3	0	0	1	0
χ^2	0.9	0	0	0.1	0.4	0	0	0.3	0	0.2	0	0	0	0	1	0.7	0	0.7	0.1	0.5	0	0.7	0	0.9

Bold: Statistical significance at the 0.05 level using chi-squared test (χ^2).

Table A14. Chromophobe RCC and RO CNVs classification and type identified.

	ChRCC								RO						p-value
Patient #	294	295	296	297	300	301	302	303	309	311	313	315	317	327	χ^2
	Pathogenic														
DEL	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0.72
DUP	0	7	0	0	0	3	3	8	0	1	1	0	0	4	
LOH	3	0	4	4	4	1	11	0	3	0	0	3	0	0	
	Likely Pathogenic														
DEL	1	0	6	1	0	3	0	0	2	1	1	1	2	0	0.16
DUP	1	21	0	1	2	14	0	31	2	5	1	0	0	16	
LOH	0	0	0	0	1	0	5	0	0	0	0	0	0	0	
	Benign														
DEL	0	1	1	2	0	0	2	0	0	0	0	2	2	0	-
DUP	0	0	0	0	0	0	0	0	0	0	0	1	1	0	
LOH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	Uncertain Significance														
DEL	6	4	17	3	13	0	13	0	11	0	2	11	6	1	0*
DUP	19	38	3	4	7	27	7	20	4	8	2	4	10	34	
LOH	1	2	2	6	3	0	115	0	0	0	1	0	0	0	

*Statistical significance at the 0.05 level using chi-squared test (χ^2).

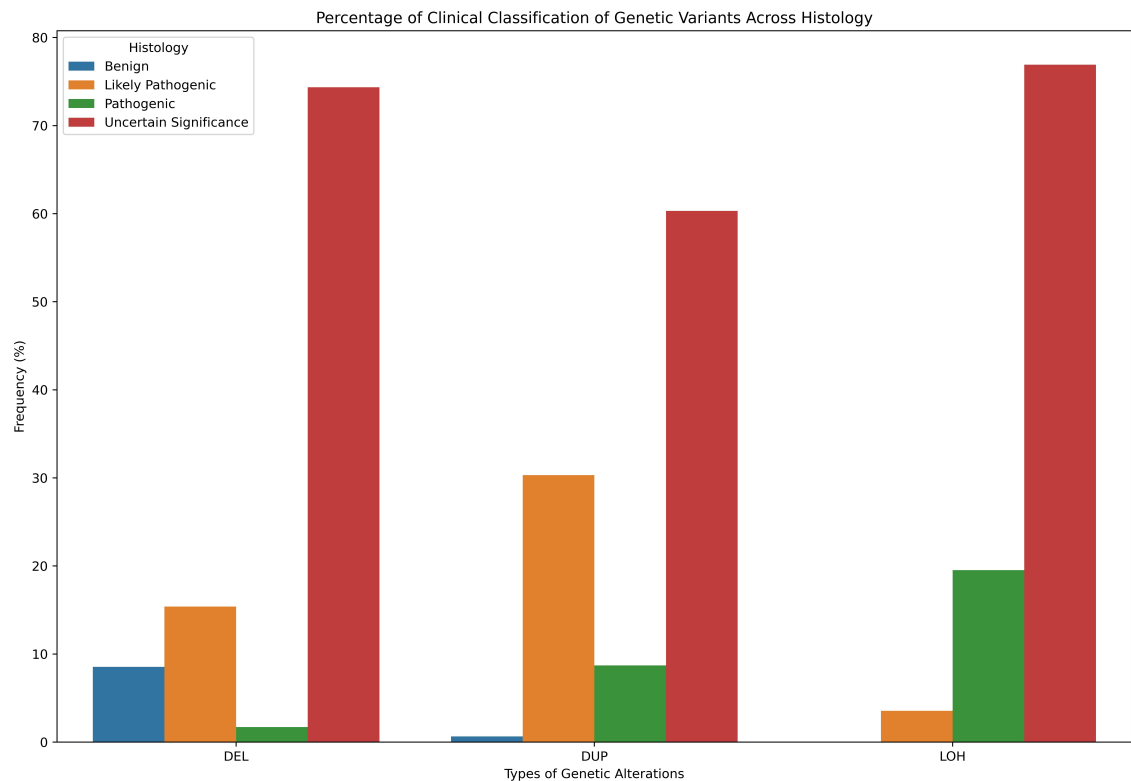


Figure A12. Distribution of clinical classifications of genetic variants across different types of genetic alterations. The bar chart represents the frequency (in percentage) of genetic variants classified as Benign, Likely Pathogenic, Pathogenic, and of Uncertain Significance for three types of genetic alterations: Deletions (DEL), Duplications (DUP), and Loss of Heterozygosity (LOH). The y-axis shows the frequency percentage, while the x-axis categorises the genetic alterations. The chart highlights that the majority of variants within DUP and LOH are classified as having Uncertain Significance, whereas DEL has a more varied distribution across all classifications.

Table A15. The diagnostic performance of the RF model using different cut-offs for Pearson’s correlation coefficient (*r*).

Correlation (<i>r</i>)	ACC	SPE	SEN	AUC	MCC	F1
>0.55	81.25	87.50	75.00	85.00	0.63	0.80
>0.50	68.75	87.50	50.00	83.00	0.40	0.62
>0.45	62.50	87.50	37.50	80.00	0.29	0.70
>0.40	75.00	87.50	62.50	78.00	0.52	0.71
>0.30	68.75	87.50	50.00	89.00	0.40	0.62

Table A16. Calculation of the power function to determine the required sample size.

p_1 (ChRCC)	p_2 (RO)	E	Z
0.286	0.714	0.05	1.96
N_i			627.576

Note: N_i is the sample size of each independent sample, p_1 is the proportion of the first sample, p_2 is the proportion of the second independent sample, Z is the Z-score of the confidence interval, and E is the margin of error. From the samples obtained in the first project, a total of 35 samples were analysed. Of these, 10 were identified as chromophobe (ChRCC) and 25 as renal oncocyoma (RO). The proportions for each subtype were calculated, yielding 29% for ChRCC and 71% for RO, which correspond to proportions of 0.29 and 0.71, respectively. The study determined the margin of error to be 0.05 and set the confidence interval at 95%. Using this confidence interval, the corresponding Z-score was calculated. Based on these parameters, the power function was computed, indicating that a sample size of 627.58 would be required for each subtype to achieve the desired statistical power.

Table A17. Representation of Z-scores for two-tailed hypothesis testing at various confidence levels.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9600	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9992
3.2	.9993	.9993	.9993	.9993	.9993	.9993	.9994	.9994	.9994	.9994
3.3	.9994	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995	.9995
3.4	.9995	.9995	.9995	.9995	.9995	.9995	.9995	.9995	.9996	.9996

Note: To determine the Z-score corresponding to a 95% confidence interval, we start by recognising that a 95% confidence level leaves 5% of the distribution in the tails (since $100\% - 95\% = 5\%$). Because the Z-distribution is symmetrical, this 5% is split evenly between the two tails, leaving 2.5% in each tail. Thus, to find the Z-score, we need to locate the value on the Z-table that corresponds to a cumulative probability of 97.5% (which is $95\% + 2.5\%$). This cumulative probability represents the area under the curve to the left of the Z-score. Looking at the Z-table; We first find 0.9750 in the body of the table, which is the closest value to our desired cumulative probability. This value corresponds to a Z-score of 1.9 in the row and 0.06 in the column, which when added together gives us a Z-score of 1.96. Therefore, the Z-score associated with a 95% confidence interval is 1.96.

Appendix A.2. Technical Lab Work

Appendix A.2.1. Samples Preparation

(FFPE Samples):

Materials:

- Microtome
- Microtome Blades
- Floating Out Water Bath
- Cold Plate/Ice Tray
- Hot Plate
- Slides

- Forceps
- Paint brush

Methods:

We precisely segment tissue using the Leica RM2235 Rotary Microtome. The microtome's configuration ensures the ideal clearance angle for sectioning, specifically angled to accommodate the microtome RM55 blades. A specimen clamp is utilised to accommodate various cassette clamp sizes. By releasing the specimen clamp lever, the orientation of the specimen clamp can be adjusted using the dials. The thickness of the section can be altered simply by turning the knob, with the selected thickness set at 10 μm . After configuring the microtome, tissue blocks are created. This process involves removing any excess wax from the tissue cassette while ensuring the knife guard and brake are securely in place for safety. Additionally, the knob is adjusted to the desired thickness.

Using the coarse driving wheel, we cautiously advance the tissue block close to the blade after aligning it with precision. To trim the block carefully, ensuring a visible whole tissue face, we typically use a new blade or the unused portion of an old blade. Once trimming is complete, the tissue blocks are placed face down on the ice tray or cold plate, ready for sectioning. To maintain consistent section thickness throughout the sectioning process, we finely adjust the adjustment knob. For sectioning, either a brand-new blade or the unused portion of a blade is utilised. The required tissue block is placed and aligned with the blade for sectioning.

We advance the cassette clamp to create continuous ribbon-shaped pieces by slowly twisting the smooth-turning wheel. A small paintbrush is used to remove the ribbon from the blade, and forceps are employed to lift the sections for further processing. These sections are delicately floated onto a clean water bath after being cut.

We emphasise the importance of avoiding over-expansion by minimising the time sections spend on the water bath to reduce tissue damage. Using a pencil, pertinent information for each section is noted on the frosted portion of the slide. The slide is then placed on a heated plate to accelerate drying and improve section adherence. After each set of sections, we ensure to wipe down the surface of the water bath thoroughly, ensuring no remnants from the previous tissue block are left behind.

After finishing each block, it is essential to clear excess wax from the microtome to maintain equipment cleanliness and avoid any possible carry-over. To promote good section adhesion to the slides and aid in drying, the cut section slides are placed on the hot plate. We ensure to engage the brake and blade guard when our service is completed. Used blades are disposed of in the Sharp's container if necessary. All wax and debris are removed from the microtome, and the water bath is completely emptied and cleaned. Lastly, depending on specific needs, the sliced sections can be stored in slide baskets either at room temperature or in the oven. Figures [A13](#) and [A14](#) represents labelled FFPE tissue samples on each slide for ChRCC and RO respectively.

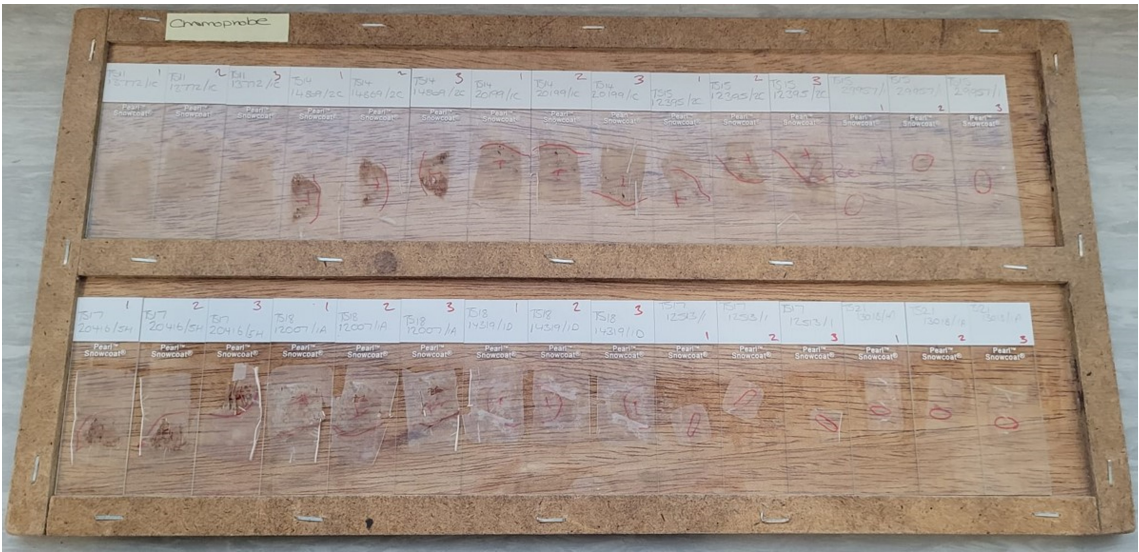


Figure A13. Representation of labelled FFPE tissue samples on each slide for Chromophobe patients.

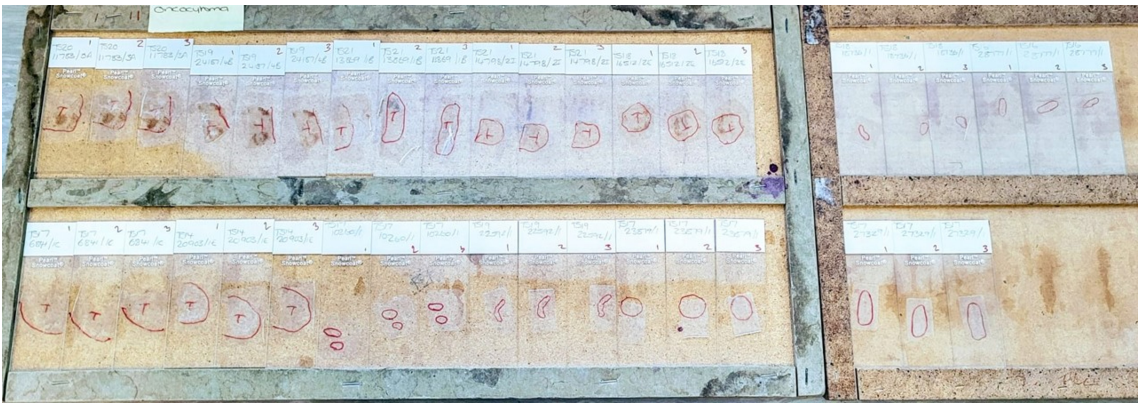


Figure A14. Representation of labelled FFPE tissue samples on each slide for oncocytoma patients.

(Fresh-Frozen Samples):

Tissue samples are obtained from patients who have given their consent to initiate this process [19]. Once collected, the samples undergo a careful examination to confirm their authenticity. The next step involves transferring these samples to Pathology Reception. Here, the pathologist holds the final decision regarding tissue dissection. To aid in sample orientation, the pathologist may annotate tissue margins following NHS SOP NLHR009, Surgical Cut Up. The pathologist separates both healthy and diseased tissue into labelled cryovials and provides essential information if excess tissue is available and permitted for retention.

The tissue samples in the labelled cryovials are then snap-frozen in liquid nitrogen, with the precise freezing time duly noted on the pathology form. The separated material, designated as cancerous and normal, should be placed into two distinct cryo-boxes and promptly stored in the freezer at -80 degrees Celsius. Once the boxes are filled, they should be evenly distributed between the two freezers specified to be kept at -80°C.

Ultimately, two copies of the Pathology Form are produced for Tissue Bank documentation. One copy is retained in the "Collected Tissue Pathology and Consent Forms" folder, while the original signed Consent Form and a copy are deposited in the "Awaiting cut-up" folder [19].

Appendix A.2.2. (DNA Extraction)

An essential method in molecular biology, DNA extraction involves separating and purifying DNA molecules from biological materials such as tissues, blood, or cells [105]. For this study, a protocol from the Maxwell RSC DNA FFPE kit was followed for DNA extraction [106]. Since we had different types of tissue samples, including FFPE and fresh-frozen samples, we adopted two different methods for DNA extraction from each sample type [106,107].

(FFPE Preparation and Preprocessing for DNA Extraction):

Materials for FFPE samples:

- FFPE tissue samples (of volume 2.0 mm³).
- Maxwell RSC DNA FFPE Kit
- Micro-centrifuge
- Vortex
- Razor blades
- Micro-tubes (1.5-2.0ml)
- Pipettors and pipette tips
- Heating blocks
- Deionised or nuclease free water

Slide-mounted tissue sections were scraped for five minutes. Subsequently, the FFPE tissue sections were inserted into a microcentrifuge tube (patients had 3 slides sectioned for FFPE tissue samples, so we scraped all 3 slides and added them to the tube) and centrifuged for fifteen seconds, collecting the sample at the bottom of the tube. Tissue sections made of FFPE with a maximum volume of 2 mm³ were utilised [106].

For the manual preprocessing step, sample tubes were filled with 300 μ l of mineral oil and vortexed for ten seconds. The samples were then heated to 80°C for 2 minutes and allowed to cool to room temperature. Following the instructions in the Maxwell Manual, a master mix of Lysis Buffer, Proteinase K Solution, and Blue Dye was carefully prepared. Each sample tube was then filled with 250 μ l of the master mix, and vortexing was performed for 5 seconds. Layers were separated by centrifugation at 10,000 $\times$ g for 20 seconds, with close attention paid to any pellets that might have formed in the aqueous layer. After a 30-minute incubation at 56°C, the sample tubes were transferred to an 80°C heating block for a 4-hour incubation period [106].

Following this phase, the samples were allowed to cool to room temperature and then further incubated for five minutes. Concurrently, cartridge preparation was initiated, following the instructions provided in the protocol. The blue, aqueous phase was finally separated after centrifugation at maximum speed for five minutes. It was promptly transferred to well #1 of a Maxwell8 FFPE Cartridge. The cartridges were placed in the deck trays with well #1 positioned away from the tubes used for elution. The laboratory technique proceeded with accuracy, completing each stage as it progressed and meeting all standards [106]. Following that, the Maxwell instrument run was performed for FFPE samples.

(Fresh-Frozen Sample Preparation for DNA Extraction):

Materials for Fresh-Frozen Samples:

- Fresh-frozen tissue samples (of volume 2.0 mm³).
- Vortex
- Pipettors and pipette
- Micro-tubes
- Dry heat block
- Deionised or nuclease free water
- 1x Phosphate-buffered saline

We initiated the tissue lysis procedure by labelling the incubation tubes and adjusting the temperature to 56°C. After transferring 5–50 mg of tissue, we centrifuged at top speed for 15 seconds to gather tissue pieces at the bottom of the tube. Subsequently, we carefully changed the pipette tips in each tube to avoid cross-contamination and added 300 μ l of Nuclease-Free Water and 30 μ l of Proteinase K (PK) Solution. Following each addition, a 10-second vortexing phase was carried out. Then, we vortexed once more after adding 300 μ l of Lytic Enhancer (LE2) using the same tip-changing procedure. Without shaking, the incubation was conducted for a minimum of 16 hours at 56°C.

Following the incubation, we vortexed each tube for ten seconds. To remove any remaining undigested material, we next centrifuged at maximum speed for five minutes. Next, we carefully poured all of the supernatant, making sure not to include any pelleted material, into fresh tubes. We dispensed 300 μ l of Lysis Buffer into these new tubes, replacing the pipette tips after each addition. We completed another 10-second vortex and readied the cartridges. To ensure homogeneity, the tissue lysate samples were transferred to separate cartridges into the largest well (well #1). During each sample transfer, pipette tips were carefully changed to prevent cross-contamination. The samples were combined with the binding solution by aspirating and dispensing at least ten times [107].

Afterwards we proceeded to the preparation of the Maxwell RSC Genomic DNA Cartridges. We arranged the cartridges on the deck tray outside the instrument in preparation for the purification procedure. Firmly snapping each cartridge into place, we positioned them in the deck trays farthest away from the elution tubes, ensuring they were placed in the largest well, well #1. After confirming that each cartridge was fully inserted into the deck tray on both sides, we carefully removed the seal, ensuring all adhesive residue and sealing tape were removed from the top of the cartridge.

Additionally, one plunger was inserted into well #8 of each Maxwell RSC Cartridge, and 15 μ l of RNase A Solution was added to well #3 of the cartridges. An empty elution tube was placed in each cartridge in the deck trays, and to complete the setup, we added 50–200 μ l of elution buffer to the bottom of each tube [107].

Appendix A.2.3. (Maxwell Instrument Run)

We powered on the Maxwell Instrument and Tablet PC, logged in, and double-tapped the desktop icon to open the Maxwell software in order to start the extraction process. Every movable part of the instrument was checked and adjusted by hand. Next, we selected "Start" from the 'Home' menu. The Research Sample Concentrator (RSC) Genomic DNA method was the first one we chose on the 'Methods' screen, and we double-checked that it was the right method before clicking the "Proceed" button. We provided the necessary expiration and kit lot information when asked. We chose the cartridge locations for the extraction run and validated the Maxwell RSC Genomic DNA technique at the top of the "Cartridge Setup" page. After entering any relevant sample tracking data, we continued.

We placed the deck trays on the Maxwell Instrument platform after ensuring that all the steps on the Extraction Checklist had been completed when the door was opened [107]. These steps included making sure there were samples in cartridge well #1, loaded cartridges, uncapped elution tubes with Elution Buffer, and plungers in well #8. The extraction run started when the "Start" button was pressed, and the door closed as the platform withdrew. The device completed the purification run, displaying the stages that were still in progress and the projected amount of time left on the screen.

When it finished, the method-end message appeared, and we opened the door by following the on-screen directions. We confirmed that the plungers were in cartridge well #8 and consulted the Maxwell Instrument Technical Manual for the Clean-Up procedure in case the plungers had not been removed from the plunger bar. We quickly removed the deck trays after the run to prevent the eluate from evaporating, sealed the DNA-containing elution tubes, and disposed of the cartridges and

plungers as hazardous waste. The Maxwell instrument was run twice separately for both the FFPE and fresh-frozen samples [107].

Following that, we use a Qubit Flex fluorometer to measure the DNA. We performed broad range analysis on double-stranded DNA samples by adding stranded 1, stranded 2, and 1 μ l of sample DNA before running the sample measurement. To determine the amount of DNA in each sample, we went through this procedure with each one. 10 ng/ μ l is the minimal acceptable result for DNA quantity. Figures A15, A16 and A17 represent of the lab work for DNA extraction.

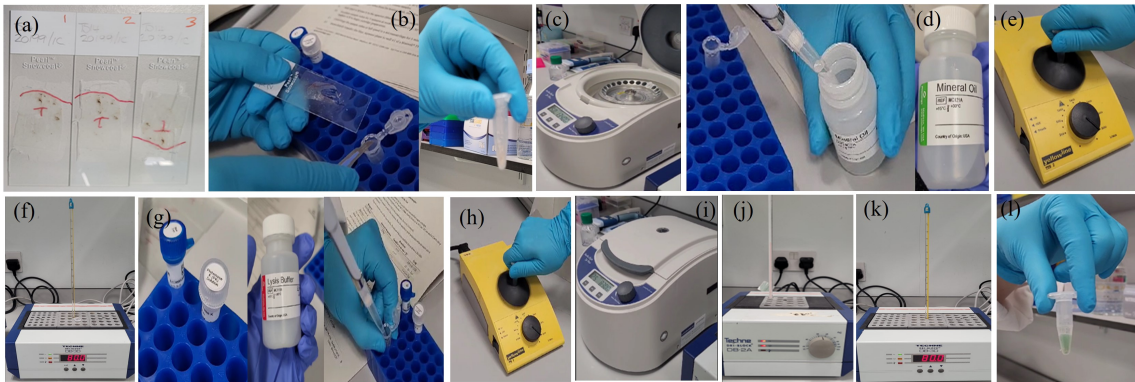


Figure A15. Representation of the lab work for DNA extraction. (a) Use 3 FFPE slides for each patient. (b, c) Scrape the FFPE sample tissue slides, put them into a microcentrifuge tube, and centrifuge for 15 seconds. (d, e) Add 300 μ l of mineral oil and vortex for 10 seconds. (f) Heat the sample to 80°C for 2 minutes. (g) Make a master mix of 224 μ l of Lysis Buffer, 25 μ l of Proteinase K, and 1 μ l of Blue Dye; add it to the sample. (h) Vortex for 5 seconds. (i) Centrifuge for 20 seconds. (j) Incubate in a heat block at 56°C for 30 minutes. (k) Heat the sample at 80°C for 4 hours. (l) Remove the samples and leave them at room temperature for 5 minutes.



Figure A16. Representation of the lab work for DNA extraction using the Maxwell instrument. (a) Add 10 μ l of RNase A solution to the sample and incubate for 5 minutes. (b) Manual preparation of the Maxwell FFPE cartridge: Open the Maxwell instrument. Place the cartridges to be used in the deck tray. For each cartridge, place a plunger into well # 8. For each cartridge in the deck tray, place an empty elution tube. (c) Add 50 μ l of Nuclease-Free Water to the elution tube. (d) Centrifuge the incubated DNA for 5 minutes. (e) Select the type of tissue used for Maxwell to extract DNA (FFPE in this case). (f) Transfer the blue part of the sample containing DNA into well # 1. (g) It takes 40 minutes for the Maxwell to extract the DNA. (h) Take the samples and store them in the fridge. The water and the DNA are visible in the tube.

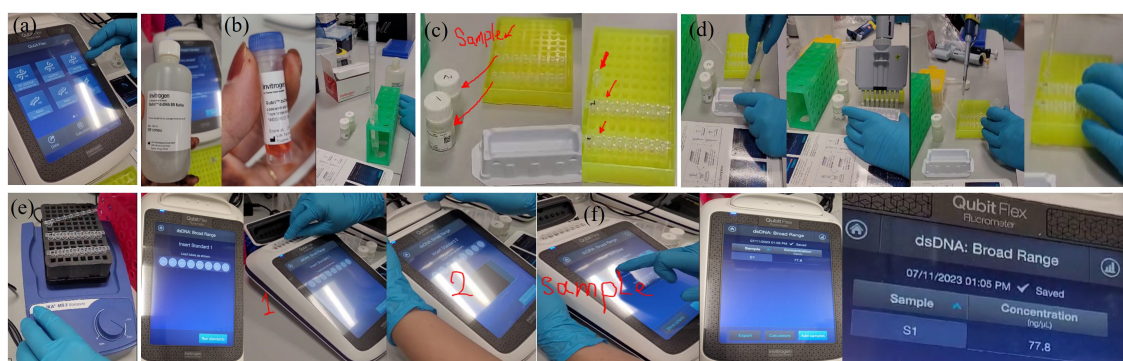


Figure A17. Representation of the lab work for DNA extraction measurement using the Qubit Flex fluorometer and preparing a working solution. (a) Select double stranded DNA and a broad range option. (b) Add 20 μl of dye and 3980 μl of buffer to a tube. (c) Two standards, type 8 and type 1, and one sample are needed. (d) Add the prepared solution to the white plate. For the standards preparation, put 120 μl of the prepared solution in each tube, then add 10 μl of each standard (standard 1 and standard 2) to the tubes. The last empty tube is for the sample; add 199 μl of the prepared solution and 1 μl of the sample to that last tube. (e) Vortex for 3 seconds, then let it sit for two minutes. (f) To measure the DNA, add Standard 1 first, then Standard 2, and finally the sample of 1 μl . The measurement for this sample gave 77.8 ng/ μl . The minimum accepted result is 10 ng/ μl .

Appendix A.2.4. (BeadChip DNA assay)

BeadChips rapidly and accurately examines up to one million single nucleotide polymorphism (SNP) sites on a single BeadChip using the potent Infinium Assay [15]. ~850K SNPs with enriched coverage for 3262 dosage-sensitive genes are included in the Infinium CytoSNP-850K v1.4 BeadChip method. 50mer (nucleotides long) SNP probes are used for high target specificity, which improves low-level mosaic identification and provides precise breakpoint estimation for copy number variations and absence of heterozygosity (AOH) events [108].

With as few as 10 consecutive probes, the BeadChip can reliably identify CNV and AOH calls because to its high 15 $\times$ bead redundancy for a high signal-to-noise ratio. With 848,902 total markers and a call rate of > 98%, the BeadChip takes 200 ng of DNA input refer to Table A18.

The Infinium cytosnp-850k v1.4 beadchip datasheet also provides marker information, which includes the total number of markers (848,902), RefSeq genes (467,422), Absorption, Distribution, Metabolism, and Excretion (ADME) genes (15,153), and more refer to Table A19 [108]. This process further involves following steps as mentioned in Infinium CytoSNP-850K BeadChip Reference Guide [26].

Table A18. Infinium CytoSNP-850K BeadChip product information.

Feature	Description
Species	Human
No. of samples per BeadChip	8
DNA input requirement	200 ng
Assay chemistry	Infinium HD Super
SNP replicates	15×
No. of SNPs to call CNV	10
Instrument support	iScan System
Total no. of markers	848,902
Sample throughput per week	960
Scan time per sample	5 min
Data performance	iScan System
Call rate	99.89%
Reproducibility	99.99%
Log R deviation	0.0929

Table A19. Infinium CytoSNP-850K v1.4 BeadChip marker information.

Marker categories	No. of markers (iScan System)
Total no. of markers	848,902
RefSeq genes	467,422
RefSeq +/- 10 kb	541,515
ADME genes	15,153
ADME +/- 10 kb	18,590
COSMIC genes	418,131
HLA markers	5145
HLA genes	276
GO genes	137,873
Exonic regions	68,801
Promoter regions	26,814 cont..
X chromosome markers	29,894
Y chromosome marker	1197
PAR/homologous markers	728

Appendix A.2.5. (DNA Amplification)

The genomic DNA (gDNA) samples are added to the MSA1 plate in this stage. To provide enough input for the assay, the samples are first denatured and neutralised in the plate, and then they are amplified during the course of an overnight incubation [26].

Materials:

- 0.1 N NaOH

- DNA samples (50 ng/ μ l)
- MA1
- MA2
- MSM
- 96-well 0.8 ml midi plate

Methods:

For the procedure of post-amplification, first, we made sure all of the consumables were ready according to Table A20, and heated the Illumina Hybridisation Oven to 37°C. Next, using the given figure as a guide, we prepared the MSA1 plate by adding 20 ng/ μ l of MA1 to particular wells and designating column 1 for samples for a single BeadChip. Next, we moved 4 μ l of the DNA sample (50 ng/ μ l) from the tubes or DNA plate to the appropriate locations in the MSA1 plate. We filled each DNA sample well with 4 μ l of 0.1 N NaOH to aid in hybridisation. A 96-well cap mat was used to seal the MSA1 plate, and it was vortexed for one minute at 1600 rpm.

After that, it was centrifuged at 280 x g. We incubated the sealing mat at room temperature for ten minutes, and then we carefully removed it and placed it aside upside down. Next, we added 75 μ l of MSM and 68 μ l of MA2 to each sample well. We performed pulse centrifugation and vortexing after resealing the MSA1 plate in the original position. As the last step in our lab work, the sealed plate was placed in the Illumina Hybridisation Oven that had been prepared for 20 to 24 hours. This marked the preparation of the material for additional analysis. Refer to Figure A18 for the procedure of DNA amplification.

Table A20. Preparation of consumables for DNA amplification.

Item	Storage	Instructions
DNA	-25 C to -15 C	Thaw at room temperature
MA1	-25 C to -15 C	Thaw at room temperature, invert 10 times to mix, and then pulse centrifuge
MA2	-25 C to -15 C	Thaw at room temperature, invert 10 times to mix, and then pulse centrifuge
MSM	-25 C to -15 C	Thaw at room temperature, invert 10 times to mix, and then pulse centrifuge

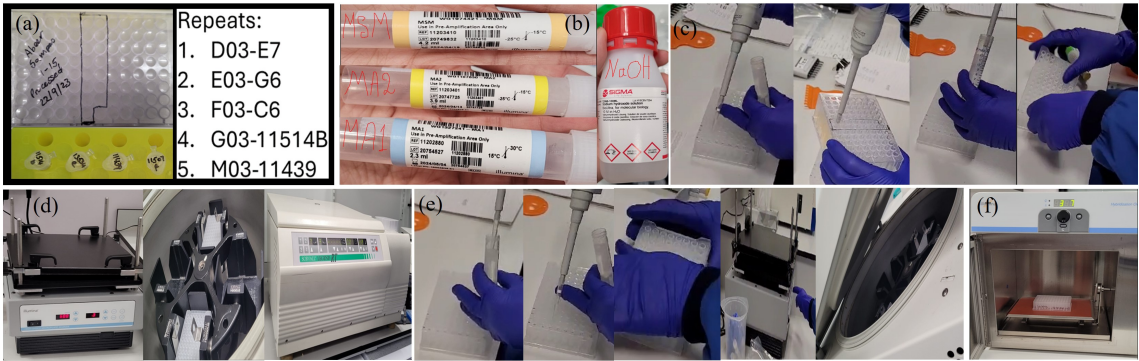


Figure A18. Representation of the lab work for DNA amplification: (a) 19 samples (FFPE and fresh tissue) with 5 repeated samples used in the study. (b) MSM, MA2, MA1, and NaOH are materials used for DNA amplification. (c) Add 20 µl of MA1 to the midi plate wells to prepare MSA1. Add 4 µl of DNA samples to the corresponding wells of the MSA1 plate. Add 4 µl of NaOH to the wells of the MSA1 plate and seal the plate. (d) Vortex for 1 minute, then centrifuge and incubate at room temperature for 10 minutes. (e) Remove the sealing, then add 68 µl of MA2 to each sample well. Add 75 µl of MSM to each sample well, then reseal, vortex, and centrifuge. (f) Incubate in the oven for 24 hours.

Appendix A.2.6. (DNA Fragmentation)

In order to prevent over-fragmentation, endpoint fragmentation is used in this enzymatic DNA fragmentation stage [26].

Materials:

- FMS (1 tube/ can hold 96 samples(we have 19 samples))

Methods:

We ensured that the consumables were ready as indicated in Table A21 and preheated the heat block with the midi plate insert to 37°C.

Table A21. Preparation of consumables for DNA amplification.

Item	Storage	Instructions
FMS	-25 C to -15 C	Thaw at room temperature, invert 10 times to mix, and then pulse centrifuge.

Following the hybridisation process, we took the MSA1 plate out of the oven and pulse centrifuged it at 280 x g. After carefully removing the sealing mat, it was put in a safe place upside down. After that, we filled each sample well in the MSA1 plate with 50 µl of FMS and replaced the cap mat in its original position. The plate was pulse centrifuged at 280 x g after being vortexed for one minute at 1600 rpm. As part of our lab protocol, we then incubated the sealed plate on the heated heat block for an hour. Refer to Figure A19 for DNA fragmentation and precipitation.

Appendix A.2.7. (DNA Precipitation)

In this stage, the DNA is precipitated using PM1 and 100% 2-propanol [26].

Materials:

- 100% 2-propanol

- PM1

Methods:

Until all of the preparations were finished, we left the MSA1 plate on the heat block. After thawing the frozen plates at room temperature, we pulse centrifuged them at 280 x g. In accordance with Table A22, we made sure that all of the necessary consumables were available and heated the heat block to 37°C.

Table A22. Preparation of consumables for DNA amplification.

Item	Storage	Instructions
PM1	-25 C to -15 C	Thaw at room temperature, invert 10 times to mix.

After that, we vortexed each sample well for one minute at 1600 rpm after adding 100 µl of PM1 and resealing the plate with the cap mat in its original position. After five minutes of incubation on the heated heat block, the sealed plate was centrifuged at 280 times the grain size. We set the centrifuge at 4°C in order to get ready for the next centrifuge step. Next, we filled each sample well with 300 µl of 100% 2-propanol and covered it with a fresh, dry cap mat. We were careful not to move the plate until the cap mat was firmly in position. After flipping the plate ten times to make sure everything was well mixed, we let it sit for fifteen minutes at 4°C in the refrigerator.

The plate was then centrifuged for 20 minutes at 3000 x g. We removed the plate and threw away the cap mat after positioning the plate next to another plate of similar weight in the 4°C centrifuge. We promptly inverted the plate to extract the supernatant and held it over an absorbent pad to treat the samples further. We didn’t want any liquid on the pad, so we let it drip onto the absorbent pad and then carefully banged the plate down. We vigorously tapped the plate while it was inverted until all of the wells were empty, which took about a minute. Ultimately, we let the pellet air dry for an hour at room temperature by setting the inverted, uncovered plate on a tube rack. Refer to Figure A19 for DNA fragmentation and precipitation.

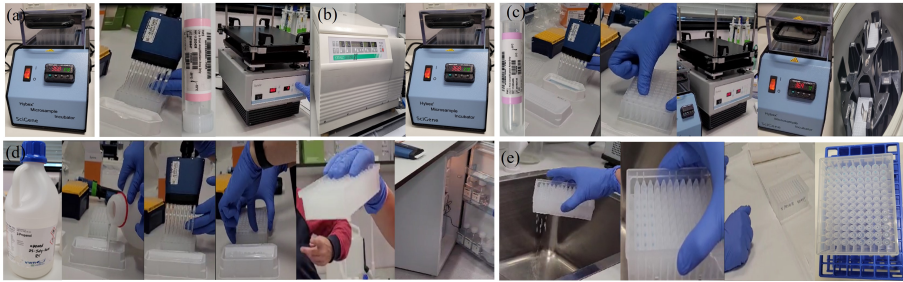


Figure A19. Representation of the lab work for DNA fragmentation and precipitation: (a) Preheat the block plate at 37°C. (b) Add 50 µl FMS to each sample well, vortex, centrifuge then heat the block for one hour. (c) Add 100 µl of PM1 to the samples, vortex, incubate for 5 min and centrifuge at 4°C. (d) Add 300 µl 100% 2-propanol to each sample well, invert 10 times to mix then incubate in fridge for 30 min. (e) Drain the plate and place it in absorbent then leave it one hour to dry.

Appendix A.2.8. (DNA Resuspension)

The precipitated DNA is resuspended in this step by using RA1 [26].

Materials:

- RA1

Methods:

Prior to use, we made sure the heat sealer was at least 10 minutes heated, and we first set the Illumina Hybridisation Oven to 48°C. We repeatedly flipped the thawed RA1 vial to dissolve its contents in order to prepare the RA1 solution. To each pellet well in the MSA1 plate, we inserted 46 μ l of RA1. We covered the plate with a foil heat seal, dull side down. We held the heat sealer sealing block down steadily and uniformly for five seconds to make sure the seal was correct. Once all the wells indentations were visible through the foil, we firmly rolled the rubber plate sealer over the plate. We resealed the plate if any of the wells were not clearly defined. Refer to Figure A20 for DNA resuspension.

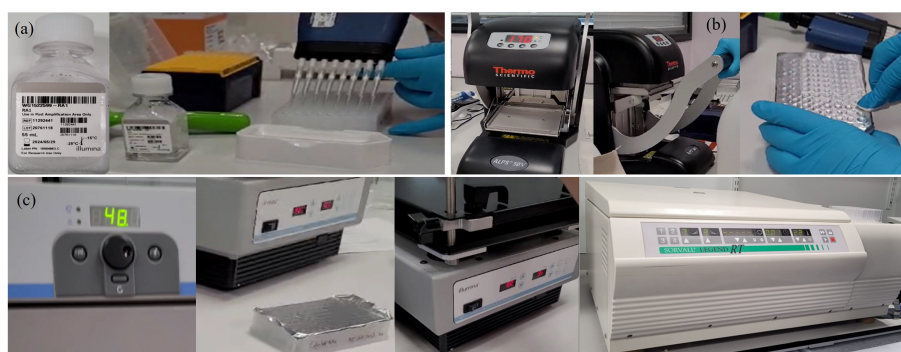


Figure A20. Representation of the lab work for DNA resuspension: (a) Add 46 μ l to the wells of the MSA1 plate. (b) Seal the plate with foil. (c) Incubate the plate at 48°C in the oven, vortex, and then centrifuge.

Appendix A.2.9. (DNA Hybridisation to BeadChip)

The resuspended, fragmented DNA is dispensed onto BeadChips in this stage. Each DNA sample is then hybridised to a portion of the BeadChip by incubation [26].

Preparation:

The Illumina Hybridisation Oven was preheated to 48°C, and the heat block was prepared to 95°C. We placed the MSA1 plate on the heated heat block and let it incubate for 20 minutes in order to denature the DNA. We proceeded with the assembling procedures concurrently with this denaturation. All of the hybridisation chamber inserts, gaskets, and chambers were set up on the benchtop. To ensure correct alignment, each gasket was placed onto the chamber and firmly pressed into place.

In order to finish the assembly, we filled each BeadChip's top and bottom reservoirs with 200 μ l of PB2. Without having to lock it, we quickly covered the chamber with the lid to stop evaporation. After that, the sealed chambers were placed on the benchtop and allowed to sit at room temperature for about an hour so that the DNA could load into the BeadChips. After the MSA1 plate had been incubated for 20 minutes, we moved it from the heat block to the benchtop and let it cool for 30 minutes at room temperature.

Method:

Load DNA onto BeadChips:

The MSA1 plate was first pulse centrifuged at 280 x g. Next, we handled the BeadChips by their ends and gently took them out of the package, keeping them away from the sample inlets. To ensure that the barcode ends aligned correctly, each BeadChip was inserted into an insert. The foil seal on the MSA1 plate was then taken off. Next, we moved 26 μ l of each sample from the MSA1 plate to the appropriate BeadChip sections. We closely monitored the loading port for any extra liquid as we let the DNA spread uniformly across the whole surface.

To make a bolus around the loading port, the area where there was no surplus liquid, we added any remaining sample from the amplification plate. These procedures guaranteed that the BeadChips were loaded precisely and carefully for our lab work.

Set Up of BeadChips for Hybridisation:

To ensure correct alignment, we put the inserts containing BeadChips into the hybridisation chamber. To keep the inserts in place, we inserted the back of the lid into the chamber and then gradually lowered the front. We made sure the lid sat squarely on the base without any gaps by firmly closing all four clamps. With the top logo facing us, the ready chamber was put into the Illumina Hybridisation Oven that had been preheated. We let it sit at 48°C for somewhere between 16 and 24 hours, depending on what our experiment needed. In order to preserve RA1 for use the next day, we stored it at 2°C to 8°C. It's crucial to remember that the MSA1 plate was disposed of safely. For the process of DNA hybridisation to BeadChip refer to Figure A21.



Figure A21. Representation of the lab work for BeadChip DNA Hybridisation: (a) Incubate the MSA1 plate in the oven at 95°C for 20 minutes to denature the DNA. (b) Assemble the hybridisation chambers. (c) Allow the plate to cool down. (d) Add 200 μ l of buffer (PB2) to the top and bottom of the BeadChip reservoirs, then cover it. (e) Centrifuge the MSA1 plate. (f) Place the three BeadChips into their inserts. (g) Transfer 26 μ l from the MSA1 plate to each section of the BeadChip, close the lid, and incubate at 48°C overnight.

Appendix A.2.10. (Resuspend XC4)

To get ready for the Extend and Stain BeadChip phase, resuspend XC4 by filling the vial with 330 ml of fresh 100% EtOH.

Preparation:

Before continuing, we took each hybridisation chamber out of the hybridisation oven and gave it a chance to cool for 25 minutes. Next, we prepared two wash dishes, filled them with 200 ml of PB1, and labelled them so that they could be quickly identified. At the same time, we measured exactly 150 ml of PB1 using a graduated cylinder and filled the Multi-Sample BeadChip Alignment Fixture. We also took out of storage the parts that we needed for the Te-Flow flow-through chamber, which included the spacers, clamps, clean glass back plates, and black frames. This meticulous approach made sure we had everything we needed and were prepared for the next phases in our lab technique, which involved washing.

Method:

Place the wash rack with its wire handle attached into one of the wash dishes that holds 200 ml of PB1.

Appendix A.2.11. (Perform Single-Base Extension)

We incubated 150 μl of RA1 five times, each time for 30 seconds. Then, 450 μl of XC1 was added and incubated for ten minutes. This was followed by 450 μl of XC2 and an identical incubation period. 200 μl of TEM was added after that, and we incubated for 15 minutes. After that, we added 450 μl of a 95% formamide/1 mM EDTA solution and incubated for 1 minute. We then repeated this process once and incubated for an additional 5 minutes. In order to match the STM tube temperature, we also altered the temperature of the chamber rack. Finally, to complete our exact and thorough washing procedure in the lab, 450 μl of XC3 was added and incubated for 1 minute, once again. Refer to Figure A22 for XC4 resuspension and the Single-Base Extension.



Figure A22. Representation of the lab work done for XC4 resuspension and the performance of the single-base extension: (a) Let the chamber cool for 10 min. (b) Add 330 ml of EtOH to the XC4 bottle, then shake well. (c) Add 200 ml of PB1 to two wash-dishes. (d,e,f) From the chamber component, remove the black frame from the BeadChip. Insert the BeadChip, into the wash-dish containing Buffer (PB1). Return the three washed BeadChips to the chamber and add buffer (PB1) make sure they are covered. Add the thin then add the glass and fix with clips. Cut the extra spacer and add the BeadChips to the chamber rack. (g) Adjust the chamber rack to be at a temperature of 44°C. Add 15 μl of RA1 and incubate for 30 seconds; repeat this 5 times. (h) Add 450 μl of XC1, then incubate for 10 minutes. (i) Add 450 μl of XC2, then incubate for 10 minutes. (j) Add 200 μl of TEM and incubate for 15 minutes. (k) Add 450 μl of formamide, then incubate for 1 minute and repeat this 1 time. Incubate for 5 minutes. (l) Change the temperature to 32°C on the STM tube. (m) Add 450 μl of XC3, then incubate for 1 min. Repeat this once.

Appendix A.2.12. (Stain BeadChips)

First, we added 250 μ l of STM to the reservoir in each chamber and incubated for ten minutes. After that, we added 450 μ l of XC3 and once again incubated for one minute. Following a 5-minute pause, we added 250 μ l of ATM and incubated for 10 minutes. Next, we added 450 μ l of XC3 and waited for an additional 5-minutes before continuing with the incubation for an additional minute. Every flow-through chamber underwent this cycle once more. The flow-through chambers were then promptly taken out of the chamber rack and set aside in alignment fixtures that were sitting on a lab bench, submerged in PB1 at room temperature.

Appendix A.2.13. (Wash and Coat BeadChips)

In order to remove BeadChips, we set up a rack to fit inside a vacuum desiccator and placed a clean tube rack on absorbent material. We filled two top-loading wash dishes, designated PB1 and XC4, with 310 ml of water, marked the water level, and then emptied it. Following that, we filled the PB1 wash dish with 310 ml of PB1. We immersed the staining rack in the PB1 wash dish, facing the locking arms and tab in our direction, in order to disassemble the flow-through chambers. We took off the two metal clamps and raised the glass back plate straight up for a deeper clean using a disassembly tool.

We took the BeadChip out of the black frame, being careful to hold it by the edges or barcode end, and carefully removed the spacer, being careful not to come into contact with the BeadChip stripes. Each flow-through chamber underwent this same procedure, wherein the BeadChips were positioned in the submerged staining rack with their arms locked and barcode facing away from us. We quickly raised and seated each BeadChip within the staining rack, lifting it up and down ten times to break the surface, in order to guarantee enough coverage and prevent BeadChip contact. We put 310 ml to the XC4 wash dish after letting the XC4 bottle shake to resuspend its contents, if necessary, after a 5-minute soak. After that, we moved the staining rack from the PB1 wash dish to the XC4 wash dish, soaking for an extra five minutes and doing the up-and-down motion ten times.

The staining rack was then taken out, set on the ready-made tube rack, making sure it was centred for even coating, and the staining rack handle was taken off. We took each BeadChip and held it by the barcode end with self-locking tweezers, then set it on a tube rack with the barcode facing up and towards us to dry. We arranged the BeadChips top to bottom and dried them for 50–55 minutes, making sure they didn't come into contact with one another or settle on the tube rack edge. Refer to Figure A23 for Staining, washing and Coating the BeadChip.



Figure A23. Representation of the lab work done for BeadChip staining, washing, and coating: (a) Add 250 μl of STM, then incubate for 10 minutes. (b) Add 450 μl of XC3, then incubate for 1 minute. Repeat this one time, then wait for 5 minutes. (c) Add 250 μl of ATM, and incubate for 10 minutes. (d) Add 450 μl of XC3, then incubate for 1 minute. Repeat this one time, then wait for 5 minutes. (e) Add 250 μl of STM, then incubate for 10 minutes. (f) Add 450 μl of XC3, then incubate for 1 minute. Repeat this one time, then wait for 5 minutes. (g) Add 250 μl of ATM, and incubate for 10 minutes. (h) Add 450 μl of XC3, then incubate for 1 minute. Repeat this one time, then wait for 5 minutes. (i) Add 250 μl of STM, then incubate for 10 minutes. (j) Add 450 μl of XC3, then incubate for 1 minute. Repeat this one time, then wait for 5 minutes. (k) Remove the flow-through chamber from the chamber rack and put the BeadChip in PB1. Move it up and down 10 times, then leave it for 5 minutes. (l) Put the BeadChip in XC4, move it up and down 10 times, then leave it for 5 minutes. (m) Clean it with ethanol and leave the BeadChip to dry for one hour.

Appendix A.2.14. (Scan and Analyse BeadChips)

The BeadChips must then be scanned using either the iScan or NextSeg 550 systems. During the scan, output files are created and saved in the designated output folder [109]. Refer to Figure A24 for BeadChip scanning.

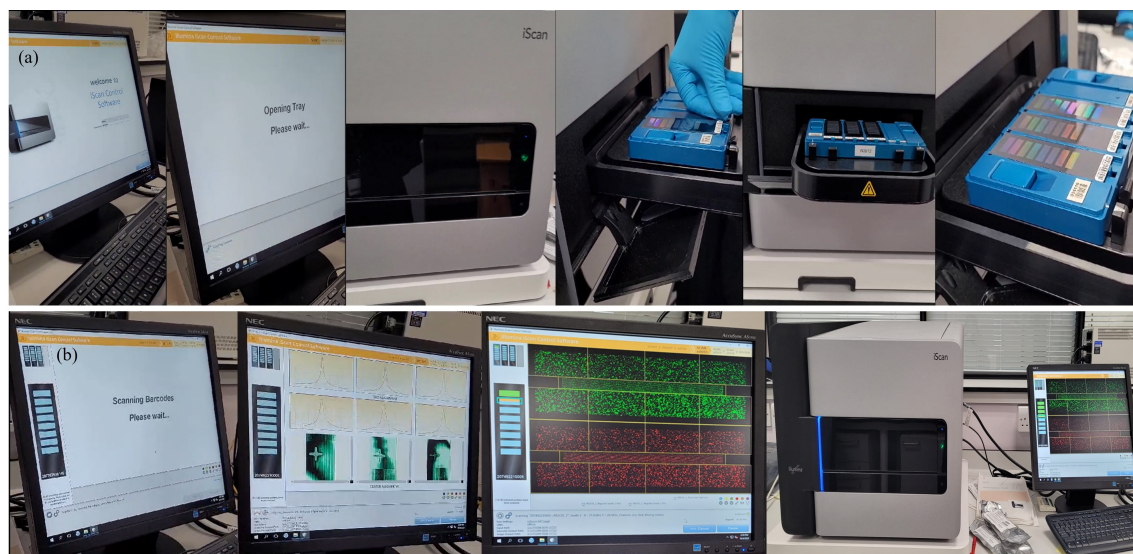


Figure A24. Representation of BeadChip scan and analysis: (a) Insert the BeadChips into the Illumina iScan machine. (b) The system scans the BeadChips, and the output files are created.

Appendix A.2.15. (SNP-based Microarrays: A Simple Summary for Non-specialists)

Genotyping using SNP-based microarrays for copy number variation involves identifying not only SNPs but also larger variations in the genome, called copy number variations. CNVs are large sections of the genome that can be duplicated or deleted in some individuals. These variations can affect gene function and sometimes influence traits or lead to diseases. SNP-based microarrays can detect CNVs by following a series of steps. First, a DNA tissue sample, is collected from the patient, just like in regular SNP genotyping. This DNA is then applied to a microarray chip that contains thousands or even millions of spots called probes. These probes are specifically designed to detect both SNPs, which represent small changes (like a single nucleotide difference), and CNVs, which involve larger genomic changes, such as sections of DNA being duplicated or deleted.

During hybridisation, DNA binds to the probes on the chip if there is a match in nucleotide sequences. For SNPs, the process detects which nucleotide is present, while for CNVs, the intensity of the hybridisation signal indicates whether extra copies or deletions of DNA segments are present. A strong signal may suggest additional copies, while a weak signal can imply a deletion. After hybridisation, the chip is scanned, and the results are processed to identify SNPs and assess CNVs. This combined data provides valuable insight into how genetic variations influence diseases, traits, and drug responses.

SNP arrays focus on specific SNP locations as they target known and important genetic variations. It provides a pre-selected target sites. SNP arrays are designed to look at the most informative spots in the genome, where variations SNPs are most likely to have an impact on traits or diseases. By focusing on these SNPs, the array can efficiently detect meaningful genetic differences without scanning the entire genome. Using SNP-based microarrays for CNV detection is advantageous for several reasons. It provides dual detection of both SNPs and CNVs in a single experiment, giving a more complete picture of genetic variation. This method is also quick and cost-effective, allowing for the study of large numbers of SNPs and CNVs simultaneously rather than requiring separate procedures. Additionally, CNVs are particularly important in disease studies, as they can play a significant role in conditions like cancer, autism, and developmental disorders, where large sections of DNA may be duplicated or deleted. SNP-based microarrays for CNV genotyping allow researchers to examine both small and large genetic changes in an individual's DNA, offering a comprehensive view of genetic variation.

References

1. Raman, S.P.; Johnson, P.T.; Allaf, M.E.; Netto, G.; Fishman, E.K. Chromophobe renal cell carcinoma: multiphase MDCT enhancement patterns and morphologic features. *American Journal of Roentgenology* **2013**, *201*, 1268–1276.
2. Wobker, S.E.; Williamson, S.R. Modern pathologic diagnosis of renal oncocytoma. *Journal of kidney cancer and VHL* **2017**, *4*, 1.
3. Choudhary, S.; Rajesh, A.; Mayer, N.; Mulcahy, K.; Haroon, A. Renal oncocytoma: CT features cannot reliably distinguish oncocytoma from other renal neoplasms. *Clinical radiology* **2009**, *64*, 517–522.
4. Rosenkrantz, A.B.; Hindman, N.; Fitzgerald, E.F.; Niver, B.E.; Melamed, J.; Babb, J.S. MRI features of renal oncocytoma and chromophobe renal cell carcinoma. *American Journal of Roentgenology* **2010**, *195*, W421–W427.
5. Wu, J.; Zhu, Q.; Zhu, W.; Chen, W.; Wang, S. Comparative study of CT appearances in renal oncocytoma and chromophobe renal cell carcinoma. *Acta Radiologica* **2016**, *57*, 500–506.
6. Saha, A.; Harowicz, M.R.; Grimm, L.J.; Kim, C.E.; Ghate, S.V.; Walsh, R.; Mazurowski, M.A. A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. *British journal of cancer* **2018**, *119*, 508–516.
7. Tamez-Pena, J.G.; Rodriguez-Rojas, J.A.; Gomez-Rueda, H.; Celaya-Padilla, J.M.; Rivera-Prieto, R.A.; Palacios-Corona, R.; Garza-Montemayor, M.; Cardona-Huerta, S.; Treviño, V. Radiogenomics analysis identifies correlations of digital mammography with clinical molecular signatures in breast cancer. *PloS one* **2018**, *13*, e0193871.
8. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; Van Stiphout, R.G.; Granton, P.; Zegers, C.M.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer* **2012**, *48*, 441–446.
9. Zhao, M.; Wang, Q.; Wang, Q.; Jia, P.; Zhao, Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC bioinformatics* **2013**, *14*, S1.
10. Pinto, D.; Darvishi, K.; Shi, X.; Rajan, D.; Rigler, D.; Fitzgerald, T.; Lionel, A.C.; Thiruvahindrapuram, B.; MacDonald, J.R.; Mills, R.; et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature biotechnology* **2011**, *29*, 512–520.
11. Zhang, F.; Gu, W.; Hurles, M.E.; Lupski, J.R. Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics* **2009**, *10*, 451–481.
12. Redon, R.; Ishikawa, S.; Fitch, K.R.; Feuk, L.; Perry, G.H.; Andrews, T.D.; Fiegler, H.; Shapero, M.H.; Carson, A.R.; Chen, W.; et al. Global variation in copy number in the human genome. *nature* **2006**, *444*, 444–454.
13. Balagué-Dobón, L.; Cáceres, A.; González, J.R. Fully exploiting SNP arrays: a systematic review on the tools to extract underlying genomic structure. *Briefings in bioinformatics* **2022**, *23*, bbac043.
14. Interpreting Infinium Assay Data for Whole-Genome Structural Variation. https://www.illumina.com/Documents/products/technotes/technote_cytoanalysis.pdf accessed (18 July 2024), 2024.
15. DNA Copy Number and Loss of Heterozygosity Analysis Algorithms. https://www.illumina.com/documents/products/technotes/technote_cnv_algorithms.pdf accessed (18 July 2024), 2024.
16. Karlo, C.A.; Di Paolo, P.L.; Chaim, J.; Hakimi, A.A.; Ostrovnaya, I.; Russo, P.; Hricak, H.; Motzer, R.; Hsieh, J.J.; Akin, O. Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations. *Radiology* **2014**, *270*, 464–471.
17. Ferro, M.; Musi, G.; Marchioni, M.; Maggi, M.; Vecchia, A.; Del Giudice, F.; Barone, B.; Crocetto, F.; Lasorsa, F.; Antonelli, A.; et al. Radiogenomics in renal cancer management—current evidence and future prospects. *International journal of molecular sciences* **2023**, *24*, 4615.
18. Posada Calderon, L.; Eismann, L.; Reese, S.W.; Reznik, E.; Hakimi, A.A. Advances in imaging-based biomarkers in renal cell carcinoma: a critical analysis of the current literature. *Cancers* **2023**, *15*, 354.
19. Tayside Biorepository. <https://www.tissuebank.dundee.ac.uk> accessed (27 April 2022), 2022.
20. Yap, F.Y.; Varghese, B.A.; Cen, S.Y.; Hwang, D.H.; Lei, X.; Desai, B.; Lau, C.; Yang, L.L.; Fullenkamp, A.J.; Hajian, S.; et al. Shape and texture-based radiomics signature on CT effectively discriminates benign from malignant renal masses. *European Radiology* **2021**, *31*, 1011–1021.
21. Yi, X.; Xiao, Q.; Zeng, F.; Yin, H.; Li, Z.; Qian, C.; Wang, C.; Lei, G.; Xu, Q.; Li, C.; et al. Computed tomography radiomics for predicting pathological grade of renal cell carcinoma. *Frontiers in oncology* **2021**, *10*, 570396.

22. Alhussaini, A.J.; Steele, J.D.; Nabi, G. Comparative Analysis for the Distinction of Chromophobe Renal Cell Carcinoma from Renal Oncocytoma in Computed Tomography Imaging Using Machine Learning Radiomics Analysis. *Cancers* **2022**, *14*, 3609.
23. Yeap, P.L.; Wong, Y.M.; Ong, A.L.K.; Tuan, J.K.L.; Pang, E.P.P.; Park, S.Y.; Lee, J.C.L.; Tan, H.Q. Predicting dice similarity coefficient of deformably registered contours using Siamese neural network. *Physics in Medicine & Biology* **2023**, *68*, 155016.
24. Python Release Python 3.6.0. <https://www.python.org/downloads/release/python-360/> accessed (29 January 2021).
25. Van Griethuysen, J.J.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H.J. Computational radiomics system to decode the radiographic phenotype. *Cancer research* **2017**, *77*, e104–e107.
26. Infinium CytoSNP-850K BeadChip Assay Reference Guide. <https://support.illumina.com/ko-kr/downloads/infinium-cytosnp-850k-reference-guide-15046990.html> accessed (27 April 2023), 2023.
27. GenomeStudio Software Downloads. https://support.illumina.com/array/array_software/genomestudio/downloads.html accessed (27 February 2024), 2024.
28. GenomeStudio 2.0 Plug-ins. <https://support.illumina.com/downloads/genomestudio-2-0-plug-ins.html> accessed (07 February 2024), 2024.
29. Microarray General Reference Materials. https://knowledge.illumina.com/microarray/general/microarray-general-reference_material-list/000002766 accessed (21 March 2024), 2024.
30. PennCNV: Copy Number Variation (CNV) detection from SNP genotyping arrays. <https://hpc.nih.gov/apps/PennCNV.html> accessed (18 July 2024), 2024.
31. International Standards for Cytogenomic Arrays. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000205.v2.p1 accessed (21 March 2024), 2024.
32. Gurbich, T.A.; Ilinsky, V.V. ClassifyCNV: a tool for clinical annotation of copy-number variants. *Scientific reports* **2020**, *10*, 20375.
33. National Center for Biotechnology Information (NCBI)-nstd102 - Clinical Structural Variants. <https://www.ncbi.nlm.nih.gov/dbvar/studies/nstd102> accessed (21 March 2024), 2024.
34. Database of Genomic Variants. <https://dgv.tcag.ca/dgv/app/home> accessed (08 April 2024), 2024.
35. UCSC Genome Browser. <https://genome.ucsc.edu/> accessed (08 April 2024), 2024.
36. CNV Xplorer. <https://cnvexplorer.com/> accessed (08 April 2024), 2024.
37. CNV ClinViewer. <https://cnv-clinviewer.broadinstitute.org/> accessed (08 April 2024), 2024.
38. BEDsect: A Tool for Feature-based Annotations of Genomic Datasets. <https://imgsb.org/bedsect/> accessed (08 April 2024), 2024.
39. UCSC Genome Browser. <https://genome.ucsc.edu/cgi-bin/hgTables>.
40. Ng, K.L.; Rajandram, R.; Morais, C.; Yap, N.Y.; Samaratunga, H.; Gobe, G.C.; Wood, S.T. Differentiation of oncocytoma from chromophobe renal cell carcinoma (RCC): can novel molecular biomarkers help solve an old problem? *Journal of Clinical Pathology* **2014**, *67*, 97–104.
41. Dvorakova, M.; Dhir, R.; Bastacky, S.I.; Cieply, K.M.; Acquafondata, M.B.; Sherer, C.R.; Mercuri, T.L.; Parwani, A.V. Renal oncocytoma: a comparative clinicopathologic study and fluorescent in-situ hybridization analysis of 73 cases with long-term follow-up. *Diagnostic pathology* **2010**, *5*, 1–6.
42. Vera-Badillo, F.E.; Conde, E.; Duran, I. Chromophobe renal cell carcinoma: a review of an uncommon entity. *International Journal of Urology* **2012**, *19*, 894–900.
43. Kim, J.K.; Kim, T.K.; Ahn, H.J.; Kim, C.S.; Kim, K.R.; Cho, K.S. Differentiation of subtypes of renal cell carcinoma on helical CT scans. *American Journal of Roentgenology* **2002**, *178*, 1499–1506.
44. Bird, V.G.; Kanagarajah, P.; Morillo, G.; Caruso, D.J.; Ayyathurai, R.; Leveillee, R.; Jorda, M. Differentiation of oncocytoma and renal cell carcinoma in small renal masses (< 4 cm): the role of 4-phase computerized tomography. *World journal of urology* **2011**, *29*, 787–792.
45. Akin, I.B.; Altay, C.; Güler, E.; Çamlıdağ, İ.; Harman, M.; Danacı, M.; Tuna, B.; Yörükoğlu, K.; Seçil, M. Discrimination of oncocytoma and chromophobe renal cell carcinoma using MRI. *Diagnostic and Interventional Radiology* **2019**, *25*, 5.
46. Kurup, A.N.; Thompson, R.H.; Leibovich, B.C.; Harmsen, W.S.; Sebo, T.J.; Callstrom, M.R.; Kawashima, A.; Atwell, T.D. Renal oncocytoma growth rates before intervention. *BJU international* **2012**, *110*, 1444–1448.

47. Chawla, S.N.; Crispen, P.L.; Hanlon, A.L.; Greenberg, R.E.; Chen, D.Y.; Uzzo, R.G. The natural history of observed enhancing renal masses: meta-analysis and review of the world literature. *The Journal of urology* **2006**, *175*, 425–431.
48. Baharzadeh, F.; Sadeghi, M.; Ramezani, M. Chromophobe renal cell carcinoma or oncocytoma: a manner of challenge in frozen section diagnosis. *BioMedicine* **2019**, *9*.
49. Shao, X.; Lv, N.; Liao, J.; Long, J.; Xue, R.; Ai, N.; Xu, D.; Fan, X. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC medical genetics* **2019**, *20*, 1–14.
50. Sebat, J.; Lakshmi, B.; Troge, J.; Alexander, J.; Young, J.; Lundin, P.; Manér, S.; Massa, H.; Walker, M.; Chi, M.; et al. Large-scale copy number polymorphism in the human genome. *Science* **2004**, *305*, 525–528.
51. Albertson, D.G.; Pinkel, D. Genomic microarrays in human genetic disease and cancer. *Human molecular genetics* **2003**, *12*, R145–R152.
52. Shaikh, T.H. Copy number variation disorders. *Current genetic medicine reports* **2017**, *5*, 183–190.
53. Füzesi, L.; Frank, D.; Nguyen, C.; Ringert, R.H.; Bartels, H.; Gunawan, B. Losses of 1p and chromosome 14 in renal oncocytomas. *Cancer genetics and cytogenetics* **2005**, *160*, 120–125.
54. Yap, N.Y.; Rajandram, R.; Ng, K.L.; Pailoor, J.; Fadzli, A.; Gobe, G.C. Genetic and chromosomal aberrations and their clinical significance in renal neoplasms. *BioMed research international* **2015**, *2015*, 476508.
55. Ohashi, R.; Schraml, P.; Angori, S.; Batavia, A.A.; Rupp, N.J.; Ohe, C.; Otsuki, Y.; Kawasaki, T.; Kobayashi, H.; Kobayashi, K.; et al. Classic chromophobe renal cell carcinoma incur a larger number of chromosomal losses than seen in the eosinophilic subtype. *Cancers* **2019**, *11*, 1492.
56. Tan, M.H.; Wong, C.F.; Tan, H.L.; Yang, X.J.; Ditlev, J.; Matsuda, D.; Khoo, S.K.; Sugimura, J.; Fujioka, T.; Furge, K.A.; et al. Genomic expression and single-nucleotide polymorphism profiling discriminates chromophobe renal cell carcinoma and oncocytoma. *BMC cancer* **2010**, *10*, 1–12.
57. Krill-Burger, J.M.; Lyons, M.A.; Kelly, L.A.; Sciulli, C.M.; Petrosko, P.; Chandran, U.R.; Kubal, M.D.; Bastacky, S.I.; Parwani, A.V.; Dhir, R.; et al. Renal cell neoplasms contain shared tumor type-specific copy number variations. *The American journal of pathology* **2012**, *180*, 2427–2439.
58. Van den Berg, E.; Van der Hout, A.; Oosterhuis, J.; Störkel, S.; Dijkhuizen, T.; Dam, A.; Zweers, H.; Mensink, H.; Buys, C.; De Jong, B. Cytogenetic analysis of epithelial renal-cell tumors: relationship with a new histopathological classification. *International journal of cancer* **1993**, *55*, 223–227.
59. Herbers, J.; Schullerus, D.; Chudek, J.; Bugert, P.; Kanamaru, H.; Zeisler, J.; Ljungberg, B.; Akhtar, M.; Kovacs, G. Lack of genetic changes at specific genomic sites separates renal oncocytomas from renal cell carcinomas. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **1998**, *184*, 58–62.
60. Wang, M.X.; Liuyu, T.; Zhang, Z.d. Multifaceted roles of the E3 ubiquitin ligase RING finger protein 115 in immunity and diseases. *Frontiers in Immunology* **2022**, *13*, 936579.
61. Amemiya, Y.; Bacopulos, S.; Seth, A. Novel Ubiquitin E3 Ligases as Targets for Cancer Therapy: Focus on Breast Cancer-Associated Gene 2 (BCA2). *Resistance to Proteasome Inhibitors in Cancer: Molecular Mechanisms and Strategies to Overcome Resistance* **2014**, pp. 317–346.
62. Pan, Z. *Identification of novel substrates of the ubiquitin E3 ligase RNF126 and characterization of its role in lipid droplet homeostasis*; University of Toronto (Canada), 2016.
63. Ehsani, L.; Seth, R.; Bacopulos, S.; Seth, A.; Osunkoya, A.O. BCA2 is differentially expressed in renal oncocytoma: an analysis of 158 renal neoplasms. *Tumor Biology* **2013**, *34*, 787–791.
64. Iakymenko, O.A.; Delma, K.S.; Jorda, M.; Kryvenko, O.N. Cathepsin K (clone EPR19992) demonstrates uniformly positive immunoreactivity in renal oncocytoma, chromophobe renal cell carcinoma, and distal tubules. *International journal of surgical pathology* **2021**, *29*, 600–605.
65. Li, G.; Gentil-Perret, A.; Lambert, C.; Genin, C.; Tostain, J. S100A1 and KIT gene expressions in common subtypes of renal tumours. *European Journal of Surgical Oncology (EJSO)* **2005**, *31*, 299–303.
66. Yusenko, M.V. Molecular pathology of chromophobe renal cell carcinoma: a review. *International Journal of Urology* **2010**, *17*, 592–600.
67. Zhu, B.; Rohan, S.M.; Lin, X. Cytomorphology, immunoprofile, and management of renal oncocytic neoplasms. *Cancer Cytopathology* **2020**, *128*, 962–970.
68. Satter, K.B.; Tran, P.M.H.; Tran, L.K.H.; Ramsey, Z.; Pinkerton, K.; Bai, S.; Savage, N.M.; Kavuri, S.; Terris, M.K.; She, J.X.; et al. Oncocytoma-related gene signature to differentiate chromophobe renal cancer and oncocytoma using machine learning. *Cells* **2022**, *11*, 287.

69. Wu, H.; Fan, L.; Liu, H.; Guan, B.; Hu, B.; Liu, F.; Hoher, B.; Yin, L. Identification of key genes and prognostic analysis between chromophobe renal cell carcinoma and renal oncocytoma by bioinformatic analysis. *BioMed Research International* **2020**, 2020, 4030915.
70. Yusenko, M.V.; Kuiper, R.P.; Boethe, T.; Ljungberg, B.; van Kessel, A.G.; Kovacs, G. High-resolution DNA copy number and gene expression analyses distinguish chromophobe renal cell carcinomas and renal oncocytomas. *BMC cancer* **2009**, 9, 1–10.
71. McGillivray, P.D.; Ueno, D.; Pooli, A.; Mendhiratta, N.; Syed, J.S.; Nguyen, K.A.; Schulam, P.G.; Humphrey, P.A.; Adeniran, A.J.; Boutros, P.C.; et al. Distinguishing benign renal tumors with an oncocyctic gene expression (ONEX) classifier. *European urology* **2021**, 79, 107–111.
72. Rohan, S.; Tu, J.J.; Kao, J.; Mukherjee, P.; Campagne, F.; Zhou, X.K.; Hyjek, E.; Alonso, M.A.; Chen, Y.T. Gene expression profiling separates chromophobe renal cell carcinoma from oncocytoma and identifies vesicular transport and cell junction proteins as differentially expressed genes. *Clinical cancer research* **2006**, 12, 6937–6945.
73. Liu, Q.; Cornejo, K.M.; Cheng, L.; Hutchinson, L.; Wang, M.; Zhang, S.; Tomaszewicz, K.; Cosar, E.F.; Woda, B.A.; Jiang, Z. Next-generation sequencing to detect deletion of RB1 and ERBB4 genes in chromophobe renal cell carcinoma: a potential role in distinguishing chromophobe renal cell carcinoma from renal oncocytoma. *The American Journal of Pathology* **2018**, 188, 846–852.
74. Molnar, A.; Horvath, C.A.; Czovek, P.; Szanto, A.; Kovacs, G. FOXI1 immunohistochemistry differentiates benign renal oncocytoma from malignant chromophobe renal cell carcinoma. *Anticancer Research* **2019**, 39, 2785–2790.
75. Ishihara, H.; Yamashita, S.; Liu, Y.Y.; Hattori, N.; El-Omar, O.; Ikeda, T.; Fukuda, H.; Yoshida, K.; Takagi, T.; Taneda, S.; et al. Genetic and epigenetic profiling indicates the proximal tubule origin of renal cancers in end-stage renal disease. *Cancer Science* **2020**, 111, 4276–4287.
76. Giesen, E.; Jilaveanu, L.B.; Parisi, F.; Kluger, Y.; Camp, R.L.; Kluger, H.M. NY-ESO-1 as a potential immunotherapeutic target in renal cell carcinoma. *Oncotarget* **2014**, 5, 5209.
77. Demirović, A.; Džombeta, T.; Tomas, D.; Spajić, B.; Pavić, I.; Hudolin, T. expression of tumor antigens MAGE-A3/4 and NY-ESO-1 in renal oncocytoma and chromophobe renal cell carcinoma. *Pathology*.
78. Coppola, F.; Mottola, M.; Lo Monaco, S.; Cattabriga, A.; Cocozza, M.A.; Yuan, J.C.; De Benedittis, C.; Cuicchi, D.; Guido, A.; Rojas Limpe, F.L.; et al. The heterogeneity of skewness in T2W-based radiomics predicts the response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Diagnostics* **2021**, 11, 795.
79. Çınarlar, G.; Emiroğlu, B.G.; Yurttakal, A.H. Prediction of glioma grades using deep learning with wavelet radiomic features. *Applied Sciences* **2020**, 10, 6296.
80. Belfiore, M.P.; Sansone, M.; Monti, R.; Marrone, S.; Fusco, R.; Nardone, V.; Grassi, R.; Reginelli, A. Robustness of radiomics in pre-surgical computer tomography of non-small-cell lung cancer. *Journal of Personalized Medicine* **2022**, 13, 83.
81. Foy, J.J.; Robinson, K.R.; Li, H.; Giger, M.L.; Al-Hallaq, H.; Armato III, S.G. Variation in algorithm implementation across radiomics software. *Journal of medical imaging* **2018**, 5, 044505–044505.
82. Linsalata, S.; Borgheresi, R.; Marfisi, D.; Barca, P.; Sainato, A.; Paiar, F.; Neri, E.; Traino, A.C.; Giannelli, M. Radiomics of patients with locally advanced rectal cancer: effect of preprocessing on features estimation from computed tomography imaging. *BioMed Research International* **2022**, 2022, 2003286.
83. Yu, H.; Scalera, J.; Khalid, M.; Touret, A.S.; Bloch, N.; Li, B.; Qureshi, M.M.; Soto, J.A.; Anderson, S.W. Texture analysis as a radiomic marker for differentiating renal tumors. *Abdominal Radiology* **2017**, 42, 2470–2478.
84. Rahim, M.A.; Hossain, M.N.; Wahid, T.; Azam, M.S. Face recognition using local binary patterns (LBP). *Global Journal of Computer Science and Technology* **2013**, 13, 1–8.
85. D'Amico, N.C.; Sicilia, R.; Cordelli, E.; Tronchin, L.; Greco, C.; Fiore, M.; Carnevale, A.; Iannello, G.; Ramella, S.; Soda, P. Radiomics-based prediction of overall survival in lung cancer using different volumes-of-interest. *Applied Sciences* **2020**, 10, 6425.
86. Santucci, D.; Faiella, E.; Cordelli, E.; Sicilia, R.; de Felice, C.; Zobel, B.B.; Iannello, G.; Soda, P. 3T MRI-radiomic approach to predict for lymph node status in breast cancer patients. *Cancers* **2021**, 13, 2228.
87. Sicilia, R.; Cordelli, E.; Merone, M.; Luperto, E.; Papalia, R.; Iannello, G.; Soda, P. Early radiomic experiences in classifying prostate cancer aggressiveness using 3D local binary patterns. In Proceedings of the 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2019, pp. 355–360.

88. Tibermacine, H.; Rouanet, P.; Sbarra, M.; Forghani, R.; Reinhold, C.; Nougaret, S.; the GRECCAR Study Group. Radiomics modelling in rectal cancer to predict disease-free survival: evaluation of different approaches. *British Journal of Surgery* **2021**, *108*, 1243–1250.
89. Yu, Y.; Li, X.; Du, T.; Rahaman, M.; Grzegorzec, M.J.; Li, C.; Sun, H. Increasing the accuracy and reproducibility of positron emission tomography radiomics for predicting pelvic lymph node metastasis in patients with cervical cancer using 3D local binary pattern-based texture features. *Intelligent Medicine* **2024**.
90. Jensen, L.J.; Kim, D.; Elgeti, T.; Steffen, I.G.; Schaafs, L.A.; Hamm, B.; Nagel, S.N. Enhancing the stability of CT radiomics across different volume of interest sizes using parametric feature maps: a phantom study. *European Radiology Experimental* **2022**, *6*, 43.
91. Scalco, E.; Belfatto, A.; Mastropietro, A.; Rancati, T.; Avuzzi, B.; Messina, A.; Valdaghi, R.; Rizzo, G. T2w-MRI signal normalization affects radiomics features reproducibility. *Medical physics* **2020**, *47*, 1680–1691.
92. Tietz, E.; Truhn, D.; Müller-Franzes, G.; Berres, M.L.; Hamesch, K.; Lang, S.A.; Kuhl, C.K.; Bruners, P.; Schulze-Hagen, M. A radiomics approach to predict the emergence of new hepatocellular carcinoma in computed tomography for high-risk patients with liver cirrhosis. *Diagnostics* **2021**, *11*, 1650.
93. Shin, J.; Lim, J.S.; Huh, Y.M.; Kim, J.H.; Hyung, W.J.; Chung, J.J.; Han, K.; Kim, S. A radiomics-based model for predicting prognosis of locally advanced gastric cancer in the preoperative setting. *Scientific reports* **2021**, *11*, 1879.
94. Bernatowicz, K.; Grussu, F.; Ligero, M.; Garcia, A.; Delgado, E.; Perez-Lopez, R. Robust imaging habitat computation using voxel-wise radiomics features. *Scientific reports* **2021**, *11*, 20133.
95. Choi, W.; Liu, C.J.; Alam, S.R.; Oh, J.H.; Vaghjiani, R.; Humm, J.; Weber, W.; Adusumilli, P.S.; Deasy, J.O.; Lu, W. Preoperative 18F-FDG PET/CT and CT radiomics for identifying aggressive histopathological subtypes in early stage lung adenocarcinoma. *Computational and Structural Biotechnology Journal* **2023**, *21*, 5601–5608.
96. Lee, J.; Yoo, S.K.; Kim, K.; Lee, B.M.; Park, V.Y.; Kim, J.S.; Kim, Y.B. Machine learning-based radiomics models for prediction of locoregional recurrence in patients with breast cancer. *Oncology Letters* **2023**, *26*, 1–10.
97. Chen, Q.; Wang, L.; Wang, L.; Deng, Z.; Zhang, J.; Zhu, Y. Glioma grade prediction using wavelet scattering-based radiomics. *IEEE Access* **2020**, *8*, 106564–106575.
98. Meijer, K. Accuracy and stability of radiomic features for characterising tumour heterogeneity using multimodality imaging: a phantom study. Master's thesis, University of Twente, 2019.
99. Ericsson-Szecsényi, R.; Zhang, G.; Redler, G.; Feygelman, V.; Rosenberg, S.; Latifi, K.; Ceberg, C.; Moros, E.G. Robustness assessment of images from a 0.35 T scanner of an integrated MRI-Linac: characterization of radiomics features in phantom and patient data. *Technology in Cancer Research & Treatment* **2022**, *21*, 15330338221099113.
100. Fernandes, C.D.; Dinh, C.V.; Walraven, I.; Heijmink, S.W.; Smolic, M.; van Griethuysen, J.J.; Simões, R.; Losnegård, A.; van der Poel, H.G.; Pos, F.J.; et al. Biochemical recurrence prediction after radiotherapy for prostate cancer with T2w magnetic resonance imaging radiomic features. *Physics and imaging in radiation oncology* **2018**, *7*, 9–15.
101. Li, Y.; Huang, X.; Xia, Y.; Long, L. Value of radiomics in differential diagnosis of chromophobe renal cell carcinoma and renal oncocytoma. *Abdominal Radiology* **2020**, *45*, 3193–3201.
102. Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics* **2006**, *7*, 1–8.
103. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PloS one* **2019**, *14*, e0224365.
104. Cawley, G.C.; Talbot, N.L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research* **2010**, *11*, 2079–2107.
105. Gupta, N. DNA extraction and polymerase chain reaction. *Journal of cytology* **2019**, *36*, 116–117.
106. Maxwell® RSC DNA FFPE Kit Technical Manual. <https://www.promega.co.uk/resources/protocols/technical-manuals/101/maxwell-rsc-dna-ffpe-kit-protocol/> accessed (27 April 2023), 2022.
107. Maxwell® RSC Genomic DNA Kit Technical Manual. <https://www.promega.co.uk/resources/protocols/technical-manuals/500/maxwell-rsc-genomic-dna-kit-protocol/> accessed (27 April 2023), 2022.

108. InfiniumTM CytoSNP-850K v1.4 BeadChip Data Sheet. <https://support.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-cytosnp850k-data-sheet-m-gl-01507/infinium-cytosnp850k-data-sheet-m-gl-01507.pdf> accessed (27 April 2023), 2023.
109. iScan System Guide. <https://support-docs.illumina.com/ARR/iScan/Content/ARR/FrontPages/iscan.htm> accessed (27 April 2023), 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.