**Table S2\_** Characterization of the UC discovery cohort (from Portugal) and validation cohort (from Belgium), alone and in combination, in terms of clinical and pathological parameters. Minor allele frequencies for each SNP in the different cohorts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | ***Discovery Cohort (N= 432)*** | ***Validation Cohort (N= 499)*** | ***Combined Cohort (N=931)*** |
|   | **n (%)** | **Median [IQR]** | **n (%)** | **Median [IQR]** | **n (%)** | **Median [IQR]** |
| **Years old at diagnostic** |   | 427 | 35 [26-46] | 469 | 33 [24-43] | 896 | 34 [25-45] |
| **Years of follow up** |   | 381 | 12 [6-17] | 469 | 21 [16-28] | 850 | 17 [11-25] |
| **Gender** | F | 241 (55.8) | - | 217 (43.5) | - | 458 (49.2) | - |
| M | 191 (44.2) | - | 282 (56.5) | - | 473 (50.8) | - |
| Missing | 0 (0.0) |   | 0 (0.0) |   | 0 (0.0) |   |
| **Family History\*** | No | 389 (90.0) | - | 469 (94.0) | - | 858 (92.2) | - |
| Yes | 14 (3.2) | - | 30 (6.0) | - | 44 (4.7) | - |
| Missing | 29 (6.7) |   | 0 (0.0) |   | 29 (3.1) |   |
| **Disease extent** | Proctitis | 155 (35.9) | - | 44 (8.8) | - | 199 (21.4) | - |
| Left-sided | 129 (29.9) | - | 157 (31.5) | - | 286 (30.7) | - |
| Pancolitis | 144 (33.3) | - | 262 (52.5) | - | 406 (43.6) | - |
| Missing | 4 (0.9) |   | 36 (7.2) |   | 40 (4.3) |   |
| **Surgery** | No | 403 (93.3) | - | 333 (66.7) | - | 736 (79.1) | - |
| Yes | 25 (5.8) | - | 161 (32.3) | - | 186 (20.0) | - |
| Missing | 4 (0.9) |   | 5 (1.0) |   | 9 (1.0) |   |
| **Need for biologics** | No | 354 (81.9) | - | 334 (66.9) | - | 688 (73.9) | - |
| Yes | 73 (16.9) | - | 165 (33.1) | - | 238 (25.6) | - |
| Missing | 5 (1.2) |   | 0 (0.0) |   | 5 (0.5) |   |
| **Prognosis\*\*** | Good | 407 (94.2) | - | 449 (90.0) | - | 856 (91.9) | - |
| Bad | 22 (5.1) | - | 35 (7.0) | - | 57 (6.1) | - |
| Missing | 3 (0.7) |   | 15 (3.0) |   | 18 (1.9) |   |
| **SNPs (alleles)** | **Minor allele** | **MAF** | **MAF** | **MAF** |
| rs1257220 (G/A) | A | 0.282 | 0.278 | 0.280 |
| rs3814022 (C/G) | G | 0.314 | 0.259 | 0.285 |
| rs4953911 (A/T) | T | 0.359 | 0.279 | 0.316 |

**IQR -** interquartile range; **\*Family history** was considered the presence of at least one first degree relatives with IBD; **\*\*Prognosis** was defined as the need of biologics within the first 2 years after diagnosis, in which patients that needed biologics in up to 2 years after diagnosis were classified as displaying a bad prognosis; **MAF –** Minor frequency allele.