

ICC analysis

The result of ICC analysis for Classic/Common-type descriptor is below:

##	type	ICC	F	df1	df2	p	lower bound
## Single_raters_absolute	ICC1	0.1466	3.062	8	99	0.0040162	0.04046
## Single_random_raters	ICC2	0.1587	3.839	8	88	0.0006504	0.05515
## Single_fixed_raters	ICC3	0.1913	3.839	8	88	0.0006504	0.06809
## Average_raters_absolute	ICC1k	0.6734	3.062	8	99	0.0040162	0.33600
## Average_random_raters	ICC2k	0.6937	3.839	8	88	0.0006504	0.41193
## Average_fixed_raters	ICC3k	0.7395	3.839	8	88	0.0006504	0.46716
##	upper bound						
## Single_raters_absolute		0.4033					
## Single_random_raters		0.4097					
## Single_fixed_raters		0.4653					
## Average_raters_absolute		0.8902					
## Average_random_raters		0.8928					
## Average_fixed_raters		0.9126					

The result of fitted random effects models is as follows:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: values ~ 1 + (1 | id) + (1 | items)
## Data: x.df
## REML criterion at convergence: 36.75
## Random effects:
## Groups Name Std.Dev.
## items (Intercept) 0.167
## id (Intercept) 0.161
## Residual 0.332
## Number of obs: 36, groups: items, 12; id, 9
## Fixed Effects:
## (Intercept)
## 0.812

## variance Percent
## ID 0.02607 0.1587
## Items 0.02795 0.1702
## Residual 0.11021 0.6711
## Total 0.16423 1.0000
```



Figure 1: Estimated ICC of presence/absence scoring of individual descriptors at the sample level with 95% CI. A descriptor is counted as present if a reviewer called it present in any grid square, and absent if they said no for all grid squares. Marked NA when the features were never present in any of 9 samples.

Gwet AC

Description of function: <https://www.rdocumentation.org/packages/irrCAC/versions/1.0/topics/gwet.ac1.raw>

```
gwet.ac1.raw(score_df)
```

```
## $est
##   coeff.name   pa   pe coeff.val coeff.se  conf.int p.value    w.name
## 1         AC1 0.737 0.29   0.6297   0.1962 (0.177,1) 0.01244 unweighted
##
## $weights
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
##
## $categories
## [1] 0 1
```

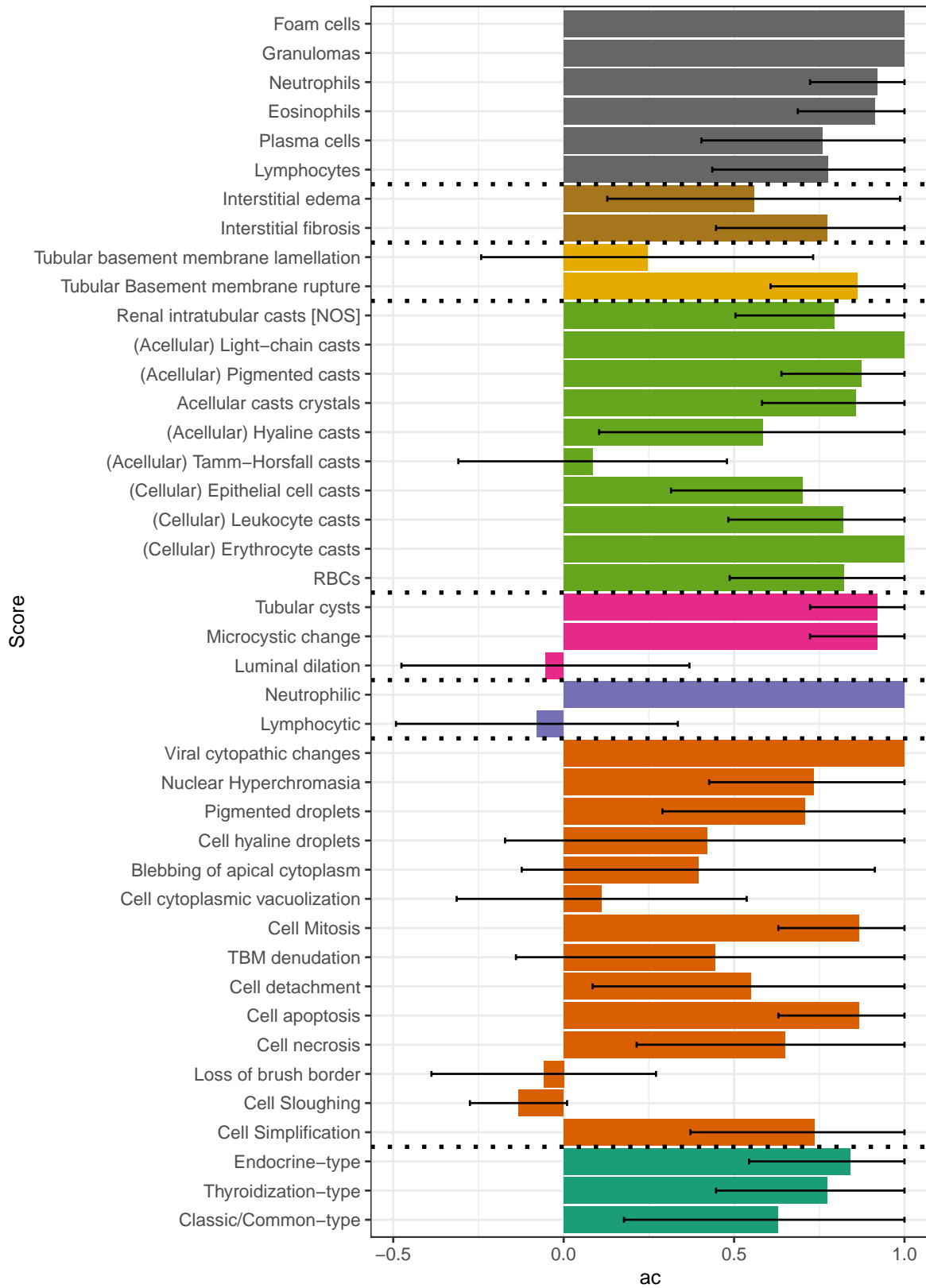


Figure 2: Estimated Gwet AC1 of individual descriptors with 95% CI. No CI was presented when there is no variation.

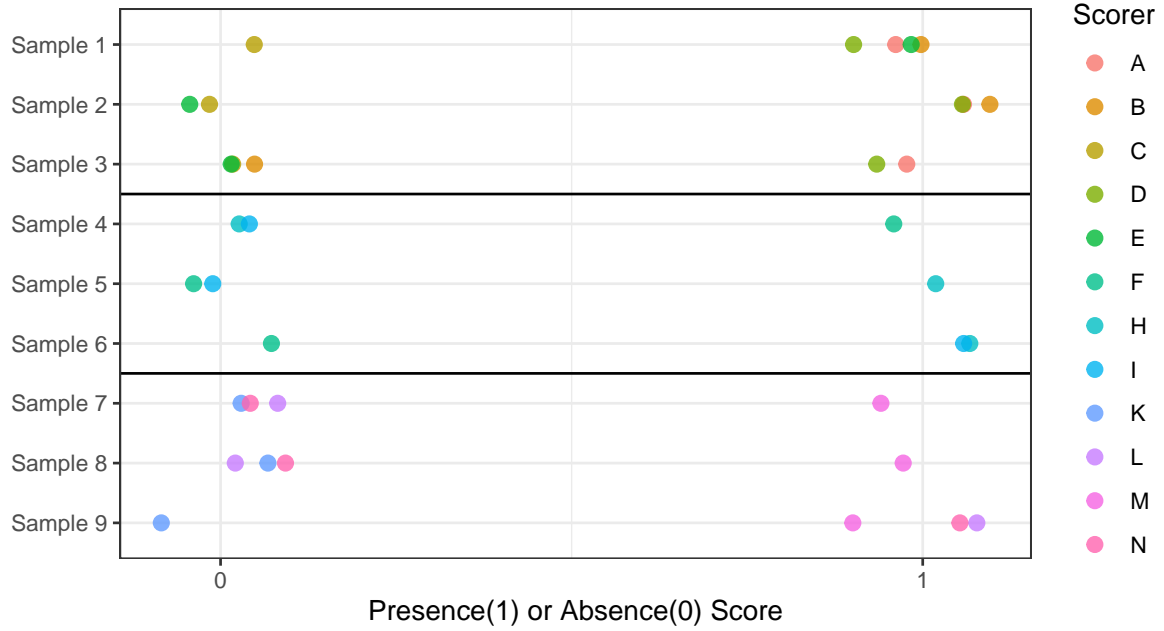
Supplementary material

Scoring

In the above analyses, presence scoring of each descriptor is as follows. A descriptor is counted as present if a reviewer called it present in any grid square, and absent if they said no for all grid squares.

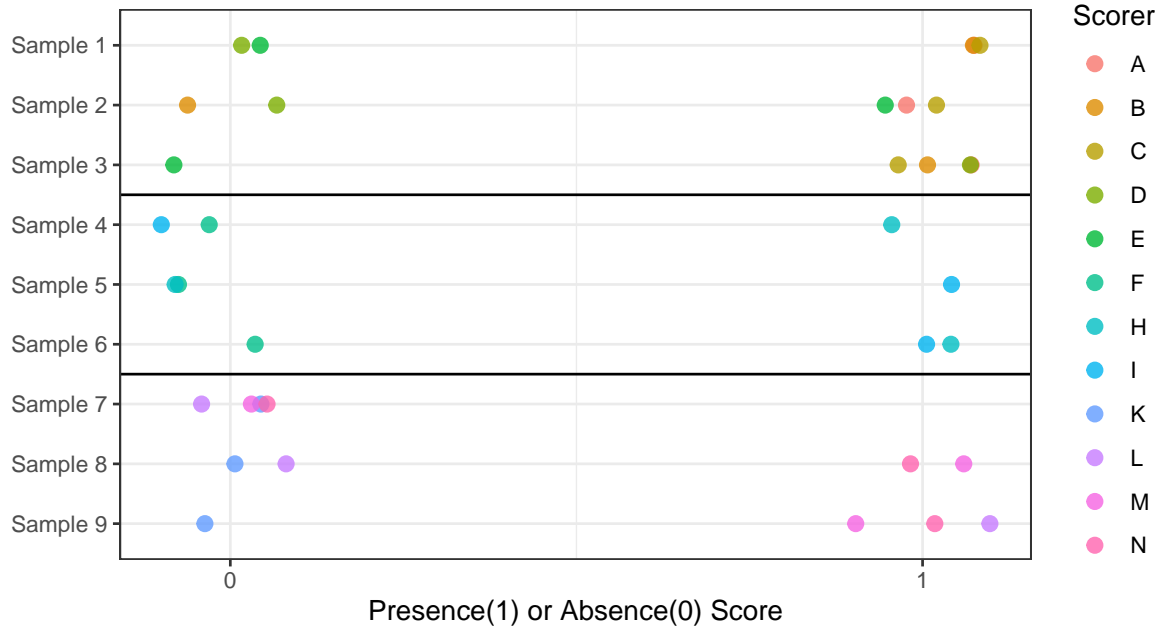
Cell Sloughing

Domain: Tubular Cell Injury (other than atrophy)



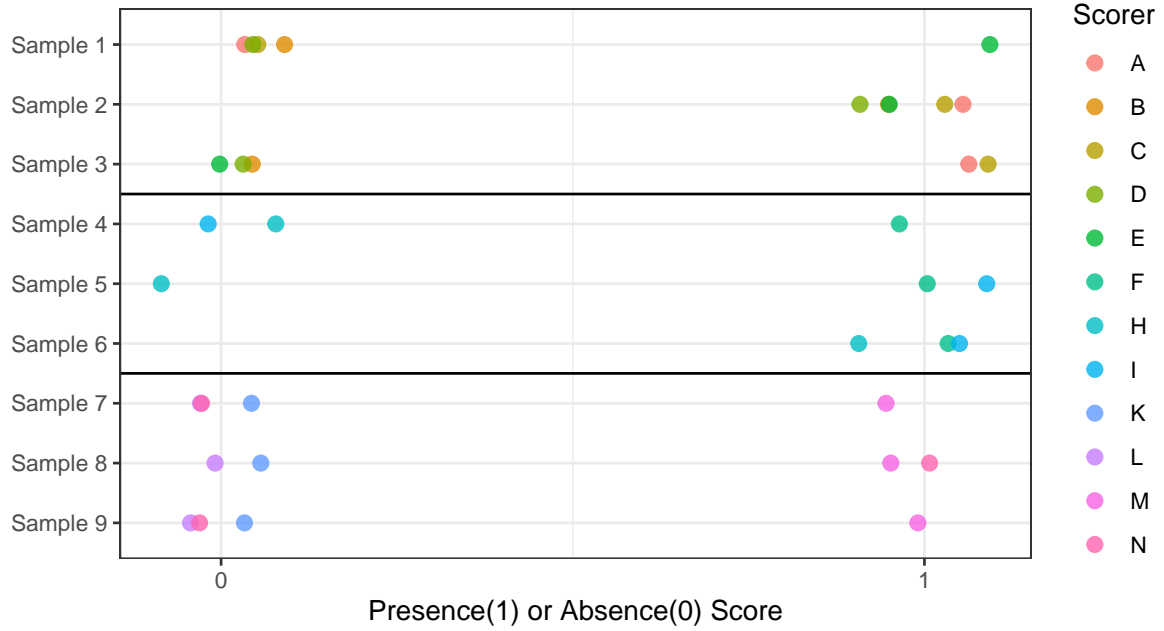
Loss of brush border

Domain: Tubular Cell Injury (other than atrophy)



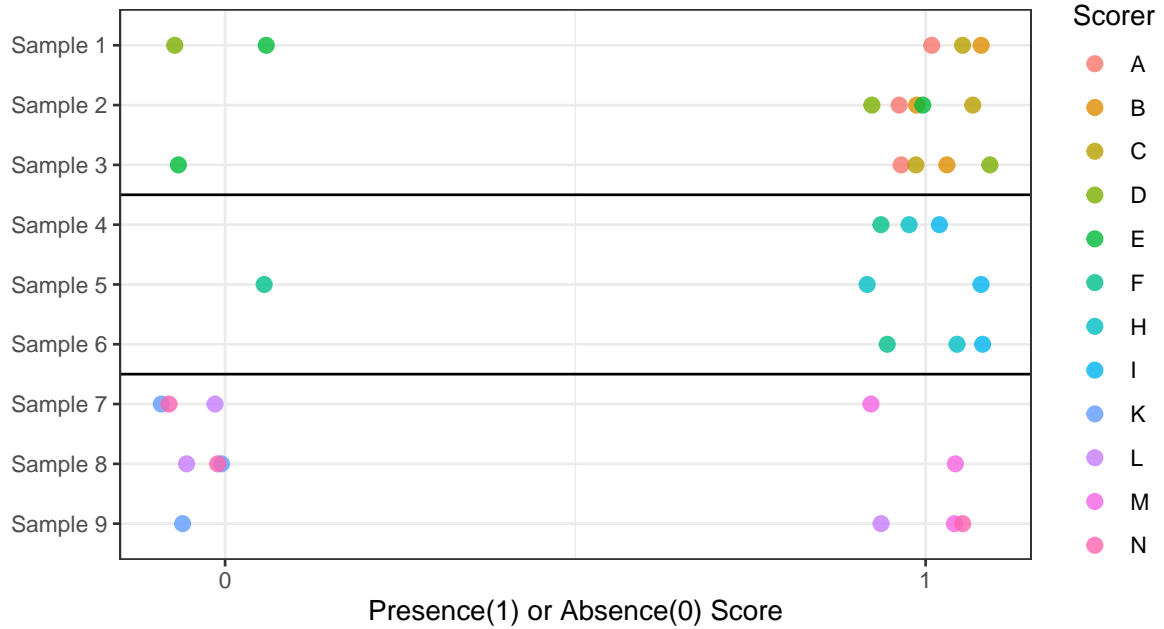
Cell cytoplasmic vacuolization

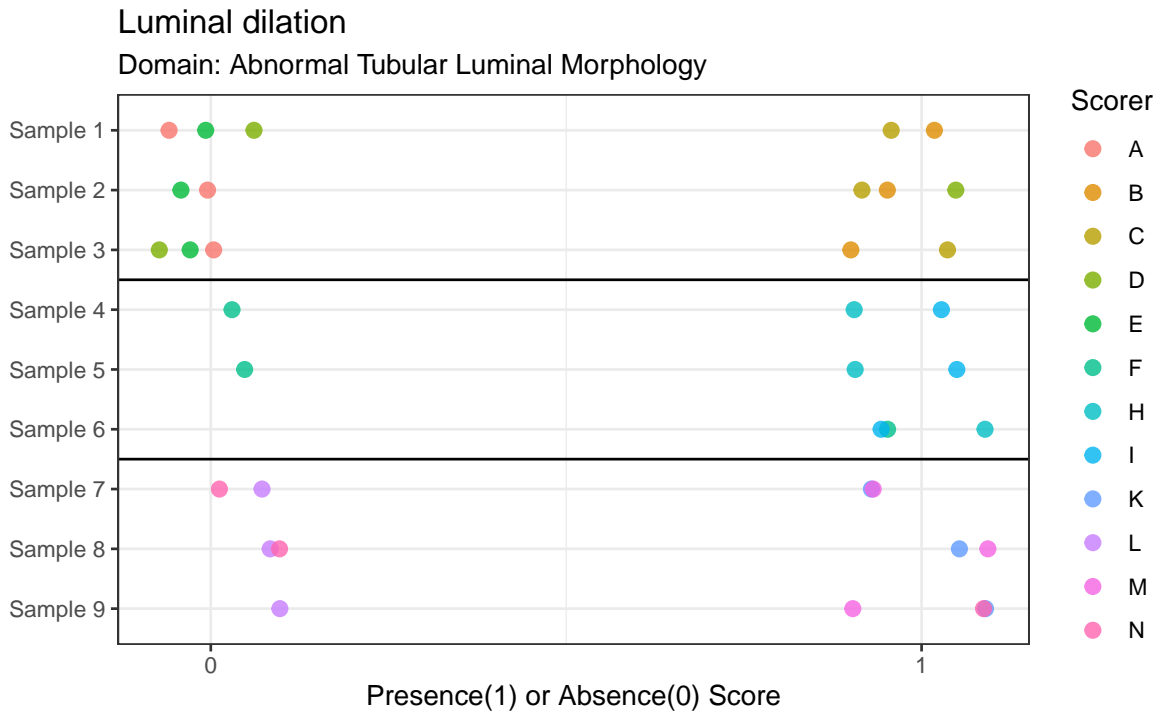
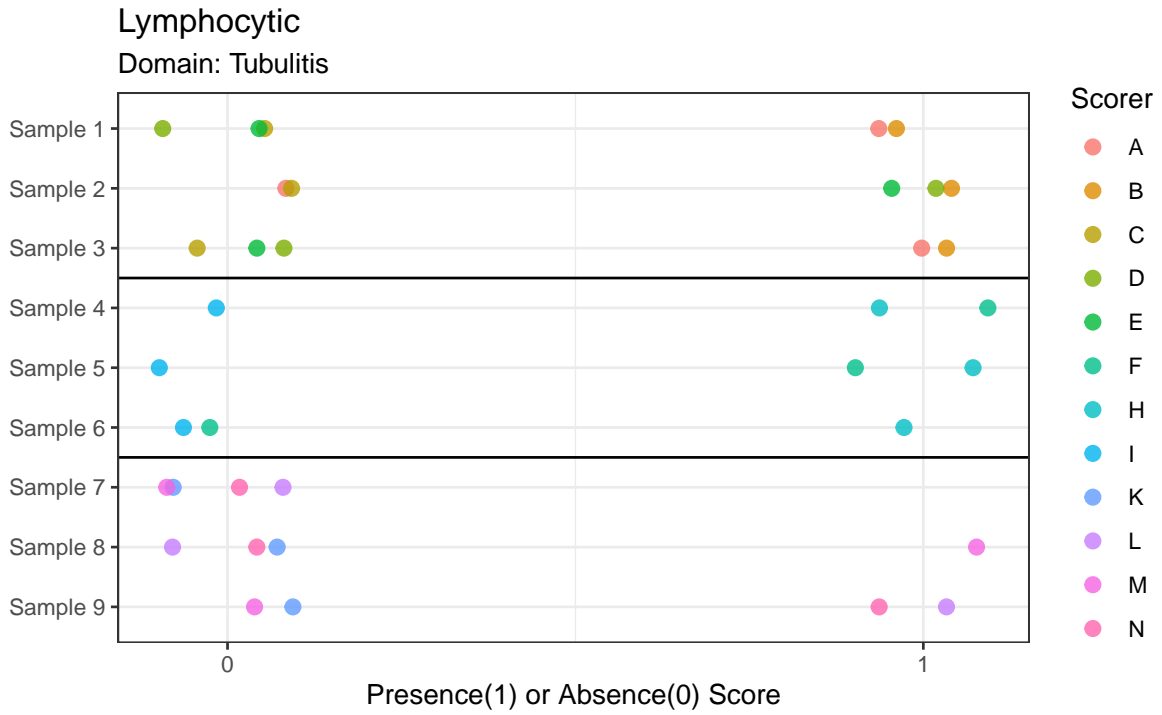
Domain: Tubular Cell Injury (other than atrophy)



Blebbing of apical cytoplasm

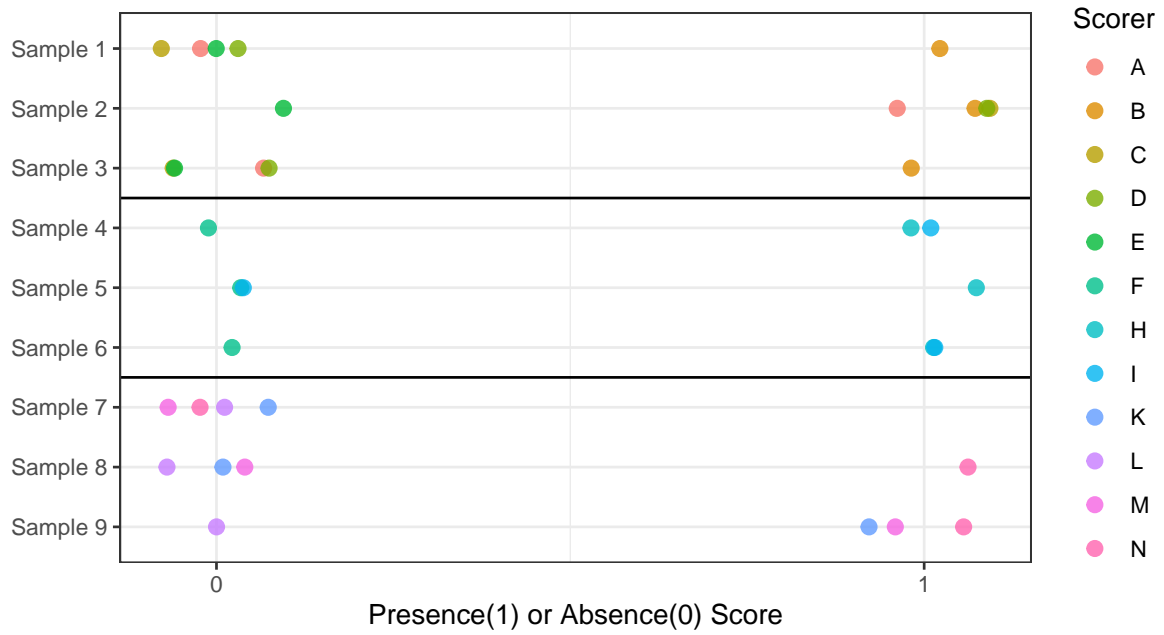
Domain: Tubular Cell Injury (other than atrophy)





(Acellular) Tamm–Horsfall casts

Domain: Intratubular Casts



Tubular basement membrane lamellation

Domain: Tubular basement membrane Morphology

