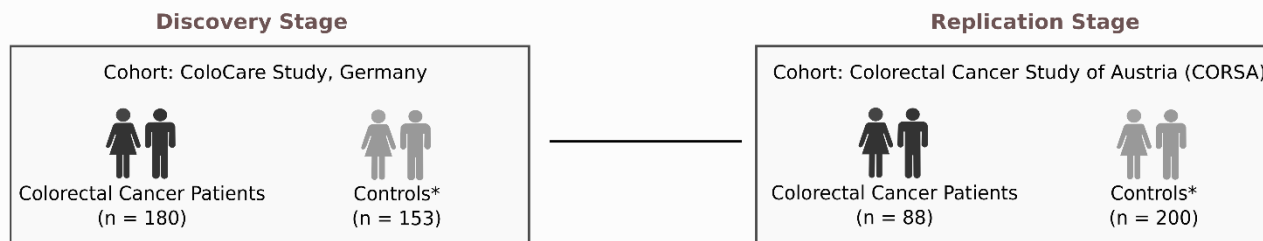


Supplementary Figure S1. Workflow and study design. (A) The study population consisted of 268 colorectal cancer patients and 353 controls (individuals not diagnosed with any colorectal malignancy), including an independent discovery (ColoCare Study: n=180 patients/n=153 controls) and replication set (CORSA Study: n=88 patients/n=200 controls). (B) Study workflow including feature finding and data alignment (raw data processing), data processing, data analysis (discovery-replication setting) and identification of replicated features.

A



B

1. Feature finding & Data alignment

- Sample analysis** UHPLC-ESI-qTOF-MS analysis
- Feature finding** Qualitative Analysis B.06.00, DA reprocessor, Mass Profiler Professional 12.1
Features n = 10,015
- Data alignment** Mass Profiler Professional 12.1 software
- Intensity filter** Features with intensity values < 10,000 in all study samples have been removed.

2. Data processing

- Missings filter** Features with missing values in >50% of samples in either colorectal cancer patients or controls in both populations were removed.
Remaining Features n = 2,669
- Blank adjustment** Features that had a minimum mean intensity below the mean intensity of blank samples were removed.
Remaining Features
n = 1,156 [Discovery set]
n = 1,148 [Validation set]
- log-transformation**
- Imputation** No imputation for univariate analysis.

3. Discovery stage

- Univariate Analysis** Multiple logistic regression models adjusted for age, sex, BMI (continuous) and smoking status.
Significant Features n = 691 [p < 0.05, FDR-corrected]

4. Replication stage

- Univariate Analysis** Multiple logistic regression models (one-sided testing) adjusted for age, sex, BMI (continuous) and smoking status.
Replicated Features n = 97

5. Feature Identification

Identified Metabolites n = 28

* Controls are defined as individuals not diagnosed with any colorectal malignancy.