

# Fibrosis stage – reference images

The following images are provided with the aim of improving inter-observer agreement in the assessment of fibrosis stage in liver biopsies, especially among non-specialist pathologists.

They provide a ‘bench mark’ of liver biopsy staging for routine diagnostic practice.

They use a scale of none/early/bridging/late stage fibrosis.

These stages are clinically important in patient management, and can be recognised in chronic liver disease of any or mixed aetiologies.

The images are of the good sized biopsies stained by Sirius Red which had either complete agreement by 8 UK liver pathologists, or had a 4:4 split. The latter represent the ‘threshold’ between adjacent stages.

Use of these images improved agreement among both the specialist liver pathologists and among a group of generalist and trainee pathologists.

# Can the Inter-Observer Agreement of Reporting Fibrosis in Medical Liver Core Biopsies be Improved with the Use of Reference Images?

## A Study Comparing Liver Subspecialists and General Histopathologists.

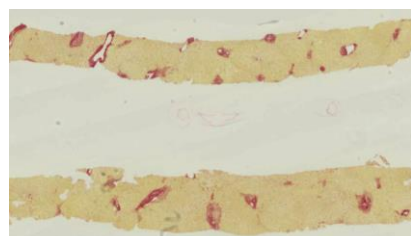
German, A.L.<sup>1</sup>; Fleming, K.A.<sup>2</sup>; Kaye, P.V.<sup>3</sup>; Davies, S.E.<sup>4</sup>; Goldin, R.<sup>5</sup>; Hubscher, S.G.<sup>6</sup>; Tiniakos, D.<sup>7</sup>; McGregor, A.<sup>8</sup>; Bellamy, C.<sup>9</sup>; Wyatt, J.I.<sup>1</sup>

<sup>1</sup> Leeds Teaching Hospitals NHS Trust, Leeds; <sup>2</sup> Green Templeton College, Oxford University, Oxford; <sup>3</sup> Nottingham University Teaching Hospitals NHS Trust, Nottingham; <sup>4</sup> Cambridge University Hospitals NHS Foundation Trust, Cambridge; <sup>5</sup> Imperial College Healthcare NHS Trust, London; <sup>6</sup> University Hospitals Birmingham NHS Foundation Trust, Birmingham; <sup>7</sup> Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle Upon Tyne; <sup>8</sup> The University Hospitals of Leicester NHS Trust, Leicester; <sup>9</sup> Edinburgh Royal Infirmary, Edinburgh.

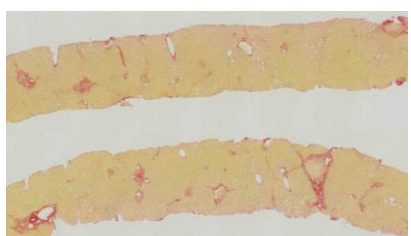
### Example Reference Images



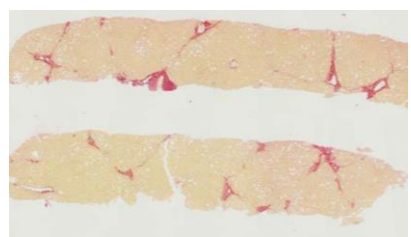
None/early



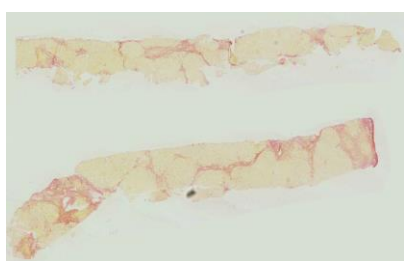
Early



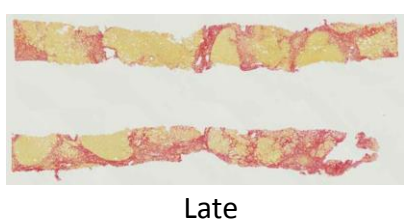
Early/Bridging



Bridging



Bridging/Late



Late

### Introduction

Liver biopsy plays an important part in assessing patients with a variety of liver diseases.

At present, a number of different scoring systems are used by histopathologists to report the level of fibrosis in liver biopsies. The system used is dependent on the underlying pathology, and as such can become complex and confusing for pathologists, especially if reporting biopsies infrequently, or if of a mixed aetiology<sup>1</sup>. There is also an inherent level of inter-observer variation between pathologists<sup>2</sup> and previous studies have provided Kappa scores for inter-observer agreement of between 0.5-0.9<sup>1</sup>.

This study was performed to assess if a set of reference images can improve the inter-observer agreement, in a simplified staging system of liver fibrosis, using four clinically relevant stages:

- None – No evidence of fibrosis
- Early – Any definite fibrosis (portal or sinusoidal) without bridging.
- Bridging – Bridging fibrosis, without nodules
- Late – Bridging fibrosis, and nodule formation

These stages are non-numeric and independent of the type of chronic liver disease.

### Method

The study was initially completed by eight subspecialist liver pathologists in two phases.

In phase one, the raters assessed the stage of liver fibrosis in 47 virtual slides stained with Sirius red, and a minimum biopsy length of 23mm, according to the four categories of none, early, bridging or late.

The responses were then collated, allowing identification of those which had either complete agreement, or a 50:50 split across stages. These were then provided as reference images for agreed and threshold examples of fibrosis stage. The purpose of the threshold examples is to provide a benchmark that separates the stages, and not considered as additional categories.

For phase two, a further 47 virtual slides were distributed, along with the reference images to use as a guide.

The slides were hosted on the Leeds University Virtual Pathology website.

General histopathology consultants who participated in the liver EQA scheme, and trainees who occasionally report liver pathology were then invited to complete both phases of the study, utilizing the same reference images for phase two.

The results were analysed for all participants, and separately for the subspecialists and generalists, to assess if any improvement was applicable to both cohorts. We identified the percentage agreement along with the the Fleiss' Kappa score and Krippendorff- $\alpha$  for phase one and two.

### Results

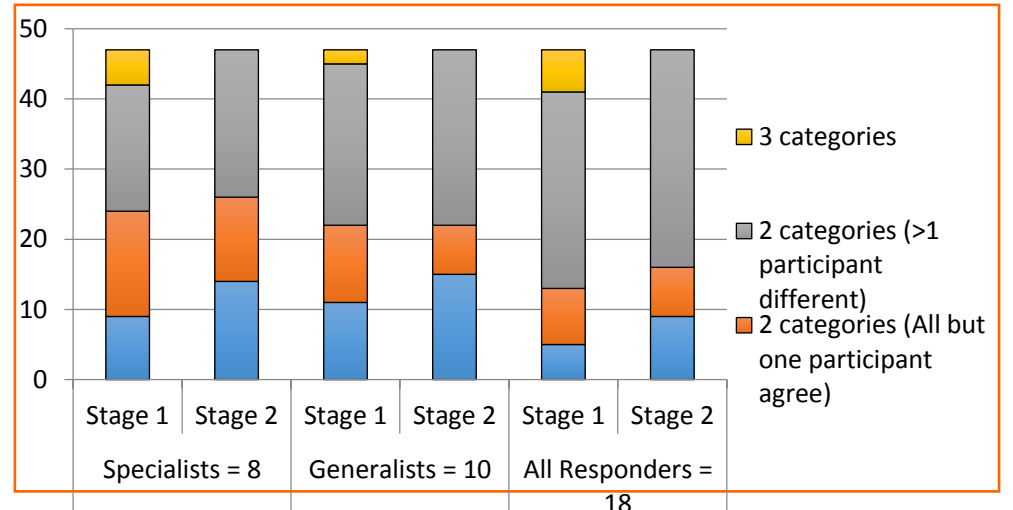
A total of eight subspecialist liver pathologists, and ten general pathologists or trainees completed both phases of the study (eighteen overall). For the general pathologists, the raters were reporting between 10 - >80 biopsies per year.

The results found a modest improvement in overall percentage agreement with the use of reference images from 67.79% to 70.08%, eliminated the spread over 3 different stages in phase 2, and an improved agreement across all four grades.

The Fleiss' Kappa and Krippendorff- $\alpha$  also confirm an overall improved agreement with the use of reference images. The Kappa score shows moderate agreement, and improved from 0.55 – 0.59 and the Krippendorff- $\alpha$  from 0.55 – 0.57.

For the subspecialist liver pathologists the results the percentage agreement improved from 65.35% to 68.84%, with the Kappa improving from 0.51 – 0.58, and Krippendorff- $\alpha$  from 0.51-0.54.

For the general pathologists and trainees, the percentage agreement improved from 69.65 to 70.97, with the Kappa improving from 0.58 – 0.60, and Krippendorff- $\alpha$  from 0.57 – 0.58.

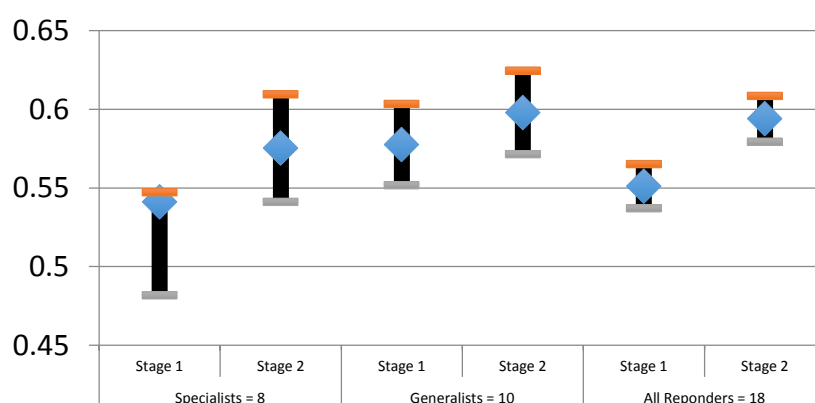


### Conclusions

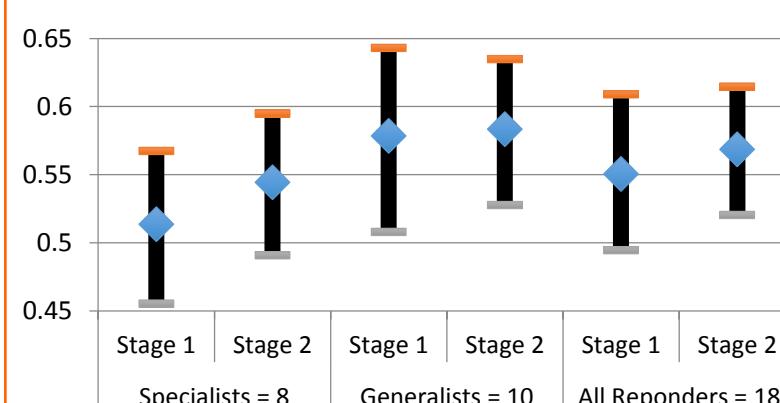
The staging of liver fibrosis has an inherent inter-observer variability. This study shows that the use of reference images when assessing the level of fibrosis, improves the level of percentage agreement across all categories, and eliminates the spread across three categories. This is seen in both subspecialist and general histopathologists

Whilst recognising that disease specific fibrosis staging systems, such as Ishak and Kleiner et al, have an important role in research, they are found to be poorly reproducible in everyday routine reporting<sup>2</sup>. We propose that the simpler 4-stage reporting, along with the use of the reference images, would improve the reproducibility and reliability of the score, amongst both generalist and specialist pathologists, which would provide our clinicians with a clinically and prognostically relevant result. The reference images would be available free online to provide an easy point of reference, and improve consistency in reporting.

### Fleiss' Kappa Score with 95% Confidence Intervals



### Krippendorff- $\alpha$ with 95% Confidence Intervals



### References

1. Goodman, Z.D. *Grading and staging systems for inflammation and fibrosis in chronic liver disease*. Journal of Hepatology; 47 (2007). 598-607.
2. Wyatt, J., Hubscher, S., Bellamy, C. *Tissue pathways for liver biopsies for the investigation of medical disease and focal lesions*. Royal College of Pathologists; March 2014. Accessed online.

# None/early

10.6759 N-E



# None/early

10.19499 N-E



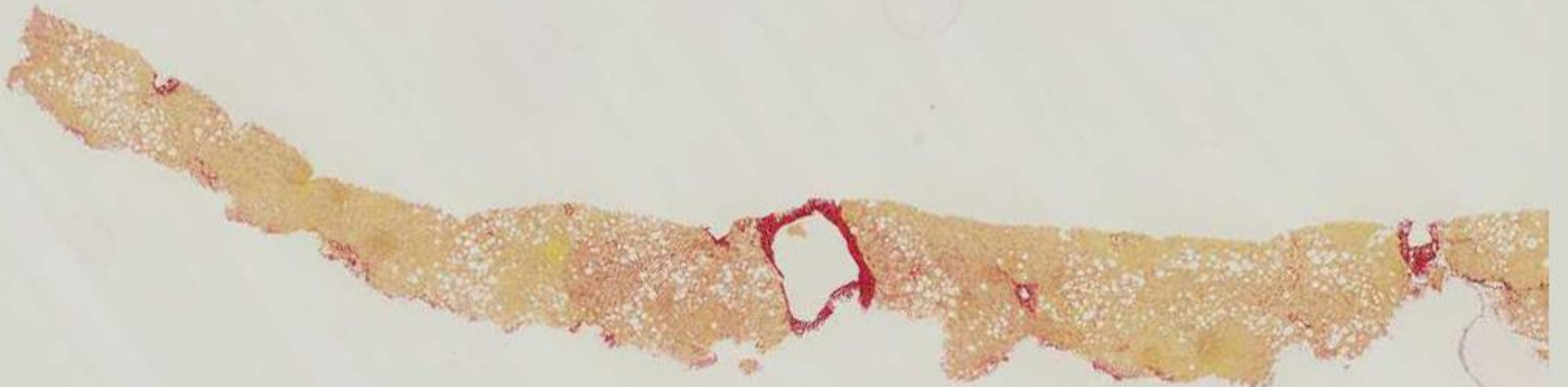
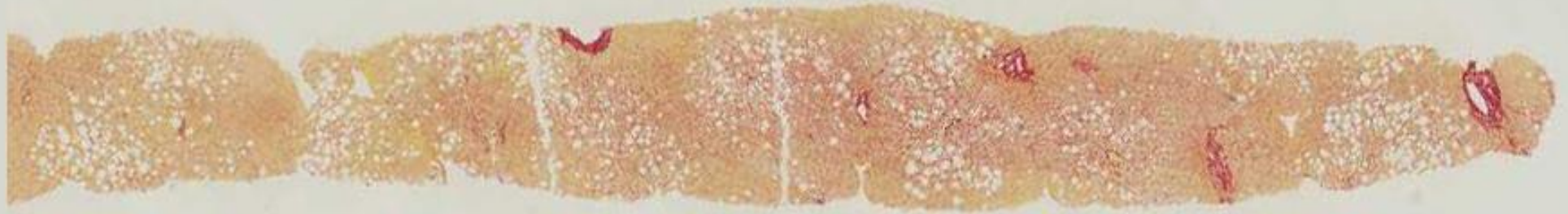
# None/early

11.41415 N-E



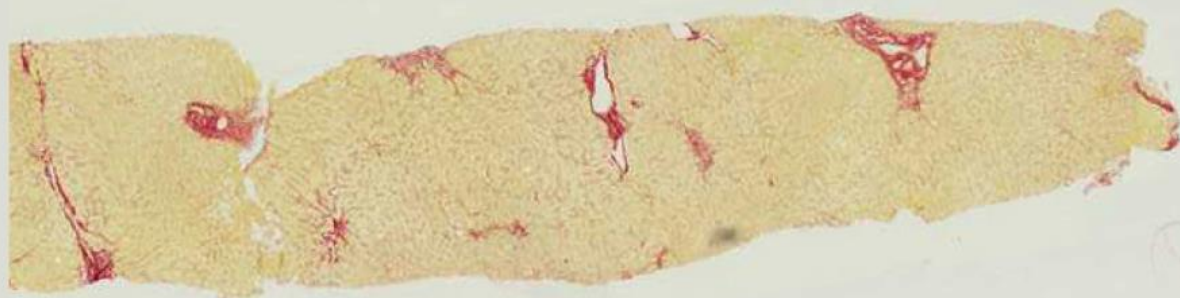
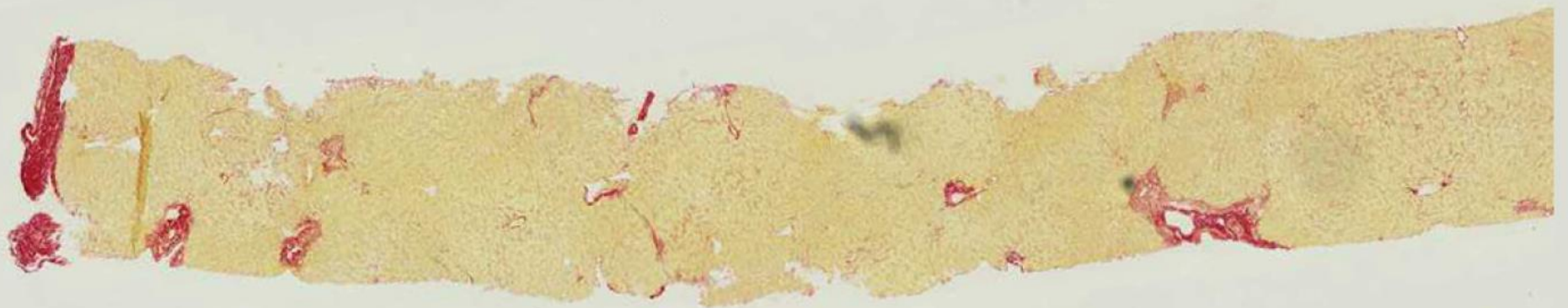
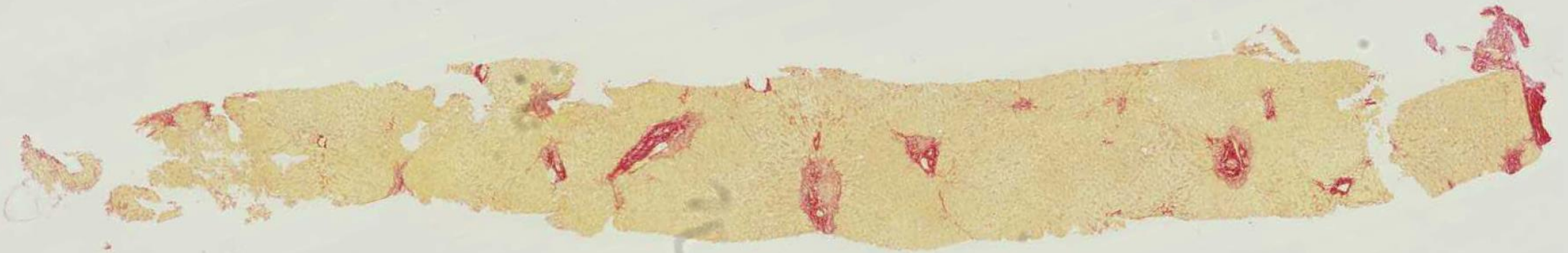
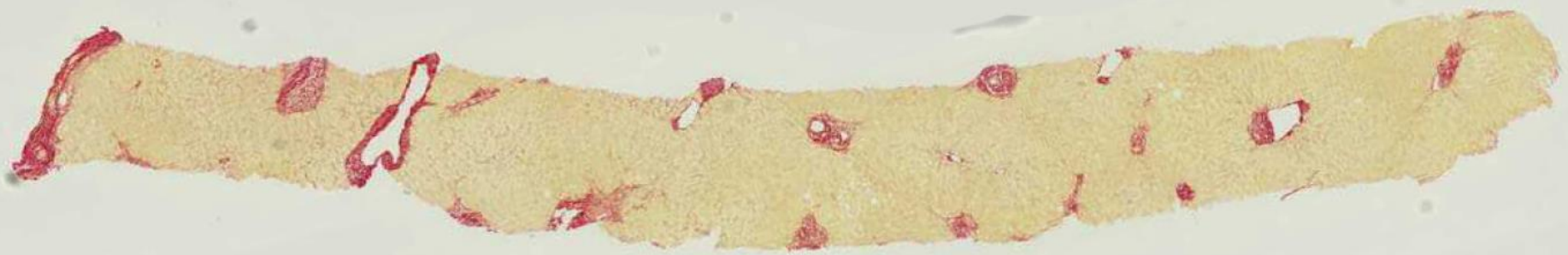
# Early

10.14246 E



# Early

12.12194 E



# Early

12.20680 E



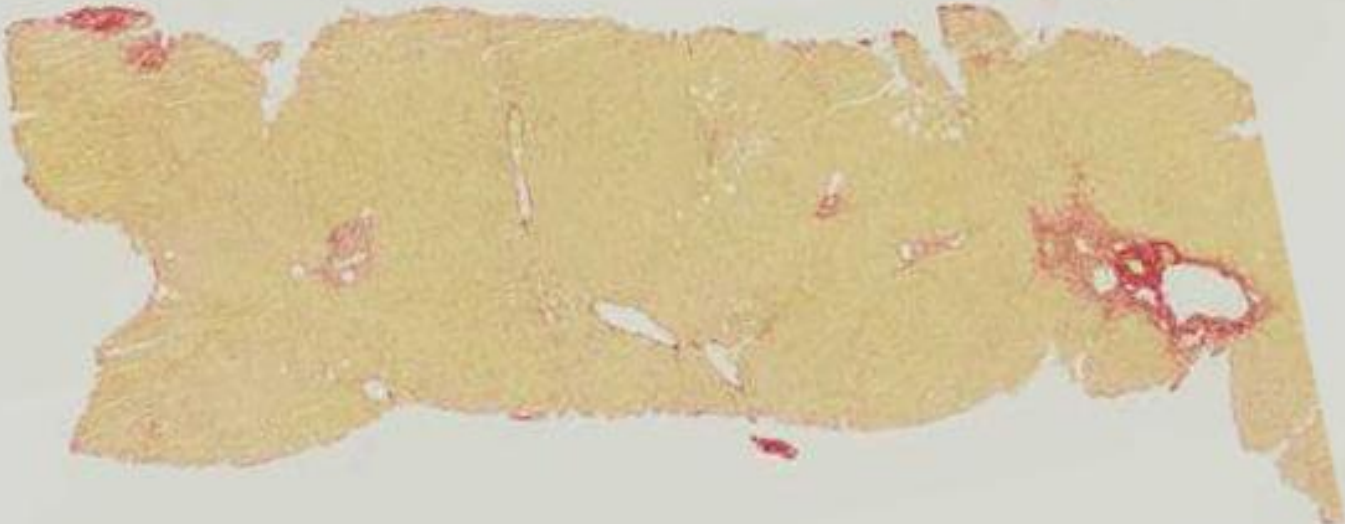
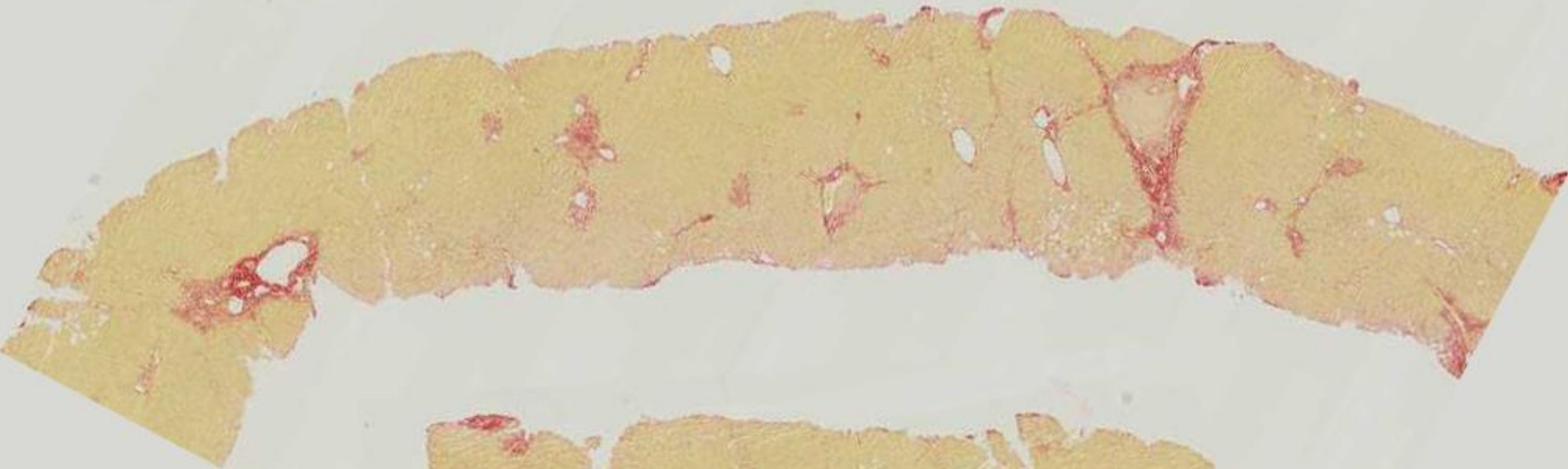
# Early/bridging

10-00799 E-B



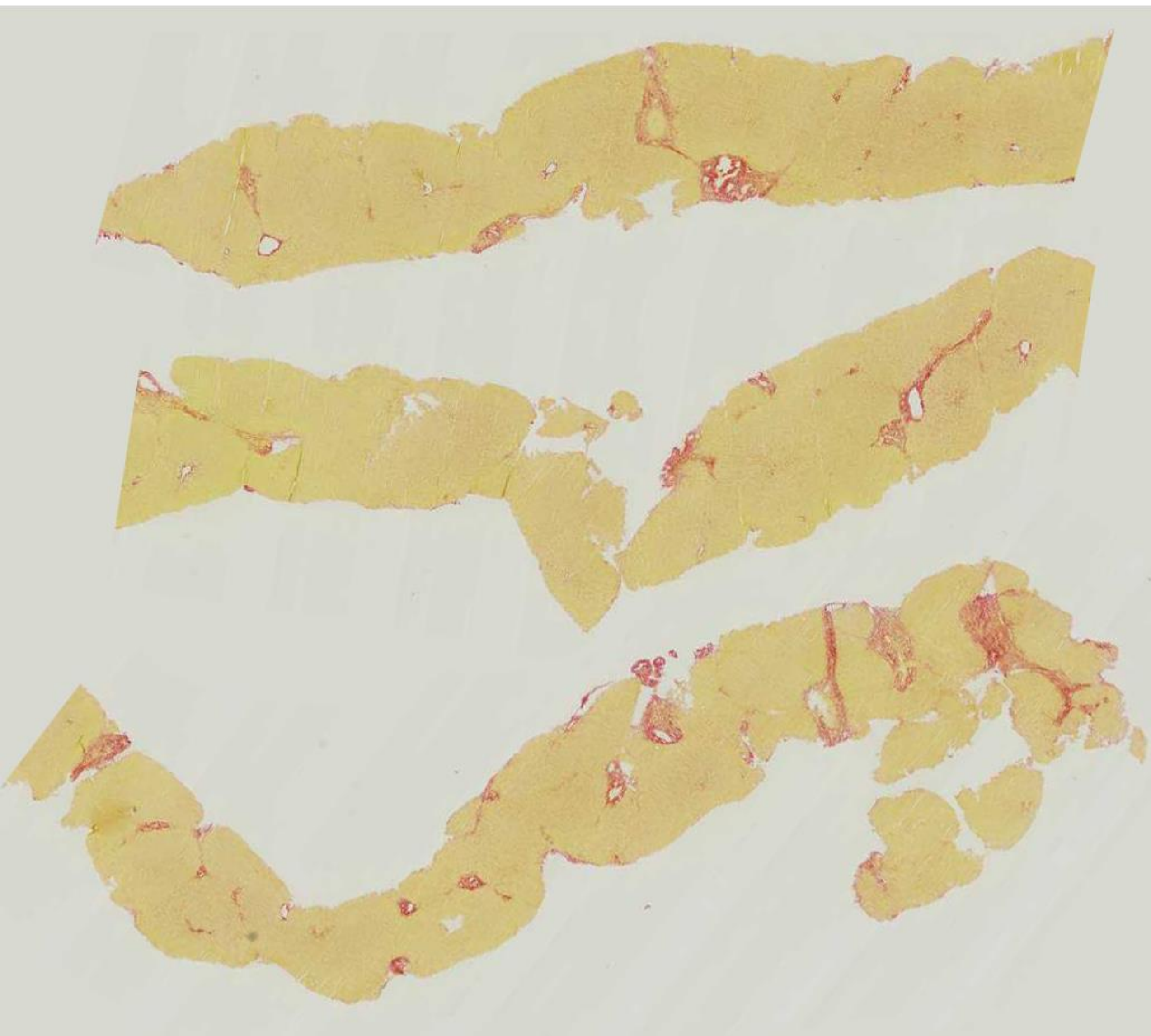
# Early/bridging

09-27945 E-B



# Bridging

10-19757 B



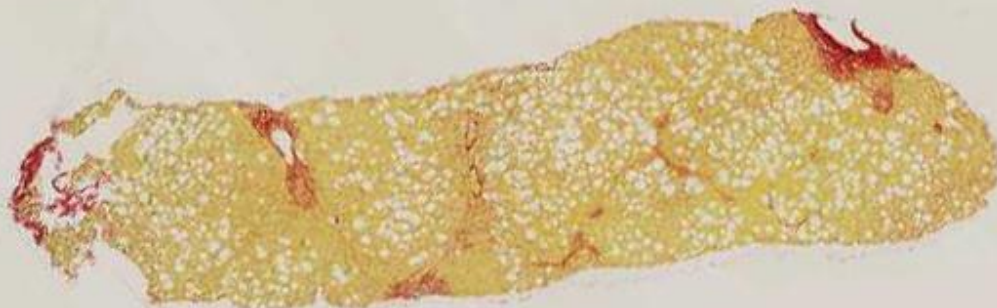
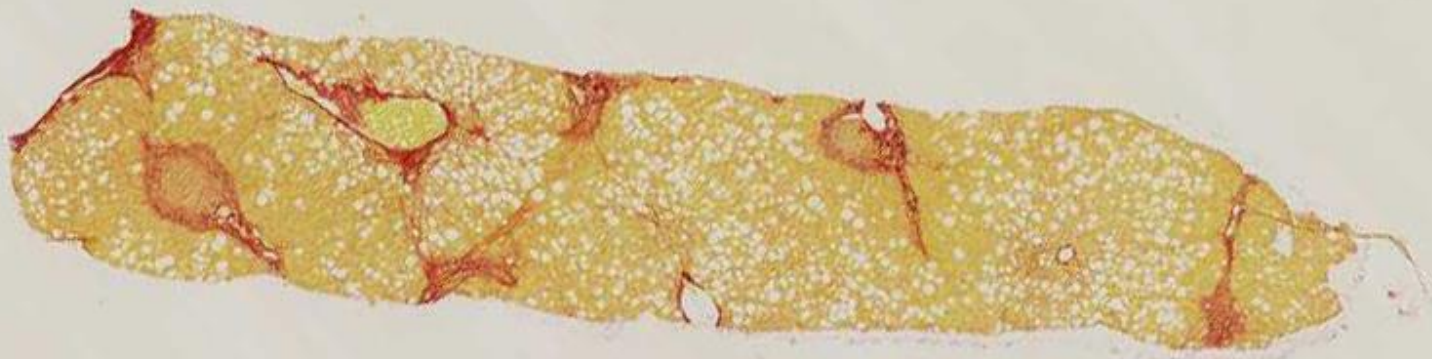
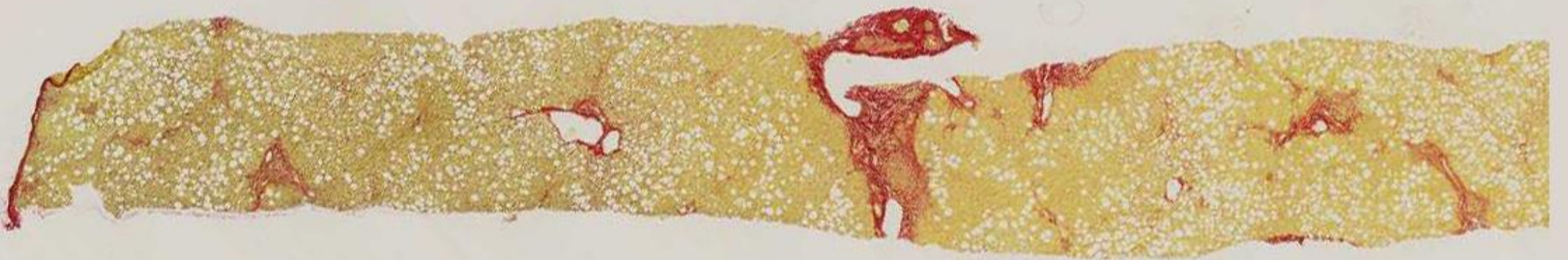
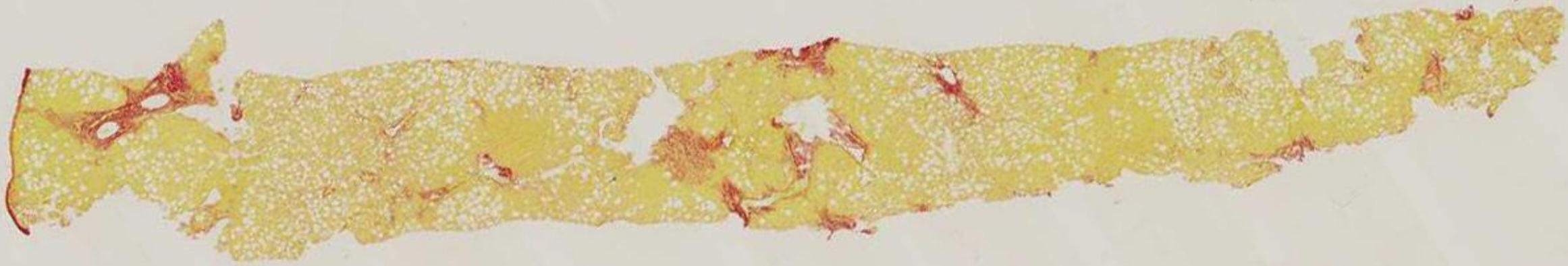
# Bridging

11.11266 B



# Bridging

10.18324 B



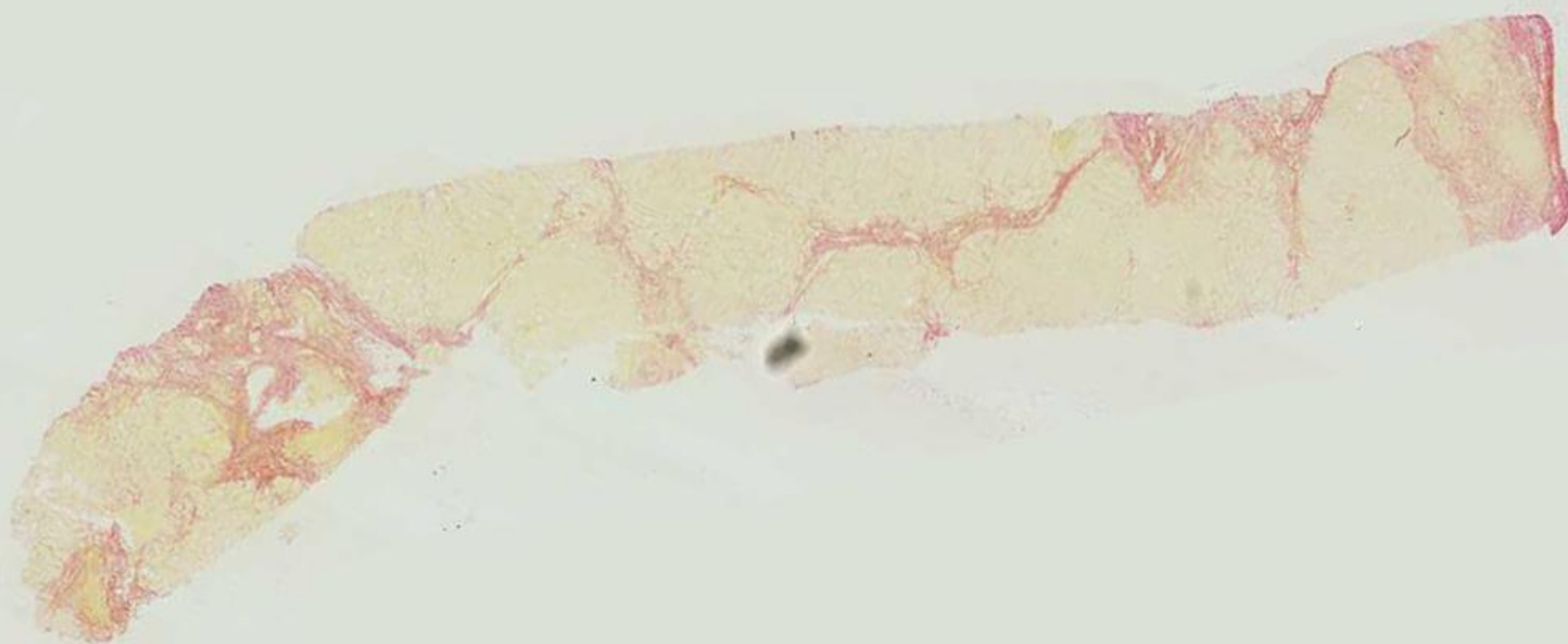
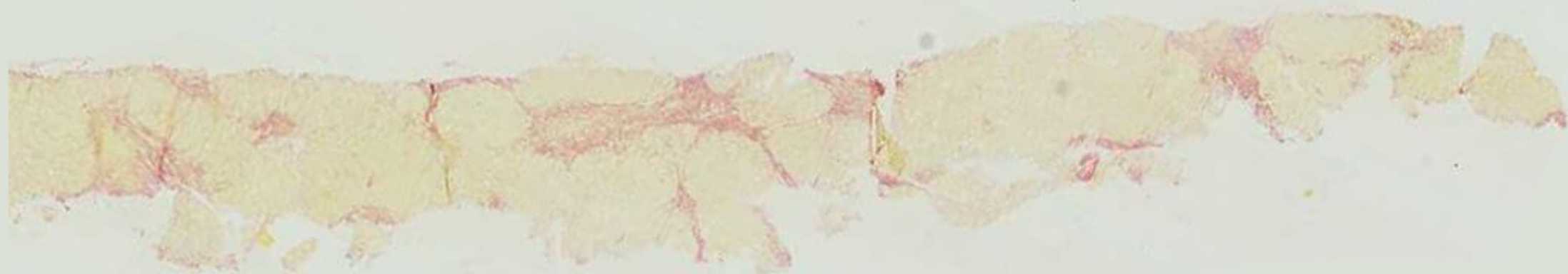
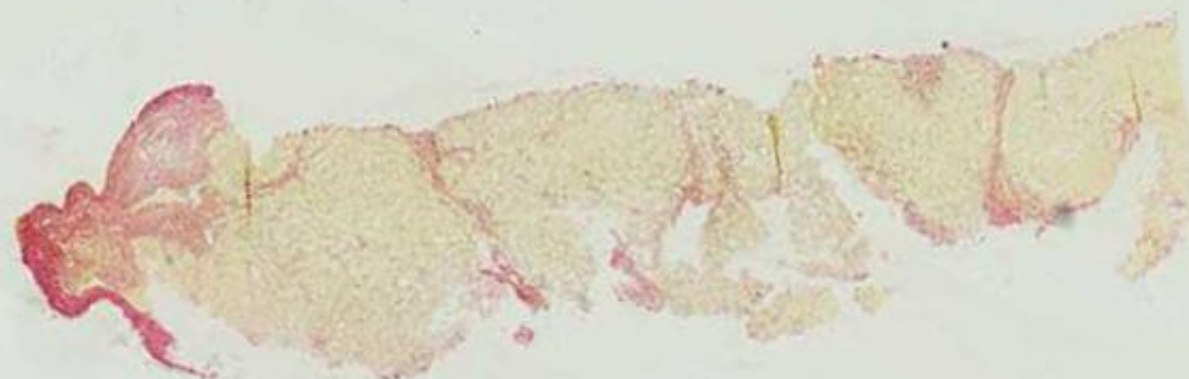
# Bridging/late

12.16976 B-L



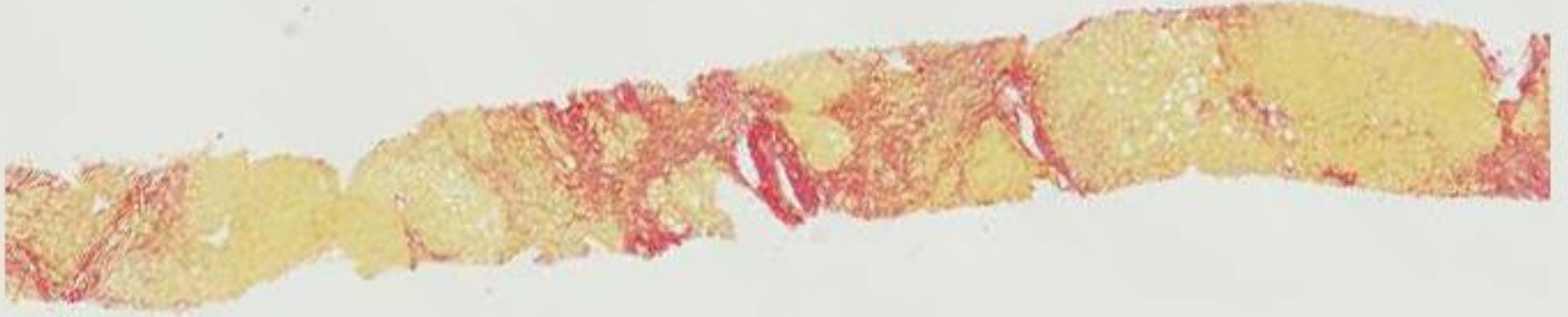
# Bridging/late

10.10729 B-L



Late

11.4053 L



# Late

11.23045 L

