Supplemental information

Synthetic whole-slide image tile

generation with gene expression profile-infused

deep generative models

Francisco Carrillo-Perez, Marija Pizurica, Michael G. Ozawa, Hannes Vogel, Robert B. West, Christina S. Kong, Luis Javier Herrera, Jeanne Shen, and Olivier Gevaert

List of Supplemental Figures

Figure S1: Synthetic samples generated by the GAN and the RNA-GAN architecture, related to Figures 3 and 4. Panel A: More examples of brain cortex and lung tissue generation using the GAN architecture. Panel B: More examples of brain cortex and lung tissue generation using the RNA-GAN architecture.

Figure S2: Google Form presented to pathologists, related to Figure 6. Example of one of the samples of the form presented to expert pathologists. Three question were asked, regarding the tissue of origin, the quality of the morphological structure and the presence of image artifacts.

Figure S3: RMSE obtained by the VAE per tissue, related to Figure 2. Violin plots of the RMSE obtained across tissues for all the samples in the test set, during the βVAE training.

Figure S4. Classification problem between glioblastoma (GBM) and lung adenocarcinoma (LUAD) with and without self-supervised pretraining, related to Figure 4. Panel A: Accuracy and F1-Score obtained across the 5-Fold CV for the model trained from scratch and the model using self-supervised learning weights. Panel B: Confusion matrices for the model trained from scratch and the model using self-supervised learning weights across all the test sets in the stratified 5-Fold CV
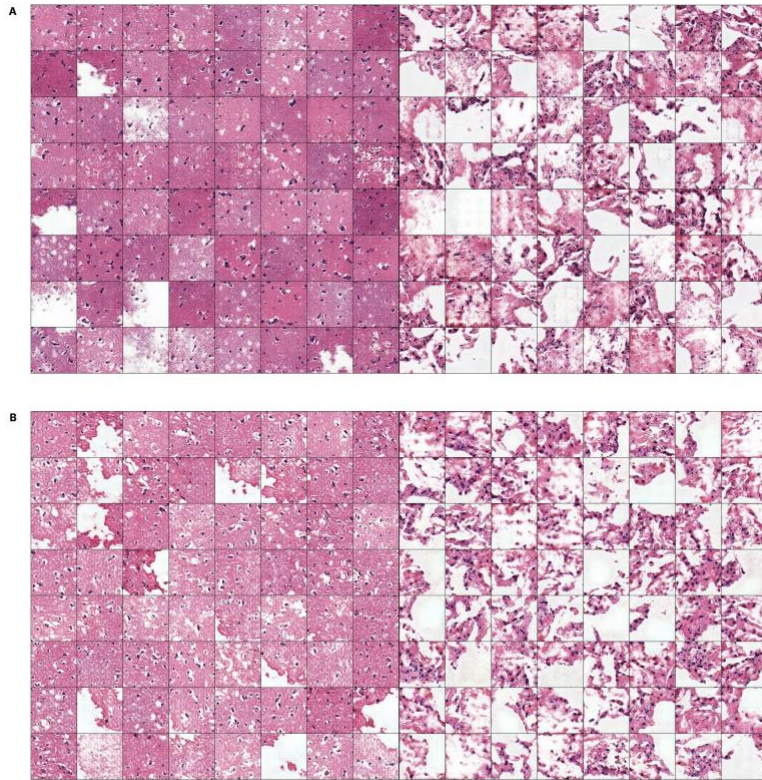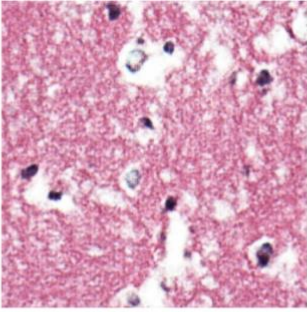
Figure S1: Synthetic samples generated by the GAN and the RNA-GAN architecture, related to Figures 3 and 4. Panel A: More examples of brain cortex and lung tissue generation using the GAN architecture. **Panel B:** More examples of brain cortex and lung tissue generation using the RNA-GAN architecture.

Figure S2: Google Form presented to pathologists, related to Figure 6. Example of one of the samples of the form presented to expert pathologists. Three question were asked, regarding the tissue of origin, the quality of the morphological structure and the presence of image artifacts.
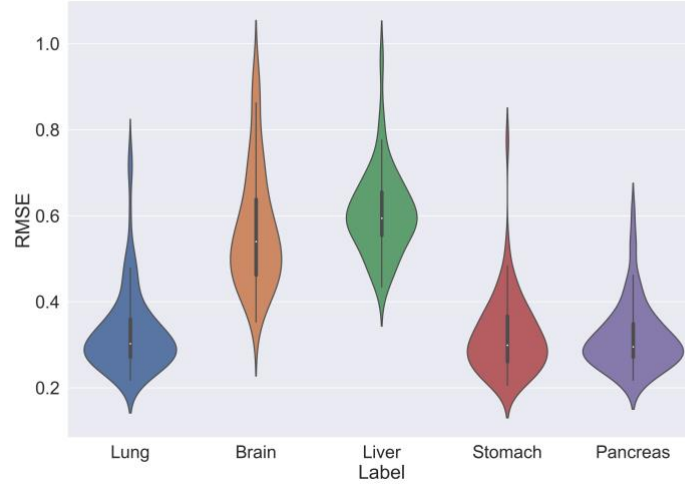
Figure S3: RMSE obtained by the VAE per tissue, related to Figure 2. Violin plots of the RMSE obtained across tissues for all the samples in the test set, during the βVAE training.
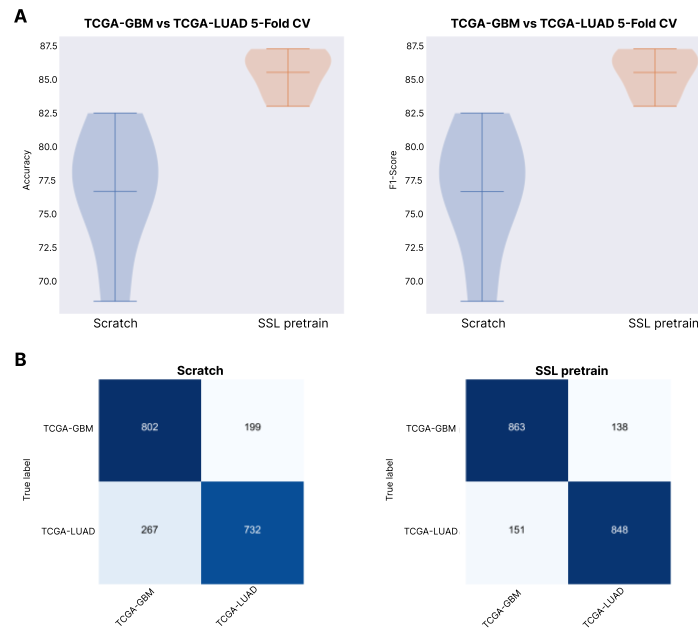

Figure S4: Classification problem between glioblastoma (GBM) and lung adenocarcinoma (LUAD) with and without self-supervised pretraining, related to Figure 4. **Panel A:** Accuracy and F1-Score obtained across the 5-Fold CV for the model trained from scratch and the model using self-supervised learning weights. **Panel B:** Confusion matrices for the model trained from scratch and the model using self-supervised learning weights across all the test sets in the stratified 5-Fold CV.