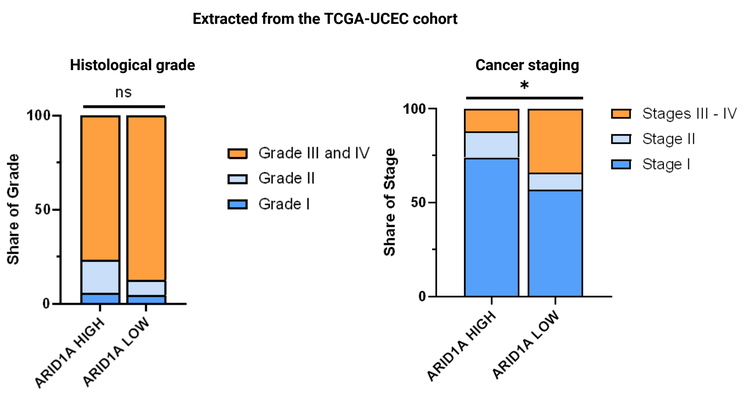
Computational dissection of the extracellular matrix in Endometrial endometrioid cancer.

Albertí-Valls M[1,3], Olave S.[1,3,6], Maiques O.[4,5], Matias-Guiu X.[2,3,6], Eritja N[1,3,6]

1.IRBLLEIDA, 2. Hospital Universitari Arnau de Vilanova, 3. Universitat de Lleida, 4. Barts Cancer Institute, 5. Queen Mary University of London, 6. Centro de Investigación en Red de Cancer (CIBERonc)

INTRODUCTION + RATIONALE

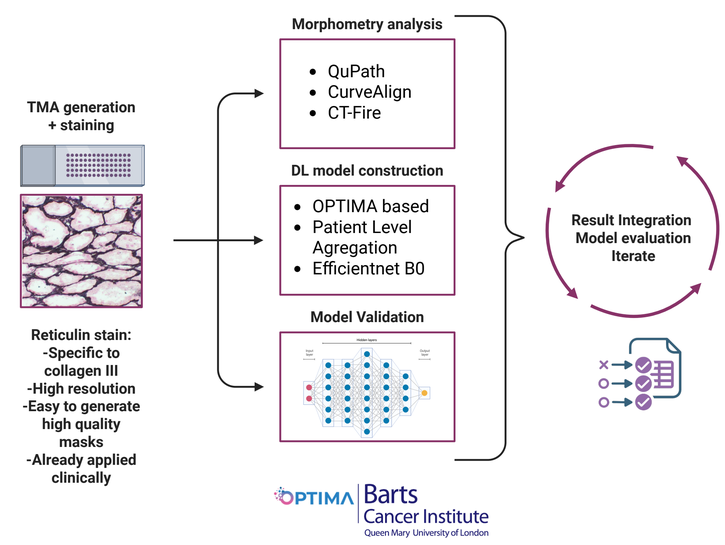
ARID1A, a core subunit of the SWI/SNF chromatin remodeling complex, is one of the **most frequently mutated genes in endometrial and ovarian cancers**. Its loss **disrupts epigenomic homeostasis,** impairing enhancer-mediated transcriptional regulation and tumor suppressor pathways (Mathur *et al.*, 2017). This epigenetic imbalance **promotes epithelial–mesenchymal transition and metastasis** (Zhang *et al.*, 2020), while also driving genomic instability and immune evasion (Wilson *et al.*, 2023). Collectively, ARID1A deficiency is **consistently linked to poor prognosis and aggressive tumor behavior** *(see figure below).*



The importance of the **tumour stromal compartment** in cradling tumour growth and facilitating cell dissemination has long been recognised. Among the underlying mechanisms, the cancer cell-mediated **paracrine activation** of **cancer-associated fibroblast**s plays a central role by driving extracellular matrix remodelling (Megino-Luque *et al.* Under Revision) Yet, the **specific architectural changes** induced in the tumor ECM have only recently begun to be disentangled and in a limited number of cancers. Crucially, these advances are now fuelling the **development of innovative prognostic and diagnostic tools** with **proven clinical potential** to improve patient outcomes in the near future.

METHODS

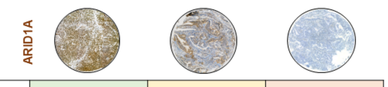
Systematic development of **ECM-based signatures** must take the lead in shaping future translational strategies. In this context, **deep learning** approaches offer a unique advantage: they can **harness large-scale data** with **strong statistical power, enabling the rapid generation of robust predictive models.** Moreover, by allowing automated integration into clinical workflows, **DL-based methods represent a powerful catalyst** for translating stromal and ECM insights into **tangible prognostic and diagnostic tools.**



We aimed to model the influence of **ARID1A expression on the extracellular matrix (ECM)** in endometrial cancer. To achieve this, we: **1.** Generated a **tissue microarray (TMA) from 180 patient samples** and stained it with **Reticulin**, a commonly used **Collagen III** stain in clinical settings. **2.** Hypothesis driven **ECM dissection** using tools like **QuPath and CurveAlign** to elucidate potential key morphometrical changes that could contextualize a DL model's performance **3**. Trained a **OPTIMA-based** EfficientNet-B0 CNN architecture in a **semi-supervised framework** to try and model previously seen characteristics into a **usable diagnostic tool.**

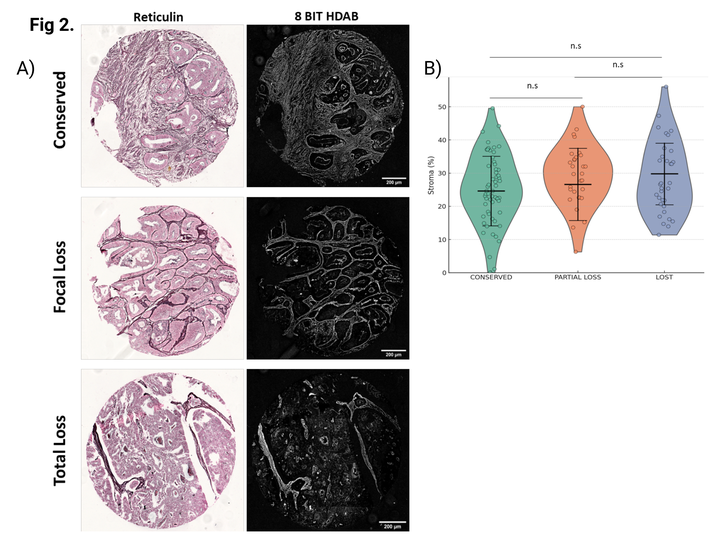
RESULTS

Fig 1.



		ARID1A Conserved	ARID1A Focal Loss	ARID1A Total Loss
Number (n)		36	21	21
Age; mean (range) years		68 (41-93)	67 (46-89)	67 (45-88)
Histological Sub-type, n (%)		Endometrioid, (100%)	Endometrioid, (100%)	Endometrioid, (100%)
Histologic differentiation, n (%)	Grade 1,	18 (50%)	8 (40%)	9 (42,8%)
	Grade 2,	11 (30%)	10 (47,6%)	9 (42,8%)
	Grade 3,	7 (20%)	3 (14,21%)	3 (14,21%)
Molecular sub-type	POLE	1 (2,77%)	0	1 (4,75%)
	MSI	10 (27,78%)	10 (47%)	11 (53%)
	P53(Abn)	2 (5,56%)	0	0
	NSMP	23 (63,9%)	11 (53%)	9 (42%)

Figure 1. Clinicopathological characteristics of the EEC TMA cohort stratified by ARID1A status.
Tissue microarrays (TMAs) comprising 78 endometrioid endometrial carcinoma (EEC) cases were analyzed and classified according to ARID1A expression into conserved (n=36), focal loss (n=21), and global loss (n=21). All tumors were of endometrioid histology. Patient age was comparable across groups (mean 67–68 years). Histologic differentiation showed a balanced distribution of Grade 1–3 tumors in all groups. Molecular subtyping revealed a predominance of NSMP and MSI cases, with rare POLE and p53-abnormal subtypes. Representative ARID1A IHC stains are shown for each category.



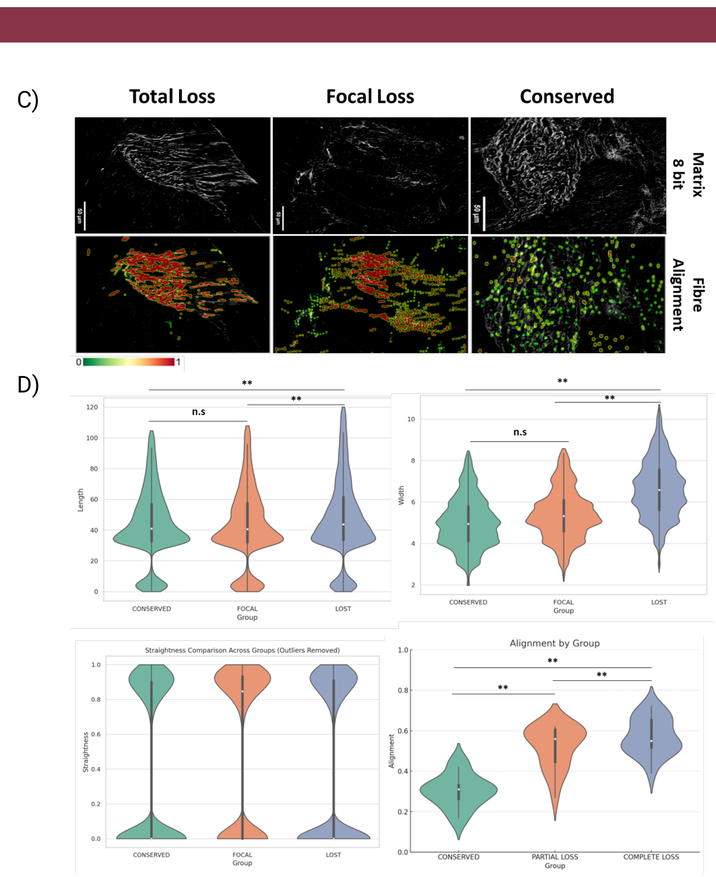


Fig. 2. ARID1A-deficient tumors show distinct ECM architecture with significantly wider, longer, and more aligned fibers, independent of stromal fraction. A) Representative reticulin stains with 8-bit masks. B) QuPath-based stromal quantification (no group differences, one-way ANOVA). C) Representative 8-bit masks with intensity maps. D) Significant differences in fiber length, width, and alignment ($p<0.05$ one-way ANOVA). **These results conform with those reported by Maiques O. et al. (2025), which demonstrated that specific ECM architectural patterns facilitate cell escape.**

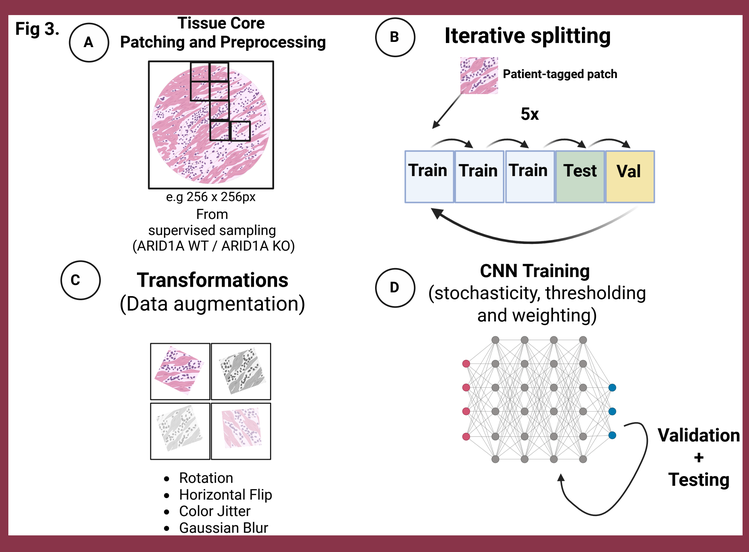


Fig 3. Workflow for model construction. A) A QuPath script was used to select and annotate tumor regions at the patient level from TMA cores, which were then patched into 256×256 tiles. B) Tiles were split into training, validation, and testing sets using a **5-fold cross-validation strategy** to ensure patient-level independence. C) Data augmentation was applied, including rotations, flips, color jitter, and Gaussian blur. D) An EfficientNet-B0 convolutional neural network (pre-trained on ImageNet) was fine-tuned using optimized hyperparameters (cross-entropy loss, learning rate scheduling, adaptive optimizer, and early stopping). The final model was evaluated through cross-validation and independent testing.

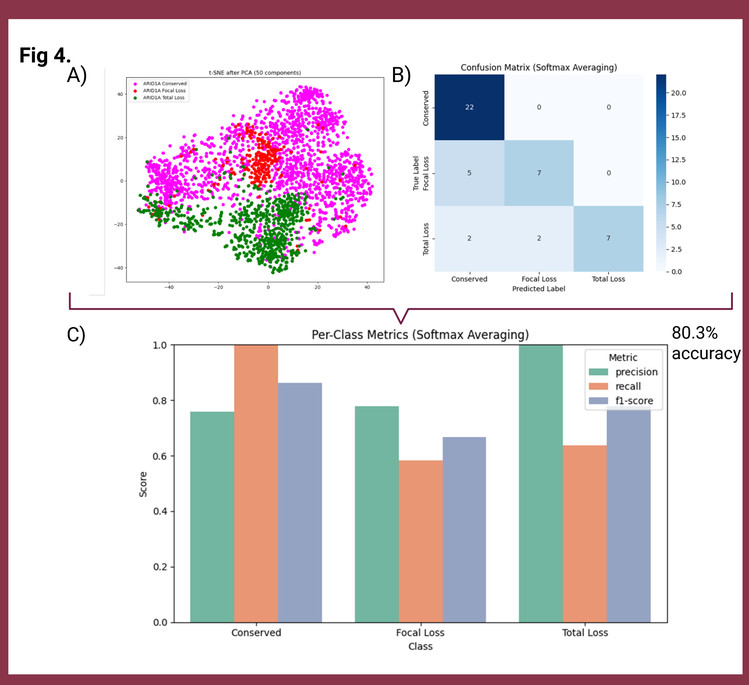


Fig. 4. Reticulin PLA model performance. A) t-SNE plot after PCA showing separation of samples by ARID1A status (conserved, focal loss, total loss) based on ECM features. B) Confusion matrix illustrating model predictions versus true labels across the three ARID1A categories. C) Per-class performance metrics (precision, recall, and F1 score) for the testing set, highlighting stronger discrimination of conserved cases compared to focal and total loss. Overall classification accuracy reached 80.3%.

CONCLUSIONS

- We observed **distinct ECM architectural differences** associated with ARID1A status.
- The current model's accuracy falls **below clinical approval thresholds** (sensitivity/specificity >95%).
- **Limitations:** Struggles with focal loss, reflecting its intermediate/transitional histology. Model bias toward classifying patches as conserved.
- **Future directions:** Incorporating whole-slide image or core-level analysis. Applying multiple-instance learning approaches. Better capturing spatial heterogeneity of histological features.

M.A-V holds a the predoctoral fellowship "AJUTS 2022 UdL PER A LA CONTRACTACIÓ DE PERSONAL PREDOCTORAL EN FORMACIÓ" from Universitat de Lleida.