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 D7.2

Report on candidate mechanisms and biomarkers for comorbid IBS

Start date of project: 2020-01-01

End Date: 2024-12-31

Duration: 60 months

Organisation name of lead contractor for this deliverable:

Partner 13: AX

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

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
4	University College Cork – National University of Ireland, Cork	IE	UCC
5	Johann Wolfgang Goethe – Universitatfrankfurt Am Main	DE	GUF
6	Katholieke Univeriteit Leuven	BE	KUL
7	Universitatsklinikum 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	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
20	Stichting Radboud Universitair Medisch Centrum	NL	RUMC

Table of Content

List of abbreviations

IBS: Irritable Bowel Syndrome
IBS-SSS: IBS Symptom Severity Scale
IDE: Individual Differential Expression
Protein A
Protein B
Protein C
Protein D
TPMS: Therapeutic Mapping System

JF: Judith Farrés (Anaxomics)
JMG: Juan Manuel García (Anaxomics)
PB: Paloma Buj (Anaxomics)

Executive Summary

This study aims to analyse the molecular differences between three distinct groups of IBS patients using systems biology approaches: IBS-Alone (Individuals with IBS and no comorbidities), IBS-Comorbid (Individuals with IBS and one comorbidity, either mental (depression or anxiety) or metabolic (fibromyalgia or chronic fatigue syndrome)) and IBS-Multimorbid (Individuals with IBS and two comorbidities, one mental and the other somatic).
Using a systems biology-based approach, we have built individual patient models from transcriptomic signatures derived from colonic biopsies from DISCOVeRiE study. We successfully constructed in-silico individual models that accurately represent IBS severity. Significant differences were observed in the downstream signalling networks between individual models from different study arms. Pathway overrepresentation analysis revealed several enriched signalling pathways in IBS-Multimorbid patient models compared to IBS-Alone. Proteins with significant differences in activation patterns between IBS-Alone and IBS-Multimorbid patient models include Protein A, Protein B, Protein C, and Protein D. The identified pathways and key proteins will be further investigated within the context of the project, utilizing data from other work packages.

Introduction

With this study we aim at analysing at a molecular level and using systems biology approaches the differences between

- IBS-Alone: individual with IBS no comorbidities
- IBS-Comorbid: individual with IBS and one comorbidity either mental (depression or anxiety) or somatic (fibromyalgia or chronic fatigue syndrome)
- IBS Multimorbid. individual with IBS and two comorbidities one mental and the other somatic

The objective is to identify mechanistic links using human protein functional interaction network-based models based on the TPMS (Therapeutic Mapping System) technology developed by AX. The analysis of the molecular data in the context of protein networks allows identifying pathways and biomarkers in a mechanistic rather than a statistical association analysis. Mechanistic biomarkers, which are rooted in the biologic mechanisms of disease, have the greatest potential for guiding clinical decision making.

To build in silico patient models representative of the different study arms we have used individual transcriptomic signatures derived from colonic biopsies. We have adjusted the in-silico response to correlate with the severity of IBS symptoms of each subject.

The resulting models have been analysed in terms of downstream signalling and differential pathways and key proteins have been identified.

IBS knowledge Set

A molecular data set summarizing existing knowledge on IBS was already presented in D7.1, is introduced here again for context. The data set was collected through manual curation of PubMed publications, following a standard protocol. This molecular description was segmented in different pathological motives, **Table 1**. AX further collected in a similar manner protein data sets of known molecular players in anxiety, depression, fibromyalgia and chronic fatigue syndrome. This knowledge sets have been used to build patient network models see the following sections.

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

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**

Individual Differentially Expressed proteins (IDES)

In order to obtain molecular profiles characteristic of the study arms of the project, that is IBS-Alone, IBS-Comorbid and IBS-Multimorbid, we have used bulk transcriptomic data streaming from the intestine biopsies, (see details in D3.3). In order to obtain an individual signature for each patient, we used transcriptomic data normalized across all samples (DESeq2 package, VST). The number of individuals available per class are summarized in **Table 2**.

Table 2 Number of individuals per study arm with transcriptomic data.

No. of Individuals	Female	Male	Total
Healthy	9	10	19
IBS-Alone	10	10	20
IBS-Comorbid	14	6	20
IBS-Multimorbid	18	2	20

For each IBS individual, a signature was derived based on a reference distribution for each gene from healthy individuals, segmented by biological sex. Upper and lower cut-offs were set at the 5th and 95th percentiles of the reference distributions. If the value of a gene in an IBS individual was higher than the upper cut-off in the reference distribution for that gene, the gene was assigned a value of +1 for that patient. Conversely, if a gene in an IBS individual was lower than the lower cut-off in the reference distribution for that gene, the gene was assigned a value of -1 for that patient. If the value was within the normal distribution, the protein was not listed in the signature. In order to use these proteins in the protein network models we have restricted the IDES to only proteins that are at 3 known links of distance to the IBS knowledge set, to remove spurious signals. And in order to increase the differences between the study arms we have selected proteins that when pooling the IDEs per study arm together and comparing them with the rest of study arms are significant ($p<0.01$ for Fisher's exact test), **Table 3**.

Table 3 Differential gens per IBS study arm streaming from IDEs. *IDE of individuals of the same study arm are pooled together and differential gens between each IBS study arm and the rest of IBS individuals, have been calculated using Fisher's exact test.*

	Fisher's exact test p-value <0.01		Fisher's exact test p-value <0.01	
Study Arm	Female (No. Genes)	Male (No. Genes)	Female (No. Genes)	Male (No. Genes)
IBS-Alone vs Other IBS	57	2	235	18
IBS-Comorbid vs Other	52	25	232	232
IBS-Multimorbid vs Oth	69	23	424	142

We did some enrichment analysis on the significant proteins from pooled IDEs per study arm. We outline here some of the relevant results. A total of 57 genes show a significant difference between IBS-Alone female group and rest of female IBS (comorbid+Multimorbid) at pvalue<0.01. When doing an overrepresentation analysis on these proteins we see that there is a significant enrichment in the Reactome pathways related with immunity (corrected p value < 0.05).

A total of 69 proteins show a significant difference between IBS-Mutimorbid female group and rest of female IBS (Alone+Comorbid) at p-value<0.01. A pathway enrichment analysis on these proteins shows a significant enrichment in a Reactome pathway.

Individual patient Models - Protein network analysis

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[1]. The models propagate a signal from a stimulus set to a response set (IBS knowledge set).

As a stimulus, we used a group of proteins from the network that were calculated using the TPMS approach to have the capacity to propagate signals effectively, maximizing interaction with other proteins in the response set, **Table 4**. By doing this type of analysis we are able to set a signalling network around the proteins known to be related to IBS. As the propagation of the signal is set to stimulate the pathogenic function of the protein in the response set, the model simulates the pathology.

Table 4 Proteins used as stimulus in the construction of individual patient models. *Proteins around IBS network Identified using the TPMS approach to have the capacity to propagate signals effectively to the response set.*

Gene name	Stimulation	Uniprot
CD14	1	P08571
NCOR1, KIAA1047	1	O75376
MAML1, KIAA0200	1	Q92585
RASSF5, NORE1, RAPL	-1	Q8WWW0
PIAS4, PIASG	-1	Q8N2W9
FST	-1	P19883

In order to build individual models, we introduce the calculated IDEs as a restriction, that is if the signal goes through a protein of the IDE the model should fulfil the signal of protein set in the IDE. The system is set to fulfil a minimum of 50% of the restriction, as some of the signals may be incompatible between them. Motives 1 and 2 of IBS knowledge set correspond to the molecular description of constipation and diarrhoea, only one of the motives was used in the models, depending on the type of IBS of the individual. At the same time, we modulated the total response (reversion of IBS knowledge set proteins) to correspond to IBS severity extracted from the study parameter IBS Symptom Severity Scale (IBS-SSS). We built a total of 20 individual models, as those where the patients with IBS-SSS score available at the time of analysis, 10 IBS-Alone individual, 3 IBS-Comorbid individuals and 7 IBS-Multimorbid individuals. This type of modelling strategy renders many possible signaling pathways between stimulus and response, most of them highly similar. We computed for each patient a total of 250 possible solutions, in order to capture as many possible variations in the signaling. The final response is given by the mean signal of the 250 solutions measured at the response set. If the signal received by a protein is the contrary of what is expected (pathological signal) the value is deduced from the total measurement, so the final in silico signal (FTsignal) is the addition of all the signals received in the response set multiplied by the expected signal and divided by the

[1] PLOS ONE. 2020;15: e0228926. doi:10.1371/journal.pone.0228926

total number of proteins of the set. If all proteins of the response set are reached and the signal they receive is to keep the pathogenic status the FTsignal will be 1, while if all the proteins of the response are reached and the signal received reverts the pathological function of the protein then the FTsignal will be of -1.Lower number than 1 could be due to the fact than not all proteins of the response set receive a signal from the input or that some receive and input that reverts the signal from the pathogenic state.

In **Figure 2** we observe the adjustment obtained between the individual patient models mean signal on the IBS knowledge set proteins and their reported IBS-SSS score. We obtained individual models that were correctly simulating in silico the phenotype, in terms of IBS severity, measured in vivo.

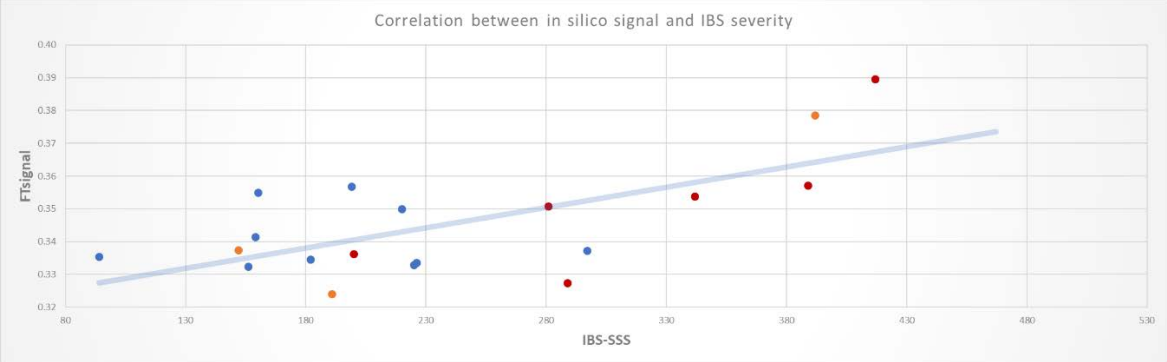

Figure 1 Correlation between the mean in silico response (FTsignal) and the IBS severity (IBS-SSS). *Dots represent individual patient models, blue correspond IBS-Alone, orange correspond to IBS-Comorbid, and red correspond to IBS-Multimorbid.*

Models' comparison

Individual in silico patient models accurately replicate the severity of IBS reported, making them a reliable representation of real patients in this regard. We have used the individual models obtained to compare IBS study arm. There are no significant differences between IBS study arms and the in-silico signal measured in the IBS knowledge set used as a response **Table 5**.

Table 5. In silico signal measured over IBS knowledge set. *Mean value of in silico signals of patient models from the same arm is represented.*

	Full TSignal on IBS	
	Mean	Std
IBS alone	0.34	0.04
IBS Comorbid	0.35	0.03
IBS Multimorbid	0.35	0.03

The same type of measurement was applied to the different pathogenic mechanisms into which we have segmented the IBS knowledge set. No differences between IBS study arms were observed, but we noted that the pathogenic mechanisms receiving the most signals were '*low-grade mucosal inflammation*', followed by '*increased intestinal permeability*'. The molecular data sets used to build individual patient models were derived from individuals' biopsies, which may make the response more representative of the intestinal aspect of the disease.

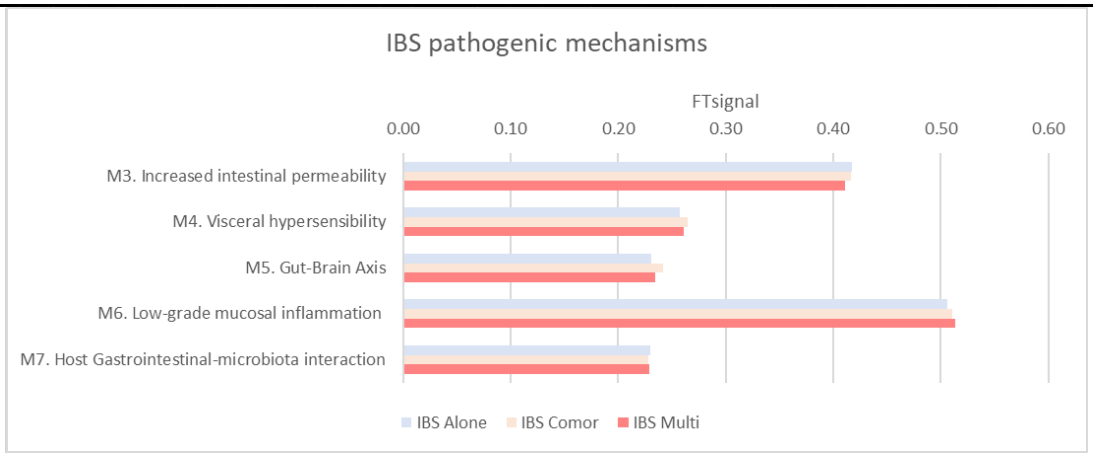

Figure 2 In silico signal measured over the individual pathogenic mechanisms within IBS knowledge set. *Mean value of in silico signals of patient models from the same arm is represented.*

We have also evaluated how the molecular knowledge sets representing the comorbidities under study, are reached in the virtual individual models. Using the same starting input, we measured the signal recovered in the protein knowledge sets representing the mental and somatic comorbidities under study. The measure *% of effectors* helps us determine whether the signal from the stimulus is effectively reaching the response proteins. This entails verifying whether the signal appropriately activates the intended proteins and inhibits the ones that are supposed to be inhibited. Patient models in the comorbid and multimorbid IBS arms exhibit a higher mean percentage of effectors, indicating greater precision compared to the IBS-alone models, **Figure 4**.

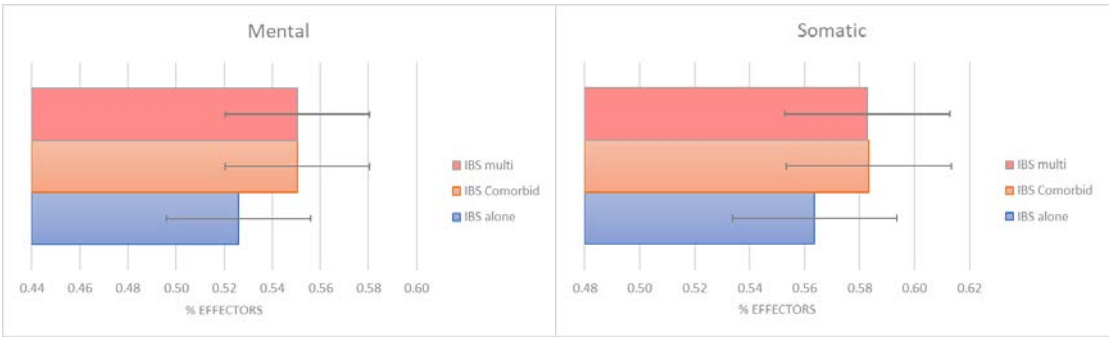

Figure 3 Proportion of proteins of the knowledge sets reached with the correct sign (pro pathologic). *Mean value of in silico signals of patient models from the same arm is represented.*

The in-silico models give a rich set of potential molecular interactions explaining the activation measured in the response set. We examined the signalling networks generated by different virtual individuals; by propagating the signal from the stimulus set of proteins to the response set (IBS knowledge set), the models identify proteins that become activated or inhibited. The models' signal range is from +1 (full activity) to -1 (no activity). For each patient model, we collected 250 possible solutions. To evaluate the signalling network per IBS study arm, we pooled the 250 solutions for each patient in the study arm.

Pairwise comparisons for each protein in the downstream signalling network were conducted using the Wilcoxon signed-rank test to identify significant differences (FDR < 0.05). Comparing the signalling patterns of IBS-Alone in silico models with those of IBS-Multimorbid revealed 225 proteins with significantly different signals. To increase certainty, we removed small signal differences and selected proteins with absolute

differences equal to or greater than 0.10. This cut-off reduced the number of proteins with different signalling between IBS-Alone and IBS-Multimorbid to 48, **Table 6**. The comparisons with IBS-Multimorbid are the ones rendering a larger number of significant differences in the models signalling. As the number of comorbid individual models is low, we concentrated the exploration in the comparison of Alone vs Multimorbid models.

Table 6 Proteins with significant differential signal between in silico patient models for IBS-Alone compared to IBS-Multimorbid

	No Proteins FDR wilcoxon< 0.05	
	Any Median difference	Median difference >|0.10|
Alone vs MutiComorbid	225	48
Comorbid vs MultiComorbid	203	23
Comorbid vs Alone	93	46

A pathway overrepresentation analysis of 48 differentially activated genes between IBS-Alone and IBS-Multimorbid patient models, rendered 17 overrepresented pathways from 4 different knowledge bases, KEGG, BioCarta, INOH and Reactome

We analysed all the signalling proteins for their classification potential, with linear regression as the base classifier. However, the cross-validated AUC values were quite low (0.56-0.58). The proteins with the higher AUC values also overlapped with the ones with higher differential signal and significant in the Wilcoxson test. The signal distribution for the four proteins with the greatest differences is shown in Figure 4. We conducted a literature review to determine if there have been any previous associations with IBS, metal or somatic diseases. Previous associations with IBS have been described specially for Protein D and Protein A.

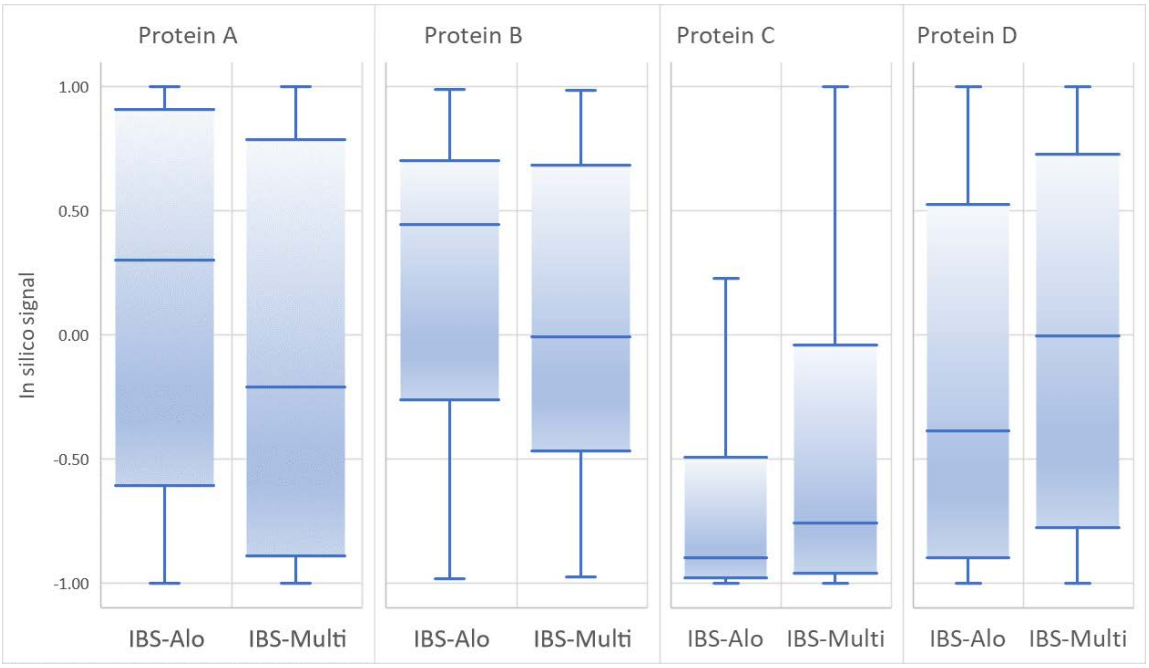

Figure 4. ***Signal distribution for the proteins with a higher difference between in silico patient models for IBS-Alone compared to IBS-Multimorbid***

Conclusions

Using individual differentially expressed proteins from biopsies, we have been able to construct in silico individual models that correctly represent the IBS severity.
The in-silico individual models, when compared by IBS study arm, did not show differences in terms of amount of total signal received at IBS knowledge set (more pathology) nor showed differences in total signal measured per each IBS pathogenic mechanisms.

The proteins known to have a role in the mental and somatic comorbidities where better reached (higher precision) in-silico individual models for IBS-Comorbid and Multimorbid compared to IBS-alone models.

Significant differences have been observed in the downstream signalling network between individual models from different study arms. Pathway overrepresentation analysis of differentially activated proteins between IBS-Alone and IBS-Multimorbid patient models revealed several enriched signalling pathways, particularly those related to insulin and immunity.

Proteins with a higher different activation pattern between IBS-Alone and IBS-Multimorbid patient models include: Protein A, Protein B, Protein C and Protein D. Previous associations with IBS have been described specially for Protein D and Protein A.

The relevant pathways and proteins identified in this study will be further explored in the analysis of data derived from other work packages of this project.