DISCOVERIE

Project No. 848228

DISCOvERIE

Development, dIagnostic and prevention of gender-related Somatic and mental COmorbitiEs in iRritable bowel syndrome In Europe

Workpackage 7
Deliverable D7.1

Report on candidate mechanisms for comorbid development and mechanistic biomarkers from existing data

Start date of project: 2020-01-01

End Date: 2024-12-31

Duration: 60 months

Organisation name of lead contractor for this deliverable: Partner 15 (EMTEC)

Project co-funded by the European Commission within the Horizon 2020 research and innovation programme under Grant Agreement No 848228		
Dissemination Level		
PU	Public	**x**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Disclaimer

This document is an output of a project grant (Grant Agreement nº: 848228) funded under H2020 Research Programme of the European Commission. The content of this document represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission or any other body of the European Union. The European Commission do not accept any responsibility for use that may be made of the information it contains.

Consortium

Partcicipant Nº	Participant Legal Name	Country	Partner's Acronym
1	Fundacio Hospital Universitari Vall d'Hebron- Institut de recerca	ES	VHIR
2	Goeteborgs Universitet	SE	UGOT
3	Stichting Katholieke Universiteit	NL	SKU
4	University College Cork – National University of Ireland, Cork	IE	UCC
5	Johann Wolfgang Goethe – Universitaet Frankfurt Am Main	DE	GUF
6	Katholieke Univeriteit Leuven	BE	KUL
7	Universitaetsklinikum Heidelberg	DE	UKL-HD
8	Alma Mater Studiorum – Universita Di Bologna	IT	UNIBO
9	VIB VZW	BE	VIB
10	Universiteit Maastricht	NL	UM
11	Semmelweis Eguetem	HU	SU
12	Universitatea de Medicina Si Farmacie Iuliu Hatieganu Cluj-Napoca	ROU	UMF
13	Anaxomics Biotech, S.L.	ES	AX
14	Hycult Biotechnology BV	NL	HB
15	Emergentec Biodevelopment GMBH	AU	EMTEC
16	Nanostring Technologies Inc	USA	NSTG
17	Cliclab Technologies Inc	ES	Cliclab
18	Followhealth SL	ES	hITc
19	BYTE Factory Systems SL	ES	BF

Table of content

1. Executive Summary 3

2. Introduction 4

3. Research procedures 6

4. Conclusion and outlook 13

1. Executive Summary

Background: DISCOvERIE studies Irritable Bowel Syndrome (IBS) in relation to comorbidities, specifically depression, anxiety, fibromyalgia and chronic fatigue syndrome. While etiology of comorbidities may ground in various confounding peculiarities, DISCOvERIE specifically aims at providing insight into possible common molecular mechanistic grounds.

Specific research goals of this deliverable: Common molecular mechanisms refer to shared or linked molecular processes being responsible for onset and development of a comorbid condition. When identified, also the selection of embedded molecular biomarkers becomes feasible, promising insight into molecular causality. When included in diagnostic tools such biomarkers may add in diagnosis and prognosis of disease, eventually also in optimizing treatment and learning about preventive measures.

Research procedures: Systematic screening for molecular mechanisms and embedded biomarker candidates used established disease molecular data sets in Systems Biology – motivated computational analysis. In addition, expert input from members of the DISCOvERIE consortium was included. Algorithmic approaches grounded in molecular feature interaction networks for modelling disease pathologies, exploring three principal ways of forming comorbidities: common molecular basis, molecular mechanistic links, and compositionality of mechanisms as model for mediated disease links.

Results: Consolidation of complementary biomarker candidate screening provided a ranked list of close to 100 candidates. Additionally, network analysis allowed consolidation of candidate mechanisms.

Utilization: Given result status provides a reference for mechanisms and biomarker candidates proposing molecular links for comorbidities in focus. This reference will see refinement along molecular phenotyping activities performed in DISCOvERIE, and serves as primer for establishing experimental tools for diagnosis/prognosis, and ultimately for optimized treatment selection in comorbid conditions.

2. Introduction

2.1 Activity background

According to the project workplan for systems biology and integromics, research executed from 01/2020 until 06/2021 aimed for contributing to two goals: First goal is identification and annotation of candidate molecular mechanisms triggering development of comorbidities. Second goal is identification of candidate molecular biomarkers embedded in identified molecular mechanisms. With such embedding we aim for proposing causal biomarkers supporting diagnostic and prognostic assays informing of onset and evolution of comorbid conditions.

The data basis used for proposing candidate mechanisms and biomarkers is assembled from background of consortium partners. Specifically, identification and annotation of biomarker candidates embedded in molecular mechanistic context of Irritable Bowel Syndrome (IBS) and shared/linked molecular mechanistic context of the following comorbidity conditions is covered:

- Major Depressive Disorder (MDD)
- Anxiety Disorders (AD)
- Fibromyalgia (FM)
- Fatigue Syndrome, Chronic (FSC)

Research activities span elements as defined in the project tasks T7.2, T7.3 and T7.4 of WP7 (Systems Biology and Integromics) of the workplan. These tasks foresee **i)** identification of mechanistic links of comorbidities using protein functional interaction networks, **ii)** applying disease molecular model interference, and **iii)** elaborating on compositionality of pathology models.

Management of activities was implemented by WP7 lead (Partner 15, EMTEC); Interim result reports were distributed and discussed with consortium members in 03/2021, with a last update presented at the General Assembly in 05/2021.

2.2 Research proposition

In general terms, a disease initiates out of a combination of genetic predisposition and environmental factors (including life style), subsequently displaying a personalized trajectory of disease evolution, manifesting on the level of clinical presentation. Such presentation shall provide characteristics allowing diagnosis, shall support statements on prognosis, and is deemed to inform on optimal therapeutic means.

The general perspective also holds for comorbidities, at first to be seen as coexistence of pathophysiologies eventually to be split into categories: **i)** disease independence, only seen due to confounders (as e.g. age), **ii)** mediated by environment (an indirect effect of triggering a second morbidity on the grounds of a given first morbidity), or **iii)** grounding in causal molecular mechanistic coupling.

Research procedures followed the latter assumption: causally linked pathologies – needing clarification of terminology: Disease *pathology* refers to a molecular representation of disease on the level of molecular processes and mechanisms, which in the case of a disease come with some level of perturbance when compared to a situation with no respective clinical presentation. The term *linked* at first refers to a correlation of perturbance in comorbidities, only a subset of which being *causal* in the sense of one pathology depends on the existence of the other pathology.

One way to approximate causality in molecular systems is involvement of the same molecular constituents (common grounds leading to different clinical readout in relation to affected organs: e.g. a certain molecular mechanism “A” displays clinically as IBS and depression).
Another way to approximate causality is embedding of a comorbid condition in a linked network where disease A mechanistically triggers disease B. Yet another way to approximate causality is in the sense of *mediated* mechanisms: Existence of a molecular mechanism “A” evident in IBS triggers some mechanism “C”, as consequence initiating mechanism “B”, with “B” displaying on the clinical level as depression.

The following Figure 1 schematically introduces the different perspectives on causal embeddedness of comorbid conditions:

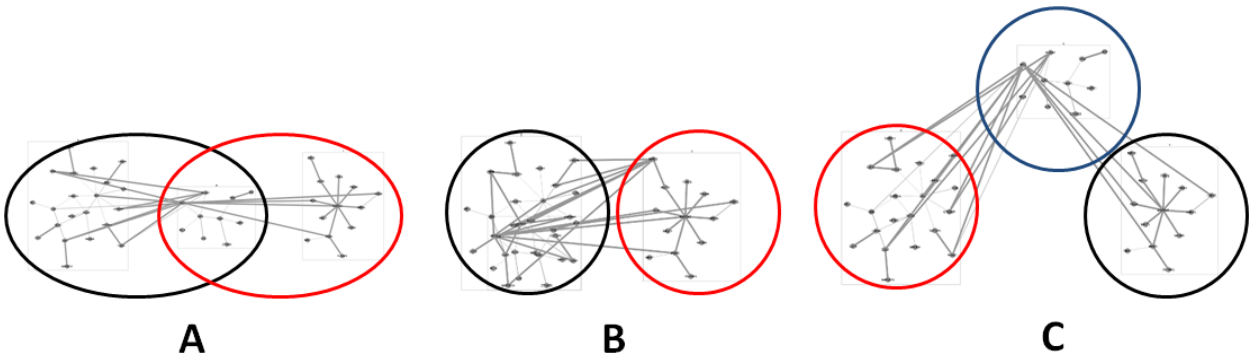

Figure 1: Causality in comorbidities. (**A**) Shared disease pathology molecular context; Context of pathology I (black) and pathology II (red) hold common aspects given by the intersection. (**B**) Embedded comorbidity; Pathology I and II integrate in an embedded pathology, indicated by the arrows linking pathology I and II context. (**C**) Mediated comorbidity; an additional molecular context (blue) links pathology I and II.

Context in Figure 1 refers to molecular interaction networks, a central concept used in research implementation. Specifically, such context may resemble individual molecular aspects, be it defined as processes, distinct motifs, or specific mechanisms. This notion directly links with the aim of searching embedded biomarkers. According to definition, a biomarker shall serve as proxy for disease-specific perturbance in molecular processes. Following the approaches outlined in Figure 1 per construction allows identification of biomarker candidates embedded in molecular processes relevant for describing comorbid conditions.

2.3 Implementation

Analysis of common grounds (Figure 1A), the embedding situation (Figure 1B) as well as of mediated mechanisms (Figure 1C) needs a way to model molecular pathology context, at Partner 13 (AX) and Partner 15 (EMTEC) realised as molecular interaction networks. Such networks are composed of nodes (molecular features considered relevant in one or both comorbid conditions). Nodes are linked by edges, in given models representing molecular feature interactions backed by evidence from experiment and/or computational inference. This approach is strictly different to general knowledge graphs (also coming as networks), where in knowledge graphs the interactions resemble other properties as e.g. frequency of co-annotation.

As causality is a central aspect for studying IBS – comorbidity mechanisms and embedded biomarkers, analysis at Partner 13 (AX) and Partner 15 (EMTEC) sticks close to a biological proxy for causality, hence the use of molecular interaction information for constituting the networks.

Data spaces used for approximating relevant mechanisms of comorbidity grounded in background data available at Partner 13 (AX) and Partner 15 (EMTEC). For fully leveraging on expertise in the project consortium, expert input together with focused literature screening was amalgamated with computational procedures.

The following Section 3 discusses research procedures.

3. Research procedures

3.1 Expert input

Biomarker candidates were discussed with and collected from consortium members and teams in a series of online meetings organized by WP7 (EMTEC) and WP8 (HB) leads. WP7 is responsible for data-driven systems biology; WP8 focuses on biomarker validation and development of diagnostic tools. As biomarker candidates are the tangible result of research outlined in the section above, teaming up of WP7 and WP8 at early project stage was considered effective for bringing computational work toward experimental testing.

Online meeting	teams/members
05/2020	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB): preparation for WP7/8 meetings with selected project partners
07/2020	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB) / Judith Farres (AX) / Ricard Farre (KUL): discussion on biomarkers proposed by KUL
07/2020	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB) / Gerard Clarke (UCC) / Kieran Rea (UCC): discussion on biomarkers proposed by UCC
10/2020	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB) / Javier Santos (VHIR) / Ana Maria Gonzalez Castro (VHIR) / Eva Hajdok (VHIR): discussion on biomarkers proposed by VHIR
12/2020	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB): preparation for WP7/8 meetings with selected project partners
12/2020	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB) / Daisy Jonkers (UM) / Daniel Keszthelyi (UM): discussion on biomarkers proposed by UM
12/2020	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB) / Judith Farres (AX) / Javier Santos (VHIR) / Magnus Simren (UGOT) / Eva Hajdok (VHIR): WP7/8 work plan and time lines
01/2021	Arno Lukas (EMTEC) / Ricardo Brandwijk (HB) / Beate Niesler (UKL-HD): discussion on biomarkers proposed by UKL-HD and others

Consolidation of these meetings resulted in a collection of 92 proposals, corresponding to 75 unique candidate biomarkers together with respective scientific references.

Additionally, scientific reviews targeted at IBS-comorbidities in focus were shortlisted. Selected references considered relevant in project focus include:

- Lessons learned--resolving the enigma of genetic factors in IBS. Gazouli et al., Nat Rev Gastroenterol Hepatol 2016 Feb;13(2):77-87.
- Longitudinal Multi-omics Reveals Subset-Specific Mechanisms Underlying Irritable Bowel Syndrome. Mars et al., Cell 2020 Sep 17;182(6):1460-1473
- Epigenetic Mechanisms in Irritable Bowel Syndrome. Mahurkar-Joshi et al., Front Psychiatry 2020 Aug 14;11:805

Literature reviewing was done at Partner 15 (EMTEC), providing 122 molecular features to include in molecular mechanism and biomarker analysis. The given list holds various feature types, including 29 SNPs (Single Nucleotide Polymorphisms), 20 miRNAs (micro-RNAs) and 73 protein coding genes.

For allowing analysis on the level of protein coding genes/protein interaction networks, SNP and miRNA information was brought to the protein coding gene level (Partner 13, AX). For SNPs the position in protein coding genes was used for assignment, for miRNAs reference databases for their association in controlling protein coding gene expression was used.

3.2 Computational analysis

Computational analysis rested on established molecular data spaces and analytics procedures of the Partners 13 (AX) and 15 (EMTEC) according to the conceptual considerations outlined in Section 2 of this document.
Computational work was designed to support two aspects:

1. evaluate/rank biomarker candidates suggested from expert discussions
2. propose additional/novel biomarker candidates

Inherent was identification and characterization of molecular mechanisms and pathways, directly derived from interaction network segments.

3.2.1 Protein network analysis/embedded comorbidity

To evaluate the candidate biomarkers proposed by project partners in a mechanistic context and identify additional candidates not previously considered, Partner 13 (AX) built a protein interaction network around proteins known to have a role in IBS, established mathematical models based on this network and used different network metrics to identify which of the candidates appear indicative for IBS and comorbidities in a molecular mechanistic way.

Method

A molecular data set summarizing existing knowledge on IBS was collected through manual curation of PubMed publications, following a standard protocol. This molecular description was segmented in different pathological motives. AX further collected in a similar manner protein data sets of known molecular players in anxiety, depression, fibromyalgia and chronic fatigue syndrome.
A human protein functional interaction network was assembled around the protein data set containing the known molecular players in IBS. The interaction data connecting the different proteins were derived from public domain data bases.

Mathematical models were built based on this network following the modelling strategy described in Jorba et al. (PLOS ONE. 2020;15: e0228926. doi:10.1371/journal.pone.0228926). Two types of modelling strategies were used for identifying cross-affectedness of IBS and comorbidities. One takes into consideration the relations between the proteins (topology of the network) to establish a relationship probability. A second strategy propagates in-silico stimuli through the topology of the protein network and identifies proteins highly affected.

In addition, the candidate list of biomarkers proposed by WP7 activities was mapped to uniprot codes, human proteins through name search, metabolites through human metabolome database and miRNA using the name of the protein associated as given by the proposing partner.

Results

Molecular data sets describing the current knowledge on IBS, anxiety, depression, fibromyalgia and chronic fatigue syndrome were assembled. Table 1 lists different functional motives of IBS and the number of entries per motive. Some of the proteins have a role in several motives; the final set contained 85 unique proteins.

MOTIVE ID	Motive Name	Effectors (No.)
1	Decreased gastrointestinal motility	15
2	Increased gastrointestinal motility	18
3	Increased intestinal permeability	34
4	Visceral hypersensibility	32
5	Gut-Brain Axis	47
6	Low-grade mucosal inflammation	22
7	Host Gastrointestinal-microbiota interaction	14
	Total entries	**182**
	Unique entries	**85**

Table 1: Molecular motifs identified for IBS, and number of protein coding genes involved.

The molecular data sets for anxiety and depression were pooled to establish the ‘mental data set’ containing 261 unique proteins. Likewise, fibromyalgia and chronic fatigue syndrome were pooled to establish the ‘somatic data set’ with 81 unique proteins.

The modelling strategy that runs in-silico stimulation of randomly selected proteins around the protein network of the IBS molecular data set and measures the propagation of the signal rendered 320 proteins with an alleged higher change of activity, thus higher involvement in the pathology. The candidate biomarkers from the consortium were also mapped in the model network to check if they were present or not.

In order to evaluate the relation of candidate biomarkers and the 320 proteins identified in the IBS signal propagation model with IBS, mental and somatic comorbidities, a second modelling strategy was used that gives a measure of relationship. The input for the model is the proteins under evaluation (including 1090 features) and as outset the collected IBS, mental and somatic protein sets. A likelihood of relation was obtained for each input protein with each of the outset data sets. Proteins were selected with a cut-off relation score that corresponds to a p value<0.1. In this way 4 types of protein groups were obtained depending on which group they showed significant relation:

- IBS+Somatic+Mental
- IBS+Somatic
- IBS+Mental
- 0 (in case no significant relation has been established)

In the case of metabolites, the signal that comes from the different proteins they are related was integrated. In the case of IBS two types of measures of mechanistic relationship were derived, first the presence of the protein in the IBS signal propagation model, and the relation score provided by the modelling strategy based on the topology.

Next to evaluation of biomarker candidates provided by consortium partners, 11 additional proteins above the relation score cut-off that were identified. In order to evaluate the potential for those proteins to be measurable in an in vitro assay, the expression profile was annotated according to Human Protein Atlas annotation scheme. **6 biomarker candidates were selected that are either secreted or membrane bound.**

3.2.2 Molecular model strategy/common grounds, composition

Approach I: Direct molecular model overlap, common ground

Partner 15 (EMTEC) utilized molecular pathology models of Irritable Bowel Syndrome (IBS), Major Depressive Disorder (MDD), Anxiety Disorders (AD), Fibromyalgia (FM) and Fatigue Syndrome, Chronic (FSC).

Specifically, a pathology model integrates molecular features reported as relevant in a disorder on a whole human proteome reference interactome, and extracts the “connected component” specific for a disorder in focus. Such network model therefore holds nodes coding for molecular features of relevance in a specific disease (with protein coding gene level as common denominator), and edges code for molecular interactions of various types. Such a model is termed a “context model” for a disease. Where applicable, such context model is transformed into a “process model” using topological characteristics of the network. In such procedure a context model is split into highly connected cliques, each of which being subsequently termed a “process”. Further details on technology aspects are provided in Mayer et al., Curr Pharm Des. 2017;23(1):29-54.

The individual phenotype molecular model characteristics are given in the following Table 2:

IBS (MeSH term: Irritable Bowel Syndrome)	**Input feature #**	**context model feature #**	**process model feature # / number of processes**
	140	88	35/6

MDD (MeSH term: Major Depressive Disorder)	**Input feature #**	**context model feature #**	**process model feature # / number of processes**
	613	513	160/16

AD (MeSH term: Anxiety Disorders)	**Input feature #**	**context model feature #**	**process model feature # / number of processes**
	176	129	43/6

FM (MeSH term: Fibromyalgia)	**Input feature #**	**context model feature #**	**process model feature # / number of processes**
	85	59	6/2

FSC (MeSH term, Fatigue Syndrome, chronic)	**Input feature #**	**context model feature #**	**process model feature # / number of processes**
	87	52	9/1

Table 2: Characteristics of molecular models covering IBS, MDD, AD, FM and FSC. The disease term reference is taken from Medical Subject Headings (MeSH) according to NCBI controlled vocabularies. "Input feature number" refers to the total count of molecular features (mapped to the level of protein coding genes") being extracted from various sources (NCBI, patent text, clinical trials references, etc.) holding annotation with the respective MeSH disease term. "context model feature #" refers to the number of input features being part of the context model (the largest connected component). The same information is also given for process models, together with the number of "processes" identified in a given context model.

With having a molecular pathology model for IBS and one for each comorbid condition of interest algorithmic "interference" of the networks becomes feasible, and subgraphs of the IBS context model being shared by a comorbidity can be identified (see Figure 1A).
Such shared aspects are considered as relevant molecular context in both diseases, i.e. a shared comorbidity mechanism.

On such basis biomarker candidates can be searched being part of such shared components of pathology models, i.e. informing on molecular mechanistic aspects assumed to be relevant in both disorders.

The four interference signatures (IBS – comorbidity pairs including MDD, AD, FM, FSC) were searched for nodes holding annotation as biomarkers. We only included candidates if in addition holding annotation (experimental evidence from publications) as biomarker (with explicit annotation in prognosis/diagnosis) in at least one of the two comorbid conditions for a given interference according to NCBI MeSH annotation.
Additionally, biomarker candidates proposed by other project teams were evaluated for being members of interference signatures.

This approach of focusing on common molecular mechanistic grounds of IBS and comorbidities identified 10 biomarker candidates.

For the interference signatures enrichment of KEGG pathways was retrieved for gaining insight into molecular mechanism context, providing mechanism-candidate biomarker assignments.

Approach II: Composite models and mediators

Partner 15 (EMTEC) pursued a second approach: derive composite molecular models from a molecular process hierarchy, and identify mediators (blood-based proteins) linking IBS and comorbidity molecular processes (see Figure 1C). The basis of the method is discussed in Lukas at al., bioRxiv 573402; doi: https://doi.org/10.1101/573402.

For realizing this approach an extensive set of general disease-associated molecular processes was extracted according to a standardized computational workflow. Data processing provided in total 9,068 molecular pathology processes across the MeSH disease vocabulary, covering 6,166 unique protein coding genes. This set of processes also included the processes identified for IBS and the individual comorbidities (for which direct interference was studied, see Table 2 above). Central idea of composite models is that co-morbidity processes may integrate at a certain level of molecular process aggregation (see Figure 1C). In such scenario closeness of morbidity processes is defined by the process aggregation level. The more processes needed to amalgamate an IBS-comorbidity process pair the less specific is the resulting aggregate with respect to a mediating situation. The metric chosen for quantifying the distance of process pairs is ratio of molecular feature overlap and total number of features seen in an aggregate. Process aggregation and process merge levels are derived via hierarchical clustering. Specifically, for each IBS process (6 in total) the interest is:

- At what aggregation level is an IBS process merging with a comorbidity process (a proxy for compositional closeness)
- What is the ratio of molecular features coming from an IBS process merging with a comorbidity process with respect to the total number of features in a merged situation (only enriched situations are relevant)
- What molecular mechanisms dominate such merged situation (expanding the molecular mechanism perspective beyond direct overlap)
- Which molecular features embedded in such merged situation are secreted and hold evidence for serving as biomarker in IBS and/or the comorbidity situation

Computation followed an all-against-all pair-wise process merge analysis. For allowing to hold on a statistical criterion with respect to relevance of process aggregation we derived the distribution for all pair-wise process merges as a score evaluating the sum of features coming from specific process aggregates in relation to the total number of features embedded in an aggregate at a certain merge level. Then all pair-wise process merge scores for all IBS – comorbidity process pairs were identified, with focus on process merges in the top 1% of the general score distribution.

Executing this procedure provided two hits in focus of IBS processes: One such aggregate embeds an IBS process and a fibromyalgia process, including in total 21 protein coding genes. A second aggregate embeds two IBS processes, one process from MDD and one process from AD, including in total 374 protein coding genes.

Screening this protein set against the Protein Atlas reference on proteins secreted in blood provided 17 novel biomarker candidates linking IBS, MDD and AD via circulating protein mediators.

4. Conclusion and outlook

All research procedures and results rested on background (existing) data with the aim of establishing a first reference on molecular mechanisms and embedded biomarker candidates linking IBS with the comorbidities in focus.

Computational methods are in place (see Section 3 of this document) for further evaluating project foreground data to come, including e.g. details of intestinal barrier function, CNS-stress axis, microbiome/metabolome and genetics/epigenetics for comorbidities in project focus.

Respective tasks of WP7 will continue until 2023, targeted at continuously expanding and refining molecular mechanisms and embedded biomarkers.

Given results status combined biomarker candidates proposed by various teams and resulting from diverse computational approaches. Consensus listing holds 100 unique candidates, including a rank according to consensus evaluation of evidence levels.

This status has allowed WP8 a first shortlisting for development of diagnostic tools, needed for quantitation of selected molecular biomarkers in clinical cohorts being established in WP2. With tools and biobanks in place, diagnostic and/or prognostic value of identified biomarker candidates with respect to onset and evolution of comorbidities becomes amenable.

Biomarker candidates were one goal of research procedures, the second goal was annotation of respective molecular mechanisms. Computational approaches utilized molecular interaction networks, and aspects considered relevant (be it from direct overlap, from process links or from mediated processes) allowed analysis for molecular mechanism and pathway enrichment, providing a ranked list with at least two levels of use in project research.

The first is in support of named candidate mechanisms, including e.g. inflammation and stress response.

A second use is in studying therapeutic drug effect, maybe also a constituting element in comorbidities. A dedicated task is included in WP7 aimed at drug mechanism of action interference with IBS and comorbidity pathologies. Combining drug effect evaluation with biomarker quantitation promises to combine diagnostic/prognostic statements and optimal drug selection.