bioRxiv preprint doi: https://doi.org/10.1101/751289; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Singular manifolds of proteomic drivers to model the evolution of inflammatory bowel disease status

Ian Morilla[a,b,1], **Thibaut Léger**[a], **Assiya Marah**[c], **Isabelle Pic**[c], **and Eric Ogier-Denis**[a]

[a]INSERM, Research Centre of Inflammation, laboratoire d'excellence Inflamex, BP 416, Paris, France
[b]Université Paris 13, Sorbonne Paris Cité, LAGA, CNRS (UMR 7539), laboratoire d'excellence Inflamex, F-93430, Villetaneuse, France
[c]Inception IBD Inc., Montreal, Canada

The conditions who denotes the presence of an immune disease are often represented by interaction graphs. These informative, but complex structures are susceptible to being perturbed at different levels. The mode in which that perturbation occurs is still of utmost importance in areas such as reprogramming therapeutics. In this sense, the overall graph architecture is well characterise by module identification. Topological overlap-related measures make possible the localisation of highly specific module regulators that can perturb other nodes, potentially causing the entire system to change behaviour or collapse. We provide a geometric framework explaining such situations in the context of inflammatory bowel diseases (IBD). IBD are important chronic disorders of the gastrointestinal tract which incidence is dramatically increasing worldwide. Our approach models different IBD status as Riemannian manifolds defined by the graph Laplacian of two high throughput proteome screenings. Identifies module regulators as singularities within the manifolds (the so-called singular manifolds). And reinterprets the characteristic IBD nonlinear dynamics as compensatory responses to perturbations on those singularities. Thus, we could control the evolution of the disease status by reconfiguring particular setups of immune system to an innocuous target state.

inflammatory bowel disease | wgcna | singular manifolds | revertible immune dynamic | compensatory perturbations

Correspondence: ***morilla@math.univ-paris13.fr***

Introduction

The response of a living system facing threats is a decision-making process depending on multiple factors. Some of those factors such as limited resources or energetic cost shape the different phenotypic status of a disease. Hence, the identification of "transition gates" (hereinafter referred as singularities) between those phenotypic phases opens a path to eventual therapeutic interventions with the aim of reconfiguring the system to a normal status. In particular, these status can be expressed as different grades of inflammation when the immune system of the gastrointestinal tract reacts to the presence of harmful stimuli. If the inflammatory conditions of the colon and small intestine become chronic, then, they are all generally grouped under the heading of Inflammatory bowel disease (IBD). IBD is an intestinal disease of unknown cause whose prevalence is in continuous growth at present. Crohn's disease (CR) and ulcerative colitis (UC) are the main variants of IBD. UC, for instance, is characterised by chronic inflammation and ulceration of the lining of the major portion of the large intestine (colon). According to the European Medicines Agency UC presents a prevalence of 24.3 per $100,000$ person-years in Europe. That means there are between 2.5 and 3 million people having IBD in the European Union (1) and this figure could be increased to 10 million worldwide in 10 years. The majority of patients are diagnosed early in life and the incidence continues to rise with $178,000$ new cases of UC each year; therefore, the particular effect of UC on health-care systems will rise exponentially. IBD is a continuous threatening state that may produce aberrant cell proliferation leading to broad epithelial alterations such as dysplasia. This scenario of chronic active inflammation in patients with UC increases the risk for the development of colorectal carcinoma (CRC) and often requires total colectomy in case of intensive medical treatment failure or positive by high-grade dysplasia. Thus, detecting and eliminating or even reverting precursor dysplastic lesions in IBD is a practical approach to preventing the development of invasive adenocarcinoma. Currently, practitioners make decision on what new therapy to be used in such cases based only on their grade of expertise in inflammatory domains. Due to the limited and largely subjective knowledge on this pathology, we were encouraged to seek the biomarker(s) whose symbiotic actions influence the molecular pathogenesis of the risk for colorectal cancer in IBD. In this work, we develop a framework for identifying those dynamical players by abstracting the IBD progression. Since this disease can be naturally observed under the prism of a "phase transition" process (2), we sample the expression profiles of patients from a manifold with singularities; evaluating the functions of interest to the IBD status geometry near these points. Given this assumption, we construct weighted protein co-expression graphs of each variant of the disease by means of a proteomic high-throughput screening consisting of two cohort of 24 patients each (discovery and validation). Next, we localise key players of any identified modules (3) that are relevant to the IBD status, i.e., control, active and quiescent. Then, we use functions associated to the eigengenes of selected proteins across patients (3) to describe the potential of protein expressions with respect to the disease status. And we lay emphasis on the behaviour of the graph Laplacians corresponding to points at or near singularities, where different transitions of disease come together. This scenario enables the identification of potential drug targets in

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/751289; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

a protein-coexpression graph of IBD, accounts for the nonlinear dynamics inherent to IBD evolution and open the door to its eventual regression to a controlled trajectory (4–6). Overall, this manuscript envisages providing clinicians with useful molecular hypotheses of dysplastic regulation and dynamic necessaries prior to make any decision on the newest course of the treatment of individual patients in IBD. And ultimately accelerate drug discovery in health-care system.

Results

WGCNA Identifies Novel Immune Drivers Causing Singularities in the Status of Disease Progression. Intuitively, one might envisage the progression of a disease as a set of immune subsystems influencing each other as response to an undesired perturbation of the normal status. In this exchange, there exist specific configurations that cause the entire system changes its behaviour or collapse. We were, then, interested in identifying the potential modulators of IBD state whose interaction may explain the disease progression as a system instead of simply investigating disconnected drivers dysregulated in their expression levels. To this end, we provide protein significance measures of protein co-expression graphs of each type of IBD disease (CR and UC). Furthermore, those are simultaneously topological and biologically meaningful measures since they are defined by $cor|x_i, S|^{\xi}$ with $\xi \geq 1$ and based on the clinical outcomes of two proteomic samples (SI Text) that capture the IBD phenotype or status, i.e., control, active and quiescent. We fixed this status as a quantitative trait defined by the vector $S = (0, 1, -1)$. And applied the Weighted Gene co-Expression Network Analysis (7, 8) between the two samples, herein considered as discovery and validation cohorts. For the sake of clarity, we only show the results obtained for the UC graph (Fig. 1). The matched inspection of its expression patterns in connectivity (Fig. 1A), hierarchical clustering of their eigengenes (9) and its eigengene adjacency heatmap (Fig. 1B-D) suggests that the most correlated expression patterns with the IBD status (Fig. 1E) are highlighted in greenyellow (97 proteins) and green (215 proteins) respectively. Whereas in CR those coloured in magenta (123 proteins) and midnightblue (41 proteins) yielded the highest correlations (Fig. S1). Nevertheless, we only kept green (UC) and magenta (CR) patterns since the others were not well preserved in the graph corresponding to the validation cohort (Fig. 2B-C and Fig. S2). Next, we wonder about the biological functions these patterns of similar protein expression to the IBD status were enrich of. To response this question, the weighted co-expression subgraph of the green (resp magenta) pattern was interrogated (Fig. 1F) using GO (10). As we expected, the green and magenta expression patterns present overabundance of IBD-related with multiple processes that are essential for the disease progression (Table S1) such as innate immune response in mucosa, positive regulation of B cell proliferation (Fig. 3C) or inflammatory response (Fig. 3E). Complementary, pathways highly related to the disease progression such as intestinal immune network for IgA production (hsa04672) (11) or known pathways such

Table 1. Manifolds incidence θ, kernel search ϕ_κ and eigengene form $f_i(x)$. First row corresponds to the optimal selection.

θ	ϕ_κ	$f_i(x)$
$\frac{cos^{-1}\left(\frac{A \cdot B}{\lVert A \rVert \lVert B \rVert}\right)}{\pi}$	$exp\left(\frac{-\lVert x_i - x_j \rVert^2}{2\sigma^2}\right)$	$\phi_S(rsv\{svd(cor\|x_i, S\|^{\xi})\})$
$\int_{-\infty}^{\infty} p_A(x) log\left(\frac{p_A(x)}{q_B(x)}\right) dx$	$tanh(\frac{1}{N} x_i \cdot x_j + \xi)$	$\phi_S(rsv\{svd(cor\|x_i, S\|^{\xi})\})$
$\int_{t_0}^{t_f} l_i(\tau) \bar{l}_j(\tau) d\tau$	$1 - \frac{\lVert x_i - x_j \rVert^2}{\lVert x_i - x_j \rVert^2 + \xi}$	$\phi_S(rsv\{svd(cor\|x_i, S\|^{\xi})\})$

Inflammatory bowel disease (hsa:05321) are also found (Table S1). In addition, the intersection of dysregulated protein sets involved in such pathways between UC and CR is very low (Fig. 3 and Fig. S3). As positive control the differential expression of well-known proteins participating in IBD such as CAMP or LYZ in UC and LCN2 or IFI16 both in UC and CR are also detected. In particular, the set composed by the proteins STAT1, AZU1, CD38 or NNMT in UC or DEFA1, IGHM, PGLYRP1 and ERAP2 in CR are robustly associated with the status of the disease (i.e. $p \sim O(e^{-3})$). These nodes, and other frequently-occurring nodes such as SYK and CD74, are attractive candidates for experimental verification. Some of these proteins work in tandem, with control sets formed by S100A9 and S100A8 identified in innate immune response process. Strikingly, some proteins such as CTSH presented in processes such as adaptive immune response or regulation of cell proliferation show similar expression between active and quiescent UC status (Fig. S4). Moreover, in CR the regulation of immune response process displayed similar expression in the active and quiescent status for all its proteins mostly belonging to the immunoglobulin heavy variable protein family that participate in the antigen recognition (Fig. S5). It should be also pointed that proteins more expressed in the quiescent than in the active status such as IGHV3-7 or IGKV2 are mainly identified during the CR progression (Fig. S6).

In the light of our computational predictions the realisation of short-term actions, sometimes pharmacologically preferable to permanent changes, on the identified novel immune drivers of IBD status can be effective. Remarkably, an experimental identification of such modifications would be not tractable otherwise.

IBD Progression Can Be Geometrically Interpretable as the Intersection of Manifolds with Boundaries. Now, we enhance the important role the geometry plays in interpreting the nature of the IBD dynamics. Notably, the geometry of IBD status can be learnt using the differential structure defined by those proteins selected as candidates in an on top space. Each IBD status is abstracted as a Riemannian manifold, and construct the graph Laplacian based on these same candidates to figure out how the system gets cross from one to other status. Hence, we describe the domain of IBD progression as intersection of three manifolds, i.e., Ω_c, Ω_q and Ω_a (Fig. S7). Then, the eigengenes associated to the candidate proteins (Fig. S8 and S9) are taken as points at or near a set of singularities in the manifolds intersection. And, later, we inspect the behaviour of their graph Laplacian to define a potential where the dynamic of IBD status through the mani-

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/751289; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

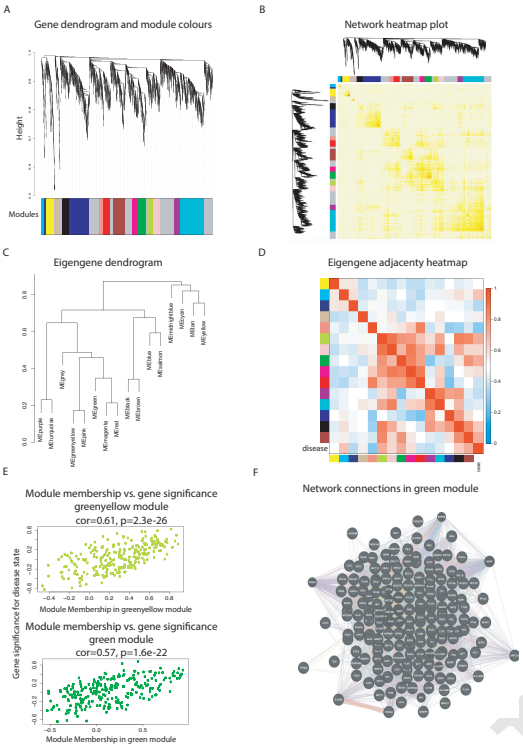

Fig. 1. Overview of the protein coexpression network analysis in UC. (A) Hierarchical cluster tree of the $3,910$ proteins analysed. The colour bands provide a simple visual comparison of module assignments (branch cuttings) based on the dynamic hybrid branch cutting method. Modules in grey are composed by "housekeeping" proteins. (B) Topological Overlap Matrix (TOM) plot (also known as connectivity plot) of the network connections. proteins in the rows and columns are sorted by the clustering tree. Light colour represents low topological overlap and progressively darker red colour represents higher overlap. Clusters correspond to squares along the diagonal. The cluster tree and module assignment are also shown along the left side and the top. (C) Hierarchical clustering dendrogram of the eigengenes based on the dissimilarity $diss(q_1, q_2) = 1 - cor(E^{(q_1)}, E^{(q_2)})$. (D) Visualisation of the eigengene network representing the relationships among the modules and the disease status. The eigengene adjacency $A_{q_1,q_2} = 0.5 + 0.5cor(E^{(q_1)}, E^{(q_2)})$. (E) Protein significance versus module membership for disease status related modules. Both measurements are highly correlated reflecting the high correlations between the IBD progression and the respective module eigengenes (i.e. greenyellow and green). (F) Graph of the green module enriched with subgraphs functionally involved in IBD progression.

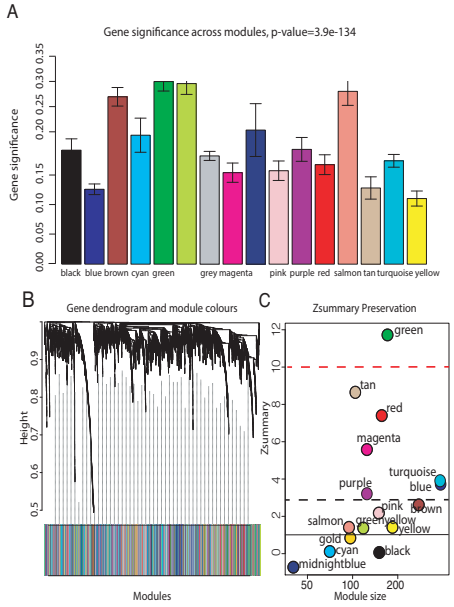

Fig. 2. Significance and preservation of UC graph modules. (A) The module significance (average protein significance) of the modules. The underlying protein significance is defined with respect to the patient disease status (B) The consensus dendrogram for discovery and validation RCH coexpression graphs. (C) The composite statistic $Zsummary$ (Eq.9.1 in (12)). If $Zsummary > 10$ there is strong evidence that the module is preserved (13). If $Zsummary < 2$, there is no evidence that the module preserved. Note that Zsummary shows a strong dependence on module size. The green UC module shows high evidence of preservation in the validation UC graph.

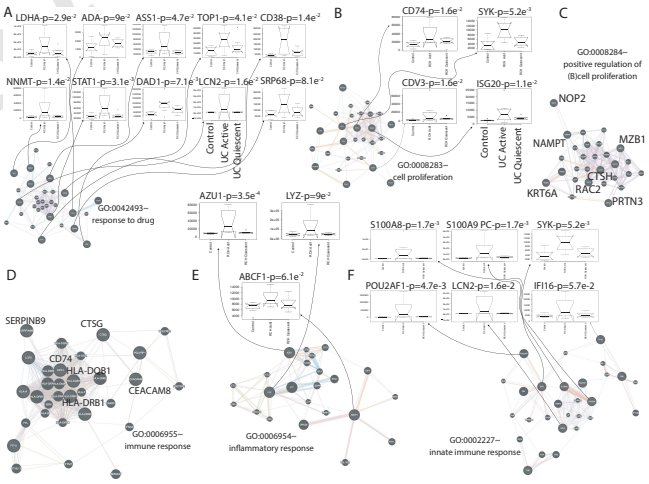

Fig. 3. Representative repertoire of graphs by enriched functions that are well preserved in UC. (A) Response to drug. (B) Cell proliferation. (C) Positive regulation of B cell proliferation. (D) Immune response. (E) Inflammatory response. (F) Innate immune response. The protein interaction graphs were constructed using Genemania (14). Edge colouring of the graphs amount to: purple stands for Co-expression, orange for Predicted, light-blue for Pathway, light-red for Physical interactions, green for Shared protein domains and blue for Co-localisation. Boxplots of expression for the most attractive drivers respect to UC status are indicated by floating arrows at the top of each subgraphs. Initial P upon the boxplot title amount to p-value associated to the IBD status correlation, whereas PC stands for positive control, i.e., proteins already described as IBD related. For the sake of simplicity, we highlighted some candidates simply by its identifiers.

bioRxiv preprint doi: https://doi.org/10.1101/751289; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

folds maybe modelled.

Let $L_{n,h}$ be the graph Laplacian appropriately scaled by $\frac{1}{\sqrt{h}}$ and constructed from the data of 38 candidate proteins selected in UC (resp. 19 in CR), using the Gaussian kernel with bandwidth h (SI Text). Yet, we identify the cross from control to active status of the disease with an intersection-type singularity since it can be naturally considered a "phase transition" as described in *Belkin et al.* *(2, 15)*. Whereas the active-quiescent pass is interpreted as an edge-type singularity since the manifold sharply changes direction (i.e., from disease to control-like status). We are particularly interested in the former scenario, which involves the intersection of the two different manifolds Ω_c and Ω_a. Thus, for a given point $x_1 \in \Omega_c$ consider its projection x_2 onto Ω_a and its nearest neighbour x_0 in the singularity. If n_1 and n_2, are the directions to x_0 from x_1 and x_2 respectively and D_1 and D_2 are the corresponding distances calculated as the Kullback-Leibler divergence (16) between each status. $L_h f(x_1)$ can be approximated by $\frac{1}{\sqrt{h}}\Phi_1(\frac{D_1}{\sqrt{h}})\partial n_1 f(x_0) + \frac{1}{\sqrt{h}}\Phi_2(\frac{D_2}{\sqrt{h}})\partial n_2 f(x_0)$. Note Φ_1,Φ_2 are scalar functions whose form and type singularity correspond. Both functions are explicitly calculated in 1 for the intersection-type singularity. To perform this calculation, we apply Theorem 2 pg. 37 in (2) first to establishing the conditions to analyse the behaviour of the graph Laplacian near the intersection of the two 8-manifolds $\overline{\Omega}_c$ (i.e. the interior of the smooth Ω_c eventually with boundary) and $\overline{\Omega}_a$ embedded in $\mathbb{R}^{24}$. Their intersection $\Omega_c \cap \Omega_a$ is a smooth manifold of dimension l (≤ 7). Let f be a continuous function over $\overline{\Omega} = \overline{\Omega}_c \cup \overline{\Omega}_a$ whose restriction $f_i := f|_{\overline{\Omega}_i}$ on $\overline{\Omega}_i$, $i \in \{c,a\}$, is C^2-*continuous*. Given a $x \in \Omega_c$ near the intersection, let x_0 be its nearest neighbour in $\Omega_c \cap \Omega_a$, and x_1 (resp. x_2) be its projection in the tangent space of $\overline{\Omega}_c$ at x_0 (resp. in the tangent space of $\overline{\Omega}_a$ at x_0). Put $\|x - x_0\| = r\sqrt{h}$. For a sufficiently small h, we have

$$L_h f(x) = \frac{1}{\sqrt{h}}\pi^{\frac{8}{2}} r e^{-r^2 sin^2\theta} p(x_0)(\partial n_1 f_1(x_0) + cos\theta \partial n_2 f_2(x_0)) + o\Big(\frac{1}{\sqrt{h}}\Big), \quad \textbf{(1)}$$

where n_1 and n_2 are the unit vectors in the direction of $x_0 - x_1$ and $x_0 - x_2$, respectively, and θ is the angle between n_1 and n_2 measured as the disease incidence during the cohorts recruitment (see Table 1). From this, we can see that for a point x on the intersection or near intersection, $L_h f(x)$ is of the form $\frac{1}{\sqrt{h}} C r e^{-r^2}$ where the coefficient C depends on the derivatives of f and the position of x.

This calculation provides us with a proper IBD potential (Fig. S10), describing the control-active set, that will be used in further dynamical enquiries.

Therapeutic reconfiguration of the IBD Complex Space. Following Cornelius et al. (4), we wanted to identify temporary actions in the expression of the selected therapeutic targets that potentially control the dynamics of IBD

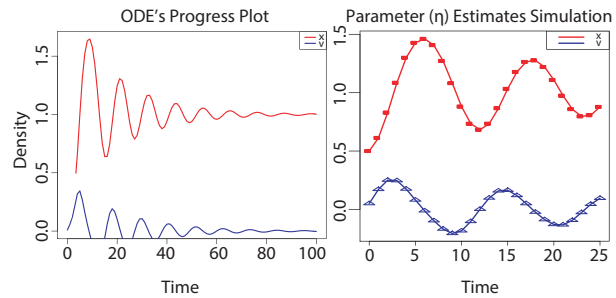

Fig. 4. Fitting our model to data. (A) Time plot of the ODE's system associated to IBD status. (B) Curves show the model for the best estimated parameters released upon a 100 bootstrap iterations, and the symbols depict the data.

status. To this end, we construct a control perturbations in an eight-dimensional system. This consists of a particle defined as the eigengene associated with one of the selected proteins across patients in their corresponding manifold of definition, i.e., Ω_c, Ω_q and Ω_a. We evaluate, then, its expression in function of the potential calculated in the previous section. The system of ordinary differential equations (ODEs) describing this particle at or near an intersection-type singularity as introduced above maybe simplified as:

$$\begin{aligned} P(\overline{x}) &= L_h f(x) \\ F(\overline{x}, v) &= -dL_h f/dx + dissipation, \\ dissipation &= -\eta * v \end{aligned} \quad \textbf{(2)}$$

Right-hand side of ODEs defining system, according to Newton's second law, i.e., if $y = (x, v)$:

$$\begin{aligned} dx/dt &= v \\ dv/dt &= F(x, v) \end{aligned} \quad \textbf{(3)}$$

Before solving the systems of ordinary differential equations defined in 3, we optimised the dissipation parameter η (Fig. 4) by performing a 100 bootstrap simulations (17). With $\eta = 0.1$, the system has two stable fixed points, one at $x < 0$ and the other at $x > 0$ (with $dx/dt = 0$ and $sd \sim O(6.5e^{-8})$). If we continue the steady state of the system from a starting point near to the origin (i.e. 0.075), there exists a bifurcation in -0.08 that well separates the two "basin of attraction" x_D and x_C representing the bistable domain (Fig. 5) of the IBD status (SI Text). This scenario holds the non-linear dynamics implicit in the progression of IBD. Importantly, it captures the iterative interventions on the expressions of the previously selected proteins required to effectively brings the system to a non-active status from an uncontrolled path of the IBD status. Let y_D and y_C be the positions of stable fixed points minimising $L_h f(x)$ at x_D and x_C respectively. We first fix an initial state near x_D to take then a state in the basin of y_D, and try to drive it to the basin of y_C. This resulted in a pertinent class of control perturbations highlighted as red arrows on the left hand side of Fig. 5. Complementary, we also calibrated the class of compensatory perturbations causing the dynamic of a state on an unbounded orbit,

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/751289; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

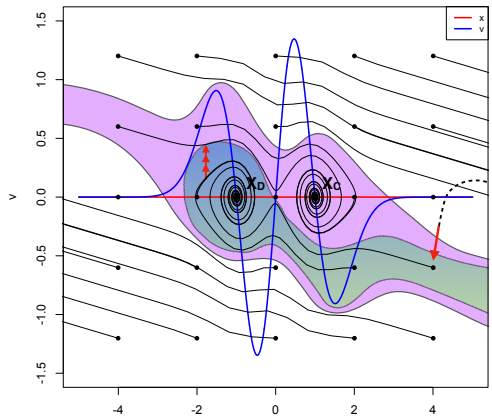

Fig. 5. Illustration of the control process in two dimensions. Light green and violet represent the basins of attraction of the stable states x_D (Disease) and x_C (Control or Latent), respectively, while white corresponds to unbounded orbits. (left and right hand side red arrows) Iterative construction of the perturbation for an initial state in the basin of x_D with x_C as a target (Left hand side), and for an initial state on the right side of both basins with x_D as a target (Right hand side). Dashed and continuous lines indicate the original and controlled orbits, respectively. Red arrows indicate the full compensatory perturbations. Individual iterations of the process are shown in the insets (for clarity, not all iterations are included).

i.e. $x \to +\infty$, be driven onto the basin of y_C. Similarly to the previous case, this class was also represented by a red arrow, but this time on the right hand side of Fig. 5. Specifically, we find that we are able to rescue the same pre-active or quiescent status above with an average distance in norm of $O(e^{-20})$ of a feasible target status. These interventions affect a small number of proteins that in turn are multi-target, which is highly desirable provided IBD status progression is believed to be in a multi-facet cellular components synchrony. The dynamics of UC and CR share a unique pattern, but involving a few different proteins in the reconfiguration of their systems. Consequently, we are able to program actions on promising therapeutic targets what is unaffordable from an experimental point of view by systematic screenings; what enhances the potential of our methodology.

Conclusions

We introduce a systematic strategy to identify potential immune drivers whose variation in expression can explain the different status displayed in the evolution of two cohorts of IBD patients. To this end, we model IBD status by the intersection of special geometric varieties called manifolds. And leverage the graph Laplacian to identify points, on or nearby their intersection, with highly specific module regulators of protein co-expression graphs. These graphs were constructed based on high-throughput proteome screenings of the two samples of IBD patients, i.e., discovery and validation cohorts. Then to make this methodology more biological meaningful, we can test our predictions by means of experiments that study how specific interventions influence the reprogramming of IBD status, either via high-throughput sequencing surveys of validation populations (18) or by immunofluorescence specific to given antibodies (19). The comparison between theory and experiment will provide insight into the functional constraints of immune system in the recognition of bio-drivers varying when facing an intestinal chronic inflammatory threat.

The continuous trade-off amongst the available resources in living systems determines, in a certain way, their response to many situations of stress. To put this mechanism of response in motion, organisms tend to deploy a large diversity of components, such as cell types or proteins (20, 21), each sensitive to a small section of their domain. For example, the colon supports the interplay between reactive oxygen and nitrogen species overproduction or cytokines growth factors, that collectively represent a pivotal role of behaviourally aspects in IBD-induced carcinogenesis (22). Likewise, the role of the binomial composed by the innate and adaptive immune system in IBD therapies involves dozens of feedback loops invoking and sustaining chronic inflammation. However, how the immune system sparks a particular response to repel an IBD threat remains confusing, though it is thought to be immune and non-immune based, with an accepted role of the gut microbiota and non-immune derived cells of the inflammatory cascade including chemokines and inflammasomes (23). In this case, the multifaceted information process limits the repertoire of the immunological machinery. To deal with those specific environmental forces, living systems wisely prioritise their resources in accordance with their costs, and constraints (24). In this work, we have shown how the immune system response in IBD is subjected to such combination of elements, what could fix the status of IBD during its evolution. Our finding reproduces an optimal framework to detect novel immune driver-type specific to IBD status and relates it to the concept of their non-linear dynamics nearby singular topological settings (25). In this context, limiting regions of phenotypic space are modelled by means of Riemannian manifolds revealing themselves suitable to reflect the important competition between resources and costs when that eventually exceeds a vital threshold in IBD. In general this unbalanced response forces the system to drive trajectories of IBD patients to undesirable status of disease (26), our model would lead those trajectories to a region of initial conditions whose trajectories converge to a desirable status –similar to the "basin of attraction" introduced in (4). The connection between the identified immune drivers-type specific and their implication in the evolution of IBD network status becomes even clearer analysing the results yielded by our dynamical model where the pass from one to other status could explain the synergy between innate/adaptative immune resources and their energetic cost and grow in relation to their success in securing resources. Although this study is a characterisation by oversimplification of the adaptive immune system detecting dysplastic lesions in IBD, we expect that our methodology and results will be instrumental also for other diseases and thus have a more wider application for the biomedical field and associated health care systems.

Materials and Methods

The calculations related to these sections were implemented using scripts based on R for weighted graphs analysis (wgcna package

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/751289; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(27)), in-house Matlab© (2011a, The MathWorks Inc., Natick, MA) functions for the analysis of singularities on manifolds and Python for nonlinear optimisation of control perturbations.

Data. A total of 48 IBD patients were distributed in two cohorts of 24 samples each composed by 8 patients stratified by status of disease, i.e., control, active and quiescent.

Proteomic Screening. LC-MS/MS acquisition in samples of $30\mu g$ of $3,910$ proteins was prepared from the groups of patient biopsies running on a NUPAGE 4-12% acrylamide gel (Invitrogen) and stained in Coomassie blue (Simply-blue Safestain, Invitrogen). Peptides and proteins identifications and quantification by LC-MS/MS were implemented by Thermo Scientific, version 2.1 and Matrix Science, version 5.1.

WGCNA. We adopted the standard flow of WGCNA (7) to constructing the protein graphs of UC and CD, detecting protein modules in term of IBD status co-expression and detecting associations of modules to phenotype i.e., control, active and quiescent disease with a soft-threshold, ξ, determined according to the scale-free topology criterion (SI Text). Gene ontology analyses coupled with bioinformatics approaches revealed drug targets and transcriptional regulators of immune modules predicted to favourably modulate status in IBD.

Limit Analysis of Graph Laplacian on Singular Manifolds. It is known by generalisation of Fourier analysis (28) that graph Laplacian converges, in the interior points of its domain, to the Laplace-Beltrami. Moreover, the eigenfunctions associated to this operator act as natural basis for the L_2 functions used to represent data on the manifold (29). The points nearby phenotypic changes of the disease space are not interior though. Hence, we must draw upon the definition of limit from (2) if we want to analyse the behaviour of our infinite graphs Laplacian. Specifically, $L_{n,h}, n = \infty$, $L_h f(x)$ in UC and CD, when x is on or nearby a singular point, h is small and the function $f : \overline{\Omega} \to \mathbb{R}$, is fixed.
For a fixed h we define L_h as the limit of $L_{n,h}$ as the amount of data tends to infinity:

$$L_h f(x) = L_{\infty,t} f(x) = \mathbb{E}_{p(X)}[L_{h,t} f(x)] f(x) = \frac{1}{h} \int_{\overline{\Omega}} K_h(x,y)(f(x) - f(y))p(y)dy. \quad (4)$$

The graph Laplacian is scaled by the Gaussian kernel K_h with bandwidth h. Note that $p(x)$ is defining a piecewise smooth probability density function on $\overline{\Omega}$. Such distribution is composed by i.i.d. 48 random samples $\mathcal{P} = \{X_1, \ldots, X_{48}\}$ corresponding to the IBD patients. Thus, if we apply to 4 the analysis already described in Section 2 of results in our intersection-type singularities on each IBD status manifold, we deduce 1.

Nonlinear Optimisation. We learn from (4) how to optimise the interventions set needed to control the evolution of our disease model. This control procedure is iterative and consists of minimising the residual distance between the target state, x^*, and the system path $x(t)$ at its time of closest approach, t_c. To ensure the existence of admissible perturbations in the system herein represented by the vector expressions 6,7 and also to limit the magnitude of the solution δx_0 of the optimisation problem 8, some few constraints must be introduced (SI Text). Then finding the particular solution, δx_0, becomes a nonlinear programming problem (NLP) that can then be properly defined as:

$$\min \quad |x^* - (x(t_c) + M(x_0', t_c) \cdot \delta x_0)| \quad (5)$$

$$\text{s.t.} \quad g(x_0, x_0' + \delta x_0) \leq 0 \quad (6)$$

$$h(x_0, x_0' + \delta x_0) = 0 \quad (7)$$

$$\epsilon_0 \leq |\delta x_0| \leq \epsilon_1 \quad (8)$$

$$\delta x_0 \cdot \delta x_0^p \geq 0 \quad (9)$$

Where the matrix $M(x_0; t)$ is the solution of the variation equation $dM = dt = DF(x) \cdot M$ subject to the initial condition $M(x_0; t_0) = 1$. And δx_0^p denotes the incremental perturbation from the previous iteration.

ACKNOWLEDGEMENTS
We acknowledge the financial support by Institut National de la Santé et de la Recherche Medicale (INSERM), Inception IBD, Inserm-Transfert, Association Franois Aupetit (AFA), Université Diderot Paris 7, and the Investissements d'Avenir programme ANR-11-IDEX-0005-02 and 10-LABX-0017, Sorbonne Paris Cité, Laboratoire d'excellence INFLAMEX.

Bibliography

1. J Burisch, T Jess, M Martinato, and PL Lakatos. The burden of inflammatory bowel disease in europe. *Journal of Crohn's and Colitis*, 7:322–337, 2013.
2. M Belkin, Q Que, Y Wang, and X Zhou. Toward understanding complex spaces: Graph laplacians on manifolds with singularities and boundaries. *MLR: Workshop and Conference Proceedings; 25th Annual Conference on Learning Theory*, 36:1–24, 2012.
3. S Horvath and J Dong. Geometric interpretation of gene coexpression network analysis. *PLoS Computational Biology*, 4, 2008.
4. SP Cornelius, WL Kath, and AE Motter. Realistic control of network dynamics. *Nature Communications*, 4(1942), 2013.
5. SP Cornelius, WL Kath, and AE Motter. Controlling complex networks with compensatory perturbations. *arXiv*, 1105(3726):1–20, 2011.
6. J Sun, SP Cornelius, WL Kath, and AE Motter. Comment on "controllability of complex networks with nonlinear dynamics". *arXiv*, 1108(5739):1–12, 2011.
7. B Zhang and S Horvath. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005.
8. P Langfelder and S Horvath. Eigengene networks for studying the relationships between co-expression modules. *BMC Systems Biology*, 1(54), 2007.
9. P Langfelder, B Zhang, and S Horvath. Defining clusters from a hierarchical cluster tree: The dynamic tree cut library for R. *Bioinformatics*, 24(5):719–720, 2007.
10. M Ashburner and al. et. Gene ontology: tool for the unification of biology. *Nat Genet*, 25(1): 25–9, 2000.
11. I Morilla, M Uzzan, D Cazals-Hatem, H Zaag, E Ogier-Denis, G Wainrib, and X Tréton. Topological modelling of deep ulcerations in patients with ulcerative colitis. *Journal of Applied Mathematics and Physics*, 5(11):2244–226, 2017.
12. S Horvath. *Weighted Network Analysis. Applications in Genomics and Systems Biology*. Springer Book, 2011. ISBN 978-1-4419-8818-8.
13. P Langfelder, R Lum, MC Oldham, and S Horvath. Is my network module preserved and reproducible? *PLoS Comput Biol*, 7(1):298–305, 2011.
14. D Warde-Farley, SL Donaldson, O Comes, K Zuberi, R Badrawi, P Chao, M Franz, C Grouios, F Kazi, CT Lopes, A Maitland, S Mostafavi, J Montojo, Q Shao, G Wright, GD Bader, and Morris Q. The genemania prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, 38:214–220, 2010.
15. M Belkin and P Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, June 2003. ISSN 0899-7667. doi: 10.1162/089976603321780317.
16. S Kullback and RA Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79—86, 1951.
17. H Hass, C Kreutz, J Timmer, and D Kaschek. Fast integration-based prediction bands for ordinary diferential equation models. *Bioinformatics*, 32(8):1204–1210, 2016.
18. Amanda E Starr, Shelley A Deeke, Zhibin Ning, Cheng-Kang Chiang, Xu Zhang, Walid Mottawea, Ruth Singleton, Eric I Benchimol, Ming Wen, David R Mack, Alain Stintzi, and Daniel Figeys. Proteomic analysis of ascending colon biopsies from a paediatric inflammatory bowel disease inception cohort identifies protein biomarkers that differentiate crohn's disease from uc. *Gut*, 66(9):1573–1583, 2017. ISSN 0017-5749. doi: 10.1136/gutjnl-2015-310705.
19. S Sedghi, F Barreau, I Morilla, N Montcuquet, D Cazals-Hatem, E Pedruzzi, E Rannou, X Tréton, JP Hugot, E Ogier-Denis, and F Daniel. Increased proliferation of the ileal epithelium as a remote effect of ulcerative colitis. *Inflammatory bowel diseases*, 22(10): 2369—2381, October 2016. ISSN 1078-0998. doi: 10.1097/mib.0000000000000871.
20. JAG Ranea, I Morilla, JG Lees, AJ Reid, C Yeats, AB Clegg, F Sanchez-Jimenez, and C Orengo. Finding the "dark matter" in human and yeast protein network prediction and modelling. *PLOS Computational Biology*, 6(9):1–14, 09 2010. doi: 10.1371/journal.pcbi.1000945.
21. JG Lees, JK Heriche, I Morilla, JA Ranea, and CA Orengo. Systematic computational prediction of protein interaction networks. *Physical Biology*, 8(3):035008, 2011.
22. TR Harris and BD Hammock. Soluble epoxide hydrolase: Gene structure, expression and deletion. *Gene*, 526(2):61 – 74, 2013. ISSN 0378-1119. doi: https://doi.org/10.1016/j.gene.2013.05.008.

DRAFT

bioRxiv preprint doi: https://doi.org/10.1101/751289; this version posted September 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

23. G Holleran, L Lopetuso, V Petito, C Graziani, G Ianiro, D McNamara, A Gasbarrini, and F Scaldaferri. The innate and adaptive immune system as targets for biologic therapies in inflammatory bowel disease. *International Journal of Molecular Sciences*, 18(2020), 2017.
24. F Ungaro, F Rubbino, S Danese, and S D'Alessio. Actors and factors in the resolution of intestinal inflammation: Lipid mediators as a new approach to therapy in inflammatory bowel diseases. *Frontiers in Immunology*, 8:1331, 2017. ISSN 1664-3224. doi: 10.3389/fimmu.2017.01331.
25. AB Goldberg, X Zhu, A Singh, Z Xu, and R Nowak. Multi-manifold semi-supervised learning. In *Journal of Machine Learning Research : Workshop and Conference Proceedings*, volume 5, pages 169–176, 2009.
26. J Sun and AE Motter. Controllability transition and nonlocality in network control. *Phys. Rev. Lett.*, 110(208701), 2013.
27. P Langfelder and S Horvath. Wgcna: an r package for weighted correlation network analysis. *BMC Bioinformatics*, 46(1):559, 2008.
28. EM Stein. *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory. (AM-63)*. Princeton University Press, 1970. ISBN 9780691080673.
29. D Chen and HG Müller. Nonlinear manifold representations for functional data. *Annals of Statistics - ANN STATIST*, 40, 05 2012.

DRAFT