METHODS

Stepwise target controllability identifies dysregulations of macrophage networks in multiple sclerosis

Giulia Bassignana[1,2], Jennifer Fransson[1], Vincent Henry[1,2], Olivier Colliot[1,2], Violetta Zujovic[1], and Fabrizio De Vico Fallani[1,2]

[1]Sorbonne University, UPMC Univ Paris 06, Inserm U-1127, CNRS UMR-7225, Institut du Cerveau et de la Moelle Epinière, Hopital Pitié-Salpêtrière, Paris, France
[2]Inria Paris, Aramis Project Team, Paris, France

ABSTRACT

Identifying the nodes able to drive the state of a network is crucial to understand, and eventually control, biological systems. Despite recent advances, such identification remains difficult because of the huge number of equivalent controllable configurations, even in relatively simple networks. Based on the evidence that in many applications it is essential to test the ability of individual nodes to control a specific target subset, we develop a fast and principled method to identify controllable driver-target configurations in sparse and directed networks. We demonstrate our approach on simulated networks and experimental gene networks to characterize macrophage dysregulation in human subjects with multiple sclerosis.

AUTHOR SUMMARY

We introduce an optimized heuristic, called stepwise target controllability, to quantify the centrality of a candidate driver node to influence the state of a network target set. We use this method to study macrophage gene network alterations in multiple sclerosis. We show that multiple sclerosis is characterized by a global loss of gene coactivation and that this is due to the dysregulation of few molecules along the driver-target pathways. These findings provide new insights into the macrophage network mechanisms underlying the pathophysiology of multiple sclerosis and provide fresh tools for the study of driver-target controllability in complex networked systems.

INTRODUCTION

For many biological systems, it is crucial to identify the units, such as genes or neurons, with the potential to influence the rest of the network, as this identification can enable describing, understanding, and eventually controlling the function of the system (Bonifazi et al., 2009; Jeong, Mason, Barabási, & Oltvai, 2001). Topological descriptors based on network science can indeed be used to quantify such influence in terms of node centrality, such as degree, betweenness, or closeness (Newman, 2010). However, these descriptors capture only the structural properties of the network and neglect their effect on the dynamics, thus limiting our understanding on the actual influencing power.

Control network theory, linking network structure to dynamics through linear or nonlinear models, has been shown to be a more principled approach for identifying driver nodes

in an interconnected system (Rugh & Kailath, 1995; Sontag, 1998). While theoretically these approaches can give a minimum set of driver nodes sufficient to steer the system into desired states, their exhaustive identification might be difficult in practice as there exists in general a very large number of equivalent controllable walks, even in relatively simple networks (Heuberger & Wagner, 2011; S. G. Wagner, 2007). In the case of criteria based on the manipulation of controllability matrices (Hautus, 1970; Kalman, 1963), the presence of many walks can for example induce numerical errors due to the different orders of magnitude in the matrix elements.

Controllability:
Ability to act upon a network for steering its state towards desired configurations.

An alternative solution has recently been proposed to circumvent this limitation, based on the possibility to map the controllability problem onto the maximum cardinality matching over the associated graph (Lin, 1974; Y.-Y. Liu, Slotine, & Barabási, 2011; Shields & Pearson, 1976). As a result, it is possible to identify a set of driver nodes—at least for directed networks—with linear, and not exponential, time complexity (Hopcroft & Karp, 1971). While this approach elegantly solves numerical issues, it can nevertheless not tell which configuration, among all the possible ones, is the most relevant. In general, there is a factorial number of equivalent configurations (with the same number of inputs), and enumerating all possible matchings (Uno, 1997) rapidly becomes unfeasible, even for simple graphs such as trees (Heuberger & Wagner, 2011; S. G. Wagner, 2007), bipartite graphs (Y. Liu & Liu, 2004), or random graphs (Zdeborová & Mézard, 2006). Thus, the research of alternative strategies to characterize the candidate driver nodes is crucial for the concrete application of network controllability tools.

Graph:
Set of nodes and edges representing the units and the links of a network.

However, they do not solve the problem of multiple driver set configurations. On the other hand, technical and experimental constraints often limit the possibility to stimulate many driver nodes in parallel, for example in gene expression modulation (Lodish et al., 2000) or brain stimulation (Hallett, 2000). In these cases, approaches that focus on the ability of single driver node to control the entire network, such as control centrality (Y.-Y. Liu, Slotine, & Barabási, 2012) or single-node controllability (Gu et al., 2015), do circumvent the multiplicity issue, but can still suffer from numerical errors and approximate results.

Driver node:
Node of the network candidate to steer the target sets state.

To overcome this impasse, we propose an integrated method that combines the advantages of the previous approaches and quantifies the capacity of a single driver node to control a predefined target set. Based on the Kalman controllability condition, our method identifies the part of the target set that can be controlled by a candidate driver. To do so, we introduce a ranking among the target nodes and we iteratively evaluate the controllability of the system by adding one target node at a time in a descending order. This eventually finds a univocal controllable configuration corresponding to the highest ranking. In the following, we first illustrate how our method, named *stepwise target controllability*, works for simple network structures and we discuss the potential benefits for directed and sparse networks as compared with alternative approaches. Then, we use it to study molecular networks of macrophage pro-inflammatory activation, derived from ontology-based reconstructions, and identify the driver-target pathway alterations using gene expression data from blood samples of patients affected by multiple sclerosis (MS) and a matched group of healthy controls (HC).

Target set:
Subset of nodes in the network that are chosen to be controlled.

RESULTS

Stepwise Target Controllability Identifies a Controllable Subset of Targets

Let $\mathcal{G}$ be a directed graph (or network) of N nodes (or vertices) and L links (or edges), and $\mathcal{T}$ an arbitrary subset of $S < N$ nodes in the network. The aim is to measure the ability of each node to drive the state of the target set $\mathcal{T}$ from a dynamical system perspective (Sontag,

1998). In the case of linear time-invariant dynamics, the number of controllable target nodes can be obtained by computing the rank of the target controllability matrix (Commault, Van der Woude, & Frasca, 2019; Gao, Liu, D'Souza, & Barabási, 2014; Murota & Poljak, 1990):

$$Q_{\mathcal{T}} = \begin{bmatrix} CB & CAB & CA^2B & \cdots & CA^{N-1}B \end{bmatrix}, \tag{1}$$

where A is the adjacency matrix of the network, B is a vector identifying the driver node, and C is a matrix selecting the rows of A corresponding to the target, or output, nodes (Material and Methods).

Walk:
Alternating sequence of nodes and edges in a network.

A *walk* in the network consists of an alternating sequence of vertices and edges.

The (i, j) entry of $Q_{\mathcal{T}}$ indicates how many walks of length $j - 1$ connect the driver to the target i (Biggs, 1974). Trivially, all the nodes not traversed by these walks do not contribute to the walk lengths and they can be neglected for the purpose of control. By removing the irrelevant nodes from the network, A becomes smaller and this results in a target controllability matrix with fewer columns. Put differently, we avoid the computation of matrix exponentials corresponding to nonexisting driver-target walks (Material and Methods). In practice, this can be of great advantage for reducing the occurrence of round-off errors during the matrix rank calculations. For example, this is the case for sparse and directed networks, where fewer nodes are reachable as compared with dense and undirected networks.

This can be easily appreciated in the following example. Let us consider a directed full binary tree with $h = 6$ levels, with the root node as the candidate driver. Without loss of generality, we randomly position a target in each level and we rank them according to their height in the tree. Then, we introduce a simple cycle among the first three nodes of the tree (Figure S1 in the Supporting Information). By construction, this configuration is controllable and the entire target set can be fully driven by the driver. However, when considering the entire network the returned rank is deficient. Instead, by removing the part of the network that is irrelevant for the control, the rank is full and we retrieve the entire controllable configuration, even in the case of larger networks, that is, up to $h = 10$ levels.

The rank of $Q_{\mathcal{T}}$ gives the number of target nodes $\tau \leq S$ controllable by the driver, but there might be in general many possible equivalent configurations. To overcome this issue, we propose a stepwise procedure that tests the controllability on subproblems of increasing size. First, we introduce a hierarchy among the target nodes and relabel them according to their importance in a descending order, that is, $t_1 \succ t_2 \succ ... \succ t_S$. Then, we create an empty auxiliary set $\mathcal{T}'$ and we sequentially include the target nodes according to their ordering. At each step, if the rank of $Q_{\mathcal{T}'}$ is full, the new target node is retained, otherwise it is removed from $\mathcal{T}'$. When all the target nodes have been visited, the algorithm returns the set of controllable targets with highest ranking (Figure 1; Material and Methods).

Our method, named *stepwise target controllability*, returns for each candidate driver not only the number of controllable target nodes τ corresponding to the configuration with highest ranking, but also the set $\mathcal{T}'$ of controllable targets.

To explore the limits of our method in terms of computational complexity, we performed a simulation analysis on synthetic random networks, which vary in number of nodes, connection density, and target size (Figure S2 in the Supporting Information). Results show that in general our method is able to retrieve a larger number of controllable targets as compared with a standard approach that computes the rank of the full controllability matrix. More specifically, when the target set contains 5% of the network nodes, results are quite stable across different
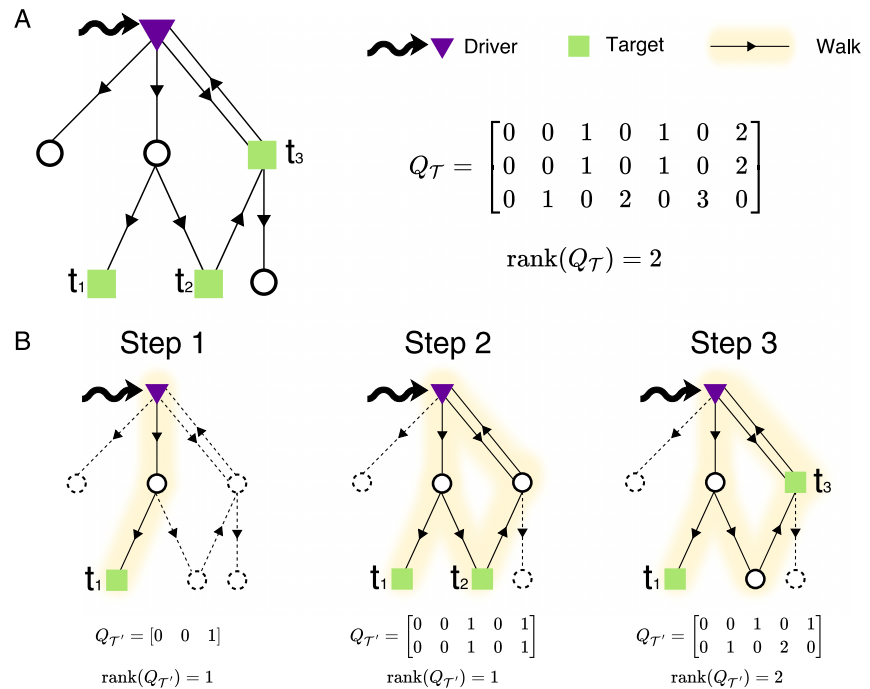
Figure 1. Working principle of stepwise target controllability. (A) A network with one driver and a target set $\mathcal{T} = \{t_1, t_2, t_3\}$ of cardinality $S = 3$. The Kalman condition informs us that only two targets are controllable from the driver, that is, $\tau = \text{rank}(Q_{\mathcal{T}}) = 2$. However, there might be up to three equivalent configurations that are controllable, that is, $\{t_1, t_2\}$, $\{t_1, t_3\}$, and $\{t_2, t_3\}$. For larger networks, the number of Kalman tests to perform can be prohibitive, that is, $\binom{S}{\tau}$. (B) By introducing a hierarchy among the target nodes, our stepwise method identifies the configuration with the most important nodes by performing only S tests (see Material and Methods). In this example, the first step considers the subgraph containing all the walks from the driver to the target set $\mathcal{T}' = \{t_1\}$. The associated controllability matrix has full rank, that is, $\text{rank}(Q_{\mathcal{T}'}) = 1$. The first target is therefore retained and the algorithm moves to Step 2, by constructing a new subgraph containing the walks from the driver to the target set $\mathcal{T}' = \{t_1, t_2\}$. The rank of the new controllability matrix is now deficient and t_2 is not retained. In Step 3, the new subgraph contains the walks from the driver to $\mathcal{T}' = \{t_1, t_3\}$. Because $\text{rank}(Q_{\mathcal{T}'})$ is full and there are no more targets, the algorithm stops and returns the controllable configuration t_1, t_3.

connection densities. For larger target-set sizes, our method works better when the connection density is relatively low (0.02–0.10). It is important also to notice that the computation of the rank starts to fail in correspondence with larger and denser networks (i.e., $N > 180$ and density > 0.20).

Driver Genes Are Homogeneously Distributed in the Macrophage Network

To test our method in a biological context, we construct a network representing the interactions between molecules involved in macrophages response to pro-inflammatory stimuli (Figure 2), with the connections between genes inferred from a previously established network based on literature (Raza et al., 2008; Robert, Lu, Law, Freeman, & Hume, 2011). This network is of interest in MS because of the chronic inflammation characteristic of the disease, and the generally destructive effects of pro-inflammatory macrophages in MS (Bitsch et al., 2000). Hence, dysregulation of macrophages may lead to aggravated inflammation and disease.
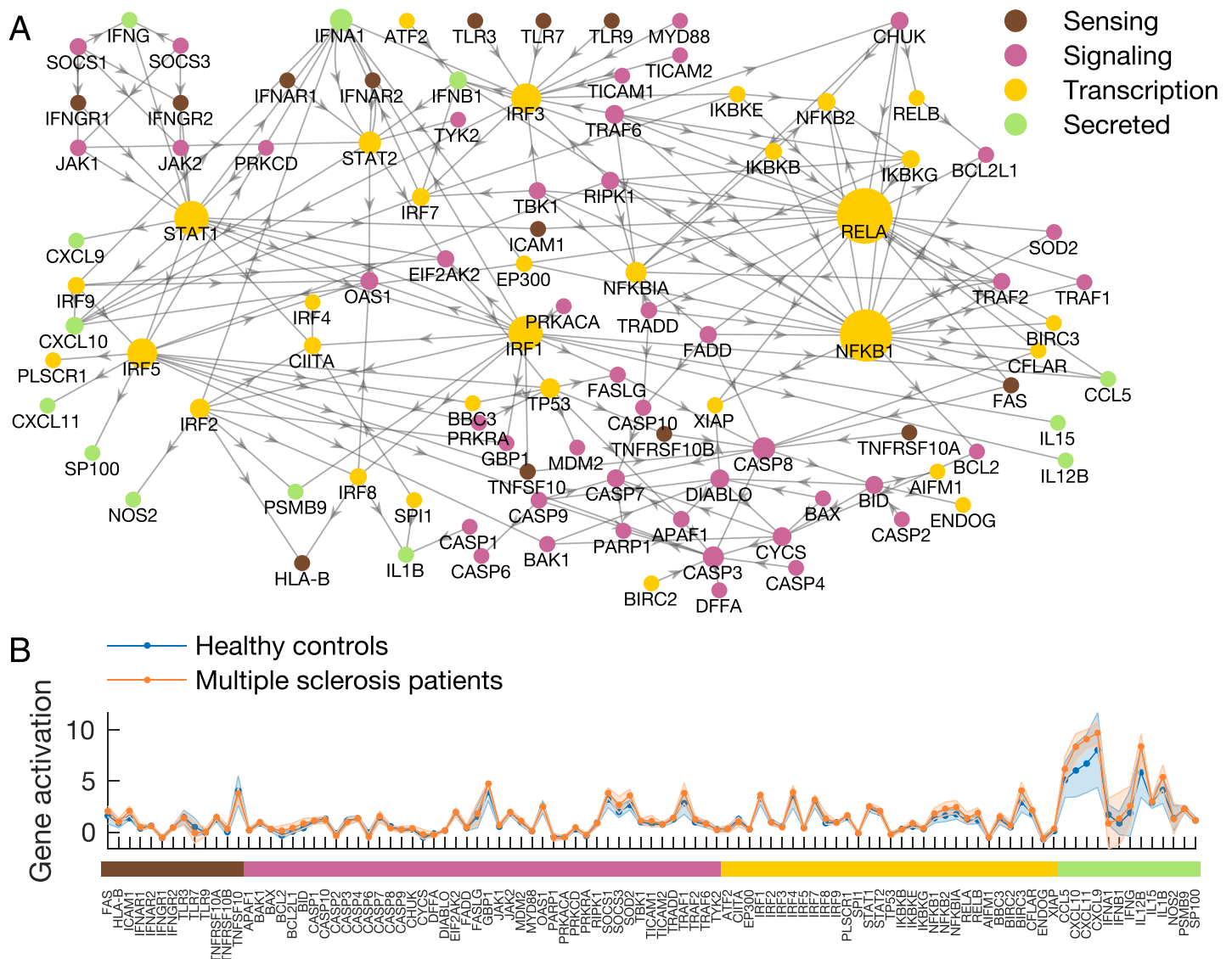
Figure 2. Molecular network and gene activation associated with the pro-inflammatory state of macrophages. (A) The molecular network reconstructed through ontology-based techniques from the macrophage.com repository (Raza et al., 2008; Robert, Lu, Law, Freeman, & Hume, 2011). The network consists of $N = 101$ nodes corresponding to genes involved in inflammation; for the sake of interpretablity, they are organized in four classes, depending on their function in the cell. *Sensing* genes are in the membrane of the cell and start a *signaling* pathway inside the cell, to the *transcription* factors, which promote the production of *secreted* molecules. There are $L = 211$ directed edges representing either activation or inhibition interactions between molecules (Material and Methods). The size of the nodes is proportional to their total degree k. (B) Gene activation computed as the ratio in expression between the "pro-inflammatory" and "alert" states, based on our RNA sequencing data, generated from monocyte-derived macrophages from blood samples of multiple sclerosis patients ($n = 8$) and healthy controls ($n = 8$) (Material and Methods). Solid lines represent group-averaged values, while transparent patches stand for standard deviation.

In order to facilitate biological interpretation of the network, we divide the nodes according to molecular function: *sensing*, *signaling*, *transcription* factors, or *secreted* molecules (Table S1). We choose the 13 *secreted* molecules as target nodes because they represent the end products of macrophage pro-inflammatory activation and enable propagation of inflammation to other cells, thus exacerbating chronic inflammation.

To establish a hierarchy among the targets, we use macrophage RNA expression data from a group of MS patients and healthy controls. The macrophages were tested with and

Fold change:
Ratio between the gene expression of diseased and healthy states.

Node degree:
Number of links of a node.

without activating stimuli to mimic the pro-inflammatory response. We measure the gene activation as the ratio of the expression between the "pro-inflammatory" and "alert" condition. We then consider the fold change Δ between the gene activation of MS patients and HC subjects (Material and Methods). Genes with larger Δ values are ranked first (Figure S3 in the Supporting Information).

Results show that 51% of the tested network nodes can control at least one target (i.e., $\tau > 0$) and that those drivers tend to be homogeneously distributed across classes (Table S2). This indicates a high redundancy in the way the target set can be controlled. Notably, target centrality values are weakly correlated (Spearman rho 0.18, $p < 0.07$) with the corresponding total node degree k, as defined in Material and Methods, indicating that the most connected genes (e.g., RELA, NFKB1) are not necessarily the ones that can most efficiently steer the state of the target set (Figure 3A).

Almost all of the driver nodes identified by our method can control the target genes CCL5, CXCL10, CXCL11, IFNA1, and IFNB1, which code for inflammatory chemokines and cytokines. This implies a high level of coregulation among these molecules, with many different actors exerting control over this regulation. Interestingly, the drivers with the highest target control centrality values (SOCS1 and SOCS3, $\tau = 10$) belong to feedback systems that control
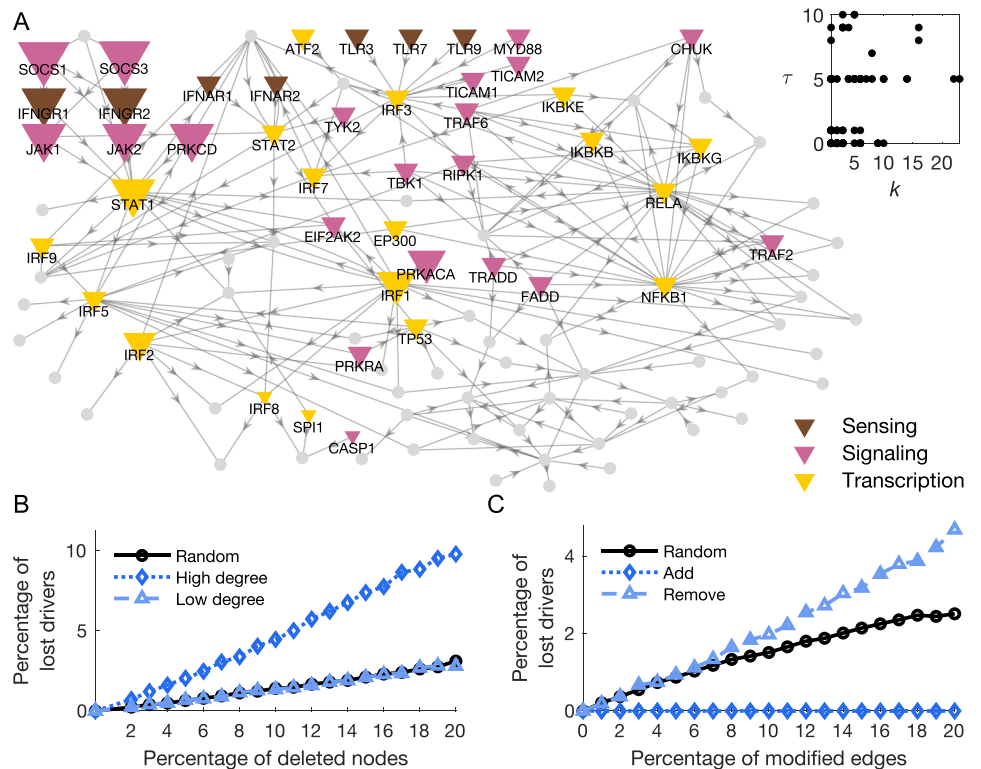
Figure 3. Gene network target control centrality and analysis of robustness for the driver nodes. (A) The size of the nodes codes the stepwise target control centrality values τ. Nodes with $\tau = 0$ are classified as not-drivers and are represented in gray. The inset shows that τ values cannot be merely predicted by node degree k (Spearman rho 0.18, $p < 0.07$). (B) The percentage of driver nodes ($\tau > 0$) that are lost when removing nodes in a random fashion (black circles), or preferentially attacking high-degree (blue diamonds) or low-degree nodes (light blue triangles). (C) The percentage of driver nodes that are lost when randomly rewiring (black circles), adding (blue diamonds), or removing edges (light blue triangles).

pro- and anti-inflammatory signal transduction by regulating the signaling process triggered in response to IFNγ (McCormick & Heller, 2015). In addition, all drivers with $\tau \geqslant 9$ can be seen in our network as a cluster of genes converging onto and including STAT1 (Figure 3A). This cluster includes the receptors of IFNγ and the signaling molecules responsible for their intracellular effects. This result matches the well-described effects of IFNγ on chemokine production (Koper, Kamińska, Sawicki, & Kemona, 2018; J. Liu, Guan, & Ma, 2005) and overall macrophage activation (Mosser & Edwards, 2008).

Robustness of driver nodes to random attacks. To assess the stability of our findings to possible errors in the network construction, we performed a robustness analysis simulating different types of alterations to its nodes and links (Material and Methods). Results show that removing nodes with higher degree k leads to a greater reduction of control centrality in the drivers compared with the removal of low-degree nodes or random removal of nodes (Figure 3B). For
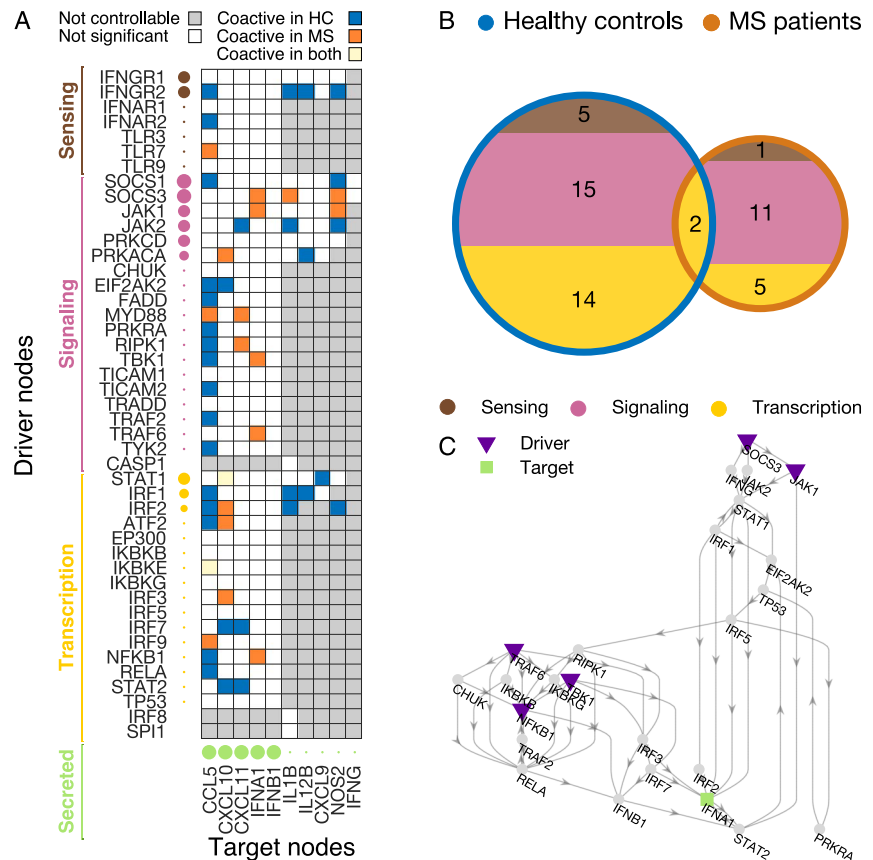
Figure 4. Altered driver-target coactivation in multiple sclerosis. (A) The coactive driver-target pairs, computed as significant Spearman correlations ($p < 0.05$) between the gene activation of controllable driver-target pairs, for the healthy control (HC - blue squares) and the multiple sclerosis group (MS - red squares). White squares indicate that there is a controllable walk from the driver to the target, but that their correlation is not significant. Gray squares mean that there is no controllable walk for driver-target pairs. The size of the circles for driver nodes codes for their target control centrality values τ. For target genes, circle sizes represent the number of driver nodes that can control them. (B) Venn diagram showing a decrease in number of driver-target coactivations in the MS patients as compared with HC. In both groups, these functional interactions tend to predominantly involve signaling genes. (C) Subnetwork of the walks from all the drivers coactivated with the target IFNA1.

example, by attacking 10% of the nodes we lose 5% of the drivers in the latter cases, while we lose 20% of the drivers when removing the most connected ones (Figure 3B). This result confirms the crucial role of hubs in biological networks in terms of resilience to random attacks (Barabási & Oltvai, 2004) and controllability (Pu, Pei, & Michaelson, 2012).

When perturbing links, the worst condition is given by their random removal. By attacking 10% of the links, around 5% of the drivers are lost. This is intuitively due to the interruption of driver-target walks and to consequent impossibility to control a node that cannot be reached. While randomly rewiring the links has an intermediate impact, adding new links has no effect on the target control centrality of the drivers (Figure 3C). This is of great advantage as it shows that our results will not change if new connections are established or provided by the literature.

Gene Dysregulation and Altered Driver-Target Coactivation in Multiple Sclerosis

Using stepwise target controllability, we detect potential directed interactions in the macrophage activation network, but we cannot quantify how changes in the driver's state affect those in the targets. To measure driver-target functional interactions, we compute the Spearman correlation between the gene activation of controllable driver-target pairs, for the HC and MS groups (Material and Methods, Table S3). We call *coactivated* the genes exhibiting a significant correlation ($p < 0.05$). Results show that in general only a moderate fraction (21%) of all the possible driver-target genes are coactivated (Figure 4A, Table S3). For both HC and MS groups these interactions tend to primarily involve signaling functions Figure 4B. However, the number of driver-target coactivations is lower in the pathological condition ($MS = 19$ versus $HC = 36$). More importantly, they differ from those observed in the HC group (Figure 4B). This is particularly evident for target IFNA1, which only exhibits coactivations with signaling and transcription drivers in the MS group (Figure 4C).

Because the macrophage network edges are fixed and reconstructed from known protein-protein interactions, differences in coactivation can be essentially attributed to altered regulation of transcription. Hence, our hypothesis is that the observed functional reorganization
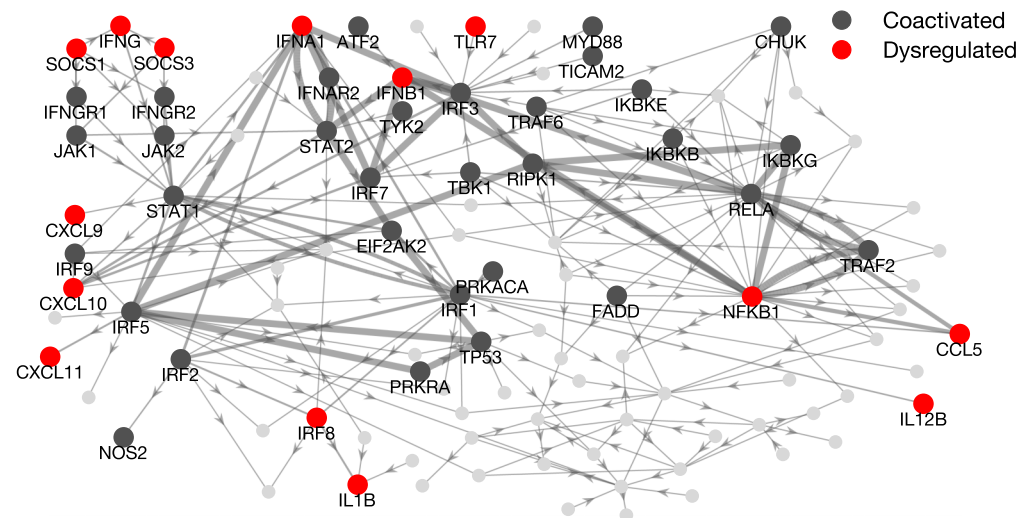
Figure 5. Pooled visualization of dysregulated genes along differentially coactivated driver-target walks. Highlighted genes indicate all nodes on walks between coactivated driver-target pairs, either in the healthy control group, or in the MS patients group. Dysregulated genes are shown in red. Edge thickness is proportional to the number of times they are traversed by walks connecting a driver to a target node (information not reported here).
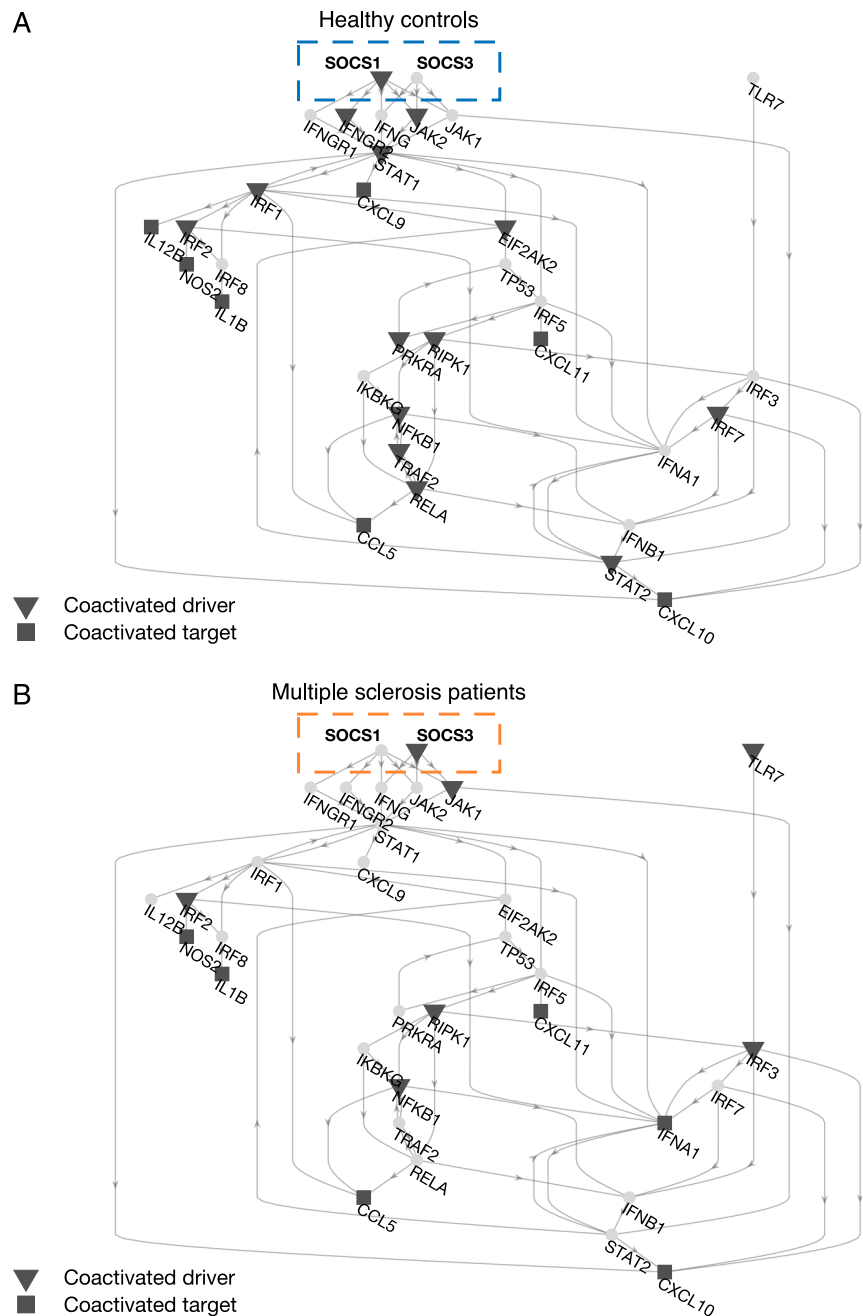
Figure 6. Dyregulated drivers and coactivation switch for SOCS-genes. The subnetwork includes all dysregulated drivers (IRF8, NFKB1, SOCS1, SOCS3, TLR7) and their controllable targets. (A) Coactivated pairs for healthy controls (HC). (B) Coactivated pairs for multiple sclerosis (MS) patients. A coactivation switch can be appreciated between the HC and MS group. SOCS1 and SOCS3 are respectively coactive and silent in HC, while they invert their role in the MS group.

can be explained by the dysregulation of specific genes along the controllable walks from the drivers to targets. To test this prediction, we examine all the pairs of genes whose coactivation appears or disappears in the MS group (Figure 4A). We found that 47/51 of these differentially coactivated pairs present at least one dysregulated gene (i.e., fold change $|\Delta|$ above the 75th percentile, Material and Methods) on the walk from the driver to the target (Figure 5, Table S4).

We find in total 14 dysregulated nodes on any of these walks. The genes that most frequently appear are NFKB1, IFNA1, and IFNB1 (36/51 walks). They are present on all walks that end with targets CCL5, CXCL10, CXCL11, IFNA1, or IFNB1, that is, the five targets that could be controlled by most drivers. This points to their dysregulation being a potent disruptor of the normal network functioning. The co-occurrence of these three dysregulated genes can be explained by a feedback loop in which NFKB1 activates IFNB1, and IFNA1 and IFBN1 both activate STAT2, which through several intermediates can influence all three genes (Figure S4 in the Supporting Information). Indeed, this stems from the fact that all these nodes belong to the main connected component of the network, that is a subnetwork in which every node is reachable from any other node.

Taken together, these results indicate that the aberrant reorganization of functional interactions in the MS group is associated with the presence of dysregulated genes along the controllable walks of the macrophage network.

Switch of SOCS-gene coactive drivers reflects dysregulated inflammatory response. Because drivers are crucial for steering the target network's state, we focus on the subnetwork specifically involving the dysregulated drivers (IRF8, NFKB1, SOCS1, SOCS3, TLR7) and the walks towards the respective controllable targets (Figure 6). By looking at how driver and target nodes are differently coactivated in healthy controls and MS patients, we obtain a much clearer description of the gene dysregulation effects. First, many of the previous results can be now appreciated in finer detail, such as (a) the reduction in number of coactivated driver-target pairs in MS, (b) the large number of targets that can be controlled by SOCS1 and SOCS3, and (c) the potential of NFKB1, IFNA1, and IFNB1 to affect the driver-target functional interactions.

Second, we report an interesting mechanism involving the drivers with the highest τ centrality values, that is, SOCS1 and SOCS3. In the HC group, SOCS1 is coactivated with the targets while SOCS3 does not exhibit any significant correlation. In the MS group, we observe the opposite, that is, SOCS1 is silent while SOCS3 becomes coactive. Because both driver genes are dysregulated, the observed "switch" mechanism could therefore be associated with the altered pro-inflammatory response of the MS group. Indeed, these two molecules are known to be strong modulators of macrophage response: SOCS1 inhibits the signaling of pro-inflammatory genes, while SOCS3 is known to be an important actor in inflammatory response, with the ratio of the two proteins determining the actual effect (Wilson, 2014).

DISCUSSION

Identification of Controllable Configurations in Complex Networks

Network controllability refers to the ability to drive an interconnected dynamical system from any initial state to any desired final state in finite time, through a suitable selection of inputs (Rugh & Kailath, 1995; Sontag, 1998). In recent years, an increasing number of research groups from different disciplines have focused their efforts on identifying the minimum set of driver nodes or quantifying the capacity of single nodes to control the entire network, as well as parts of it (Betzel, Gu, Medaglia, Pasqualetti, & Bassett, 2016; Dosenbach et al., 2007; Gu et al., 2017; Y.-Y. Liu & Barabási, 2016; Lugagne et al., 2017; Menara, Bassett, & Pasqualetti, 2019; Muldoon et al., 2016; Sun et al., 2016; Tang et al., 2017; T. Wagner, Valero-Cabre, & Pascual-Leone, 2007; Wuchty, 2014). Despite being theoretically attractive, network controllability still suffers from computational issues that limit its impact in concrete applications. This is mainly due to the presence of multiple equivalent controllable walks in a network that make

the associated controllability problem ill-posed and/or the resulting solution space very big (Rugh & Kailath, 1995; Sontag, 1998).

To reduce such complexity, we propose a method based on control centrality, which was previously designed to quantify the ability of one node to control directed networks (Y.-Y. Liu et al., 2012). First, we define the *target* control centrality to measure the controllability of a specific part of the network, that is, a predefined target set. Because edges are directed, this has the advantage of ignoring the part of the network that is not traversed by the walks connecting the driver to the target set. Second, we introduce an ordering among the target nodes and perform a stepwise controllability test with increasing size. Because of the ranking, only one controllable configuration will be identified, that is, the one with the highest ranking (Figure 1). To test the controllability of the driver-targets configuration at each step, we adopted the Kalman criterion (Kalman, 1963; Rugh & Kailath, 1995). However, the entire iterative framework is quite flexible and other methods, such as Gramian condition (Klickstein, Shirin, & Sorrentino, 2017), Popov-Belevich-Hautus criterion (J. Li, Chen, Pequito, Pappas, & Preciado, 2018), or feedback vertex set (Zañudo, Yang, & Albert, 2017), could be used as alternative controllability criteria.

The main advantage in terms of time of our method is that we identify a controllable driver-targets configuration by performing j tests when $j < S$ out of S target nodes are controllable, as opposed to a brute-force approach that requires us to test all the possible configurations with j targets and that, in the worst-case scenario, leads to $\binom{j}{S}$ tests.

In addition, because of the stepwise procedure, the method does not need to compute the rank of the full controllability matrix and this allows us to minimize possible round-off errors, at least for relatively small networks. Specifically, our results show that the computations are reliable when $N < 180$ and the connection density is < 0.2 (Figure S2). While these numbers are in general relatively low and impede to scale up to very large networks, they still represent ranges that are compatible with typical brain networks obtained from neuroimaging data (E. Bullmore & Sporns, 2009; E. T. Bullmore & Bassett, 2011; De Vico Fallani, Richiardi, Chavez, & Achard, 2014).

Control Pathways in Macrophage Molecular Networks

The study of the molecular interactions is crucial to the understanding of the basic functions of the cell such as proliferation or apoptosis (Maslov & Sneppen, 2002; Taylor et al., 2009). Determining the connection mechanisms that rule a specific biological function can significantly impact our daily life by providing new therapeutics to counteract diseases (Drier, Sheffer, & Domany, 2013; Menche et al., 2015; Tong et al., 2017). Studying molecular networks is however difficult, because in general we do not know the true functional interactions of a cell and indirect techniques such as gene coexpression are typically employed to infer such connections (Uygun, Peng, Lehti-Shiu, Last, & Shiu, 2016). Based on correlation analysis, these methods cannot inform us about the causal nature of the interactions. More importantly, the reliability of the estimated network critically depends on the number of interactions to number of data samples ratio, which is in practice very low (Uygun et al., 2016).

To overcome this limitation, we reconstruct the directed gene interactions associated with the inflammatory state of the human macrophages by adopting a novel ontology-based approach that integrates the available information from multiple datasets and results in the literature (Henry et al., 2017). Previous studies show that the number of driver nodes in biological networks is rather high because of their sparse and heterogeneous nature (Y.-Y. Liu et al., 2011; Ruths & Ruths, 2014). Consistently, we find that a large percentage of genes (51%) can control

at least one secreted molecule in the target set. Our results also confirm that, despite being crucial for global communication, hubs (e.g., RELA) are not always the most important from a network control perspective (Figure 3A). This stems from the theoretical impossibility to diversify the input signals to all the connected neighbors (Y.-Y. Liu et al., 2011). The found driver genes are heterogeneously distributed across the tested gene classes. However, our method highlights SOCS1 and SOCS3 as the drivers with the highest target control centrality values, with other IFNγ-response-related genes showing similar values. This is in line with the known effects of SOCS-genes and IFNγ on molecules secreted by pro-inflammatory macrophages (Koper et al., 2018; J. Liu et al., 2005; McCormick & Heller, 2015; Mosser & Edwards, 2008), supporting the ability of this method to identify biologically relevant drivers.

Overall, these results uncover the existence of potential causal influences from candidate driver genes to the secreted molecules in the human macrophage activation network. Because the identified driver nodes are robust to network alterations, notably when adding new links (Figure 3C, blue diamonds), the obtained results are expected to be sufficiently resilient to the integration of new gene-gene interactions. Notably, our results are also relatively robust to the deletion of links, a situation that is often associated with filtering or thresholding out the weakest or less relevant connections in biological networks (De Vico Fallani, Latora, & Chavez, 2017) (Figure 3C, light blue triangles).

From a different angle, our approach can be seen as a principled manner to focus on specific nodes (the drivers) or node pairs (driver-target) in complex networks. This might have important consequences when studying genome-wide databases where the high number of elements can make prohibitive the assessment of significant gene expressions and/or coexpressions (Bentley et al., 2008; Mullighan et al., 2007).

Dysregulated Genes and Aberrant Interactions in Multiple Sclerosis

Multiple sclerosis is an immune-mediated disease in which the immune system erroneously attacks myelin in the central nervous system. There are many neurological symptoms, including motor and cognitive deficits, that can vary in type and severity depending on the attacked central nervous system regions (Hauser, Oksenberg, & Baranzini, 2015). The role of macrophages in MS is crucial because of their ability to obtain a pro-inflammatory activation state, including the release of pro-inflammatory cytokines and leading to central nervous system tissue damage (Chu et al., 2018). Hence, dysregulation of macrophages may lead to autoimmunity and persistent inflammatory diseases (Strauss, Dunbar, Bartlett, & Phillips, 2015). While the etiology of MS is still not well understood there is a large consensus on its genetic basis and on the importance of unveiling the underlying network mechanisms (Airas, 2015).

In this study, we combined network controllability tools and gene expression data to detect the genes responsible for altering the macrophage action in multiple sclerosis. Different from standard approaches, where the attention is focused on the identification of the driver nodes in a network, we here propose an alternative way of exploiting network controllability. We first show that the macrophage inflammatory state in the MS group was characterized by a drastic alteration of the coactivations in the driver and target genes (Figure 4). Such absence of coordination was in general associated with the presence of dysregulated genes along the walks from the driver to the target node. Notably, the pathological dysregulation of NFKB1, IFNB1, and IFNA1, which belong to the same feedback cycle (Figue S4), critically affects several driver-target functional interactions (Figure 5).

Finally, our approach allows us to identify a shift mechanism for dysregulated SOCS1 and SOCS3 drivers, showing opposite coactivation patterns in MS patients compared with the healthy controls (Figure 6). These results suggest that experimentally stimulating SOCS3—a

strong inducer of pro-inflammatory response—might be more effective for moving the state of the altered secreted molecules towards physiological configurations. Taken together, these results might have practical consequences on how to design intervention strategies and counteract disease phenotype.

Methodological Considerations

Our method uses Kalman controllability rank condition (Kalman, 1963) to quantify the centrality of the driver nodes. This criterion assumes that the investigated system has a linear dynamics (Equation 2). In our case, this means that the changes in the gene activation would follow a linear trend. While this is in general not true and difficult to ascertain, it appears that results from non-linear tests are often dominated by linear relationships (Song, Langfelder, & Horvath, 2012; Steuer, Kurths, Daub, Weise, & Selbig, 2002). Furthermore, a significant fraction of the data analysis and modeling deals exclusively with linear approaches as they are simpler and easy to interpret and serve as a prerequisite of nonlinear behavior (Y.-Y. Liu & Barabási, 2016).

Another peculiarity of our approach is the assumption of time-invariant interactions in the molecular gene network. On the one hand, this assumption allows us to better exploit the well-established results and tools in network controllability (Y.-Y. Liu et al., 2011); on the other hand, it might conflict with existing literature looking for biological connectivity changes between conditions or populations such as differential gene coexpression (Choi, Yu, Yoo, & Kim, 2005). Here, we hypothesized that the activation state of each node (in terms of gene expression) could eventually change, but not the underlying network structure. Thus, our network— obtained from detailed maps of the macrophage cells—would only act as a substrate/proxy for functional interactions, such as correlated gene activities.

Our method requires a specific ordering of the target nodes. While this can be typically achieved in many biological applications—by ranking nodes according to their state (e.g., gene expression: Liseron-Monfils, Olson, & Ware, 2018; Zhao, Yang, Huang, & Holme, 2011; brain activity: Lohmann et al., 2010; Sen, Chu, & Parhi, 2019)—challenges might still remain in general. When it is not possible to impose a ranking of the nodes from external knowledge, another possibility is to derive it from the network structure taking into account, for example, node centrality values. However, there might be multiple nodes with the same centrality value, which would impede a proper ranking. In those situations, multiple node centrality measures could be integrated to get a more heterogeneous distribution (e.g., degree, strength, betweenness, closeness; Newman, 2010). Another possibility is to add equally important targets at the same iteration step and test whether they are simultaneously controllable. While this procedure is suboptimal and may underestimate the number of controllable targets, it still minimizes the computational complexity related to testing multiple driver-target combinations.

We finally notice that our method is conceived for directed networks only, where the dimensionality reduction has a real computational benefit. In fact, in the case of undirected graphs, it is not possible to remove nodes on the walks from the driver to the targets since information is bound to span the entire network. Similarly, for directed but dense networks, the possibility to focus on specific parts of the network, and reduce the computational cost, becomes lower regardless of the topology.

To conclude, it is important to mention that extensions of network controllability tools to time-varying frameworks do exist (A. Li, Cornelius, Liu, Wang, & Barabási, 2016, Zhang, Garas, & Scholtes, 2017). However, in that case networks would be inferred from gene coexpression

and therefore affected by statistical uncertainty due to sample sizes. Further research is needed to seek how to apply network controllability in presence of noisy time-varying connections.

Conclusion

In this study, we introduce a method to quantify the ability of candidate driver nodes to drive the state of a target set within a sparse and directed network. Further, we illustrate how this method works for the molecular network associated with the human macrophage inflammatory response. The obtained results reveal in a principled way the genes that are significantly dysregulated in multiple sclerosis. We hope that this method can contribute to the identification of the key nodes in biological networks to better identify pharmacological targets to counteract human diseases.

MATERIAL AND METHODS

Stepwise Target Controllability

We introduce a method to identify which target nodes in a network can be controlled from a single *driver* node. To do so, we start by considering the canonical linear time-invariant dynamics on a directed network described by the adjacency matrix $A \in \mathbb{R}^{N \times N}$:

$$\dot{\mathbf{x}}(n) = A\mathbf{x}(n) - B\mathbf{u}(n), \qquad \mathbf{y}(n) = C\mathbf{x}(n), \tag{2}$$

where $\mathbf{x}(n) \in \mathbb{R}^N$ describes the state of each node at time t, $B \in \mathbb{R}^N$ specifies the driver node, $\mathbf{u}(n) \in \mathbb{R}^N$ is its external input (or control) signal, $\mathbf{y}(n) \in \mathbb{R}^S$ is the output vector, and $C \in \mathbb{R}^{S \times N}$ is the output matrix identifying the target nodes.

Such a system is controllable if it can be guided from any initial state to any desired final state in finite time, with a suitable choice of input. A necessary and sufficient condition to assess the controllability of Equation 2 is that the controllability matrix Q,

$$Q = \begin{bmatrix} B & AB & A^2B & \cdots & A^{N-1}B \end{bmatrix}, \tag{3}$$

has full row rank, that is, $\text{rank}(Q) = N$. That is the Kalman rank condition, which basically verifies the existence of linearly independent rows in Q (Kalman, 1963; Rugh & Kailath, 1995). If so, the driver node can reach and control the dynamics of all the other nodes through independent walks of length $N - 1$ at maximum.

If it is of interest to control only a target set $\mathcal{T}$ of the network, specified in C and consisting of $S \leq N$ nodes, then Equation 2 can be reduced into a target controllability matrix $Q_{\mathcal{T}} = CQ$ (Equation 1), where C filters the rows of interest corresponding to the targets. Now, the rank of $Q_{\mathcal{T}}$ gives the number $\tau \leq S$ of nodes in the target set that can be controlled by the driver.

To identify a driver-target configuration, we further introduce a hierarchy among the target nodes, so that we can order and relabel them from the most important one to the least, that is, $t_1 \succ t_2 \succ \ldots \succ t_S$. Then we perform the following stepwise procedure for each candidate driver node:

- Step 1. *Initialization*
  - Create a temporary empty target set $\mathcal{T}' \leftarrow \{\}$
  - Set the number of controllable targets $\tau \leftarrow 0$
- Step 2. *Repeat until termination criteria are met.* For $j \leftarrow 1, \ldots, S$ do
  - Add the j-th target node to the target set $\mathcal{T}' \leftarrow \mathcal{T}' \cup \{t_j\}$

- Build the subgraph containing the nodes on walks from the driver to the targets in $\mathcal{T}'$

- Compute the rank of the target controllability matrix $Q_{\mathcal{T}'}$

- *If* rank($Q'_{\mathcal{T}}$) is full then $\tau \leftarrow \tau + 1$ else $\mathcal{T}' \leftarrow \mathcal{T}' \setminus \{t_j\}$

- $j \leftarrow j + 1$

- Step 3. *Output τ and $\mathcal{T}'$*

Eventually, the *target control centrality* τ is the number of controllable targets in $\mathcal{T}$, and the set $\mathcal{T}'$ contains the τ controllable targets with highest ranking.

Note that in general, the method may underestimate the actual number of controllable target nodes because of the occurrence of numerical errors, but not because of the stepwise procedure itself. The Matlab code associated with the stepwise target controllability is freely available at https://github.com/BCI-NET/Public.

Construction of the Macrophage Activation Network

We reconstruct the inflammatory molecular network of the human macrophage by integrating information from the macrophage signal transduction map (Robert, Lu, Law, Freeman, & Hume, 2011; Raza et al., 2008). This map contains a comprehensive, validated, and annotated map of signal transduction pathways of inflammatory processes in macrophages based on the current literature. To extract molecular interactions from this map, we used the Hermit software (Motik, Cuenca Grau, & Sattler, 2008), which implements automatic reasoning based on logical rules. We specifically used the rules implemented in the molecular network ontology to infer molecular interactions depending on the process they belong (Henry et al., 2017; Musen, 2015). Because we are interested in the inflammation process, we restricted our analysis to a specific subset of 101 genes with known roles in macrophage pro-inflammatory activation, and for which their regulation in response to pro-inflammatory stimuli could be confirmed in our data set. These genes were classified according to their function in the cell: *sensing*, *signaling*, *transcription*, and *secreted* (Table S1), as described in databases such as NCBI Gene (Agarwala et al., 2018; UniProt Consortium, 2019), and GeneCards Stelzer et al., 2016. The full network was thus reduced to include only these genes and their interactions. Because of recent studies, we also opted to exclude two edges (from SOCS3 to IFNGR1 and to IFNGR2) to represent the involved pathways (Wilson, 2014).

The resulting network contains $N = 101$ nodes and $L = 211$ unweighted directed edges representing either activation or inhibition between genes. The total degree k of each node in the network is computed by summing the number of incoming and outgoing edges:

$$k_i = \sum_{j=1}^{N} A_{ij} + \sum_{j=1}^{N} A_{ji}, \tag{4}$$

where $A_{ij} = 1$ if there is an edge between the corresponding genes, and 0 otherwise.

Collection of Macrophage mRNA Expression Data

Collection of blood for the study was approved by the French Ethics Committee and the French Ministry of Research (DC-2012-1535 and AC-2012-1536). Written informed consent was obtained from all study participants. All patients fulfilled diagnostic criteria for multiple sclerosis (Thompson et al., 2018), and individuals (multiple sclerosis patients and healthy donors) with any other inflammatory or neurological disorders were excluded from the study. Patients were included in the study only if they were not undergoing treatment.

Blood was sampled from eight MS patients and eight healthy controls in acid citrate dextrose tubes. From blood samples, peripheral blood mononuclear cells were isolated using Ficoll Paque Plus (www.gelifesciences.com) and centrifugation (2,200 rpm, 20 min). Cells were washed in PBS and RPMI +10% FCS. Monocytes were isolated with anti-CD14 microbeads (www.miltenyibiotec.com) and plated in 12-well plates (500,000 cells/well) in RPMI +10% FCS and granulocyte-macrophage colony-stimulating factor (500 U/ml) to induce differentiation into macrophages. After 72 hr, media was replaced with fresh media supplemented with granulocyte-macrophage colony-stimulating factor (500 U/ml) to maintain "alert" macrophages or IFNγ (200 U/ml) + upLPS (10 ng/ml) to induce "pro-inflammatory" activation. Cells were lysed after 24 hr and RNA was extracted with RNeasy Mini Kit (www.qiagen.com).

Transcriptome sequencing cDNA libraries were prepared using a stranded mRNA polyA selection (Truseq stranded mRNA kit, www.illumina.com). For each sample, we performed 60 million single-end, 75 base reads on a NextSeq 500 sequencer (www.illumina.com). RNA-Seq data analyses were performed by GenoSplice technology (www.genosplice.com). Sequencing, data quality, reads repartition (e.g., for potential ribosomal contamination), and insert size estimation were performed using FastQC (Andrews, 2010), Picard-Tools (http://broadinstitute .github.io/picard/), Samtools (H. Li et al., 2009) and rseqc (Wang, Wang, & Li, 2012). Reads were mapped using STARv2.4.0 (Dobin et al., 2013) on the hg19 Human genome assembly. Gene expression regulation study was performed (Noli, Capalbo, Ogilvie, Khalaf, & Ilic, 2015). Briefly, for each gene present in the FAST DB v2018_1 annotations, reads aligning on constitutive regions (that are not prone to alternative splicing) were counted. Based on these read counts, normalization was performed using DESeq2 (Love, Huber, & Anders, 2014) in R (v.3.2.5; R Core Team, 2014).

Network Modeling and Data Analysis

In the modeling framework described by Equation 2, matrix A corresponds to the molecular network and represents the time-invariant component of the system. The dynamic component is instead represented by the gene activation response in the healthy and diseased condition (Figure 2B), computed as the ratio in gene expression between the "pro-inflammatory" and "alert" condition. Specifically, $\mathbf{x}(n)$ represents the gene activation. B is a vector identifying the candidate driver. The control signal $\mathbf{u}(n)$ is outside the scope of this work. The output vector $\mathbf{y}(n)$ and the output matrix C identify the target nodes.

We select the genes belonging to the *secreted* molecules class (Table S1) as our target set $\mathcal{T}$. All the nodes in the other classes are then tested separately as potential driver nodes by computing their target control centrality τ. To enhance numerical precision, the logarithmic transformation $\log(q+1)$ is applied to the elements of the target controllability matrix $Q_{\mathcal{T}}$ (Equation 1).

The hierarchy among the target nodes is established by computing the fold change Δ between the corresponding gene activation in the two groups:

$$\Delta = \frac{\mu_{MS}}{\mu_{HC}}, \qquad (5)$$

where μ_{MS} and μ_{HC} are group-averages for MS patients and healthy controls, respectively, of the gene activation. Nodes with higher Δ absolute values are ranked first. Highly positive Δ values indicate a too strong inflammatory response (overactivation) in the MS patients with respect to the healthy controls. Highly negative Δ values indicate a too weak inflammatory response (underactivation). We define *dysregulated* genes along the controllable driver-target walks as those for which $|\Delta|$ is above the 75th percentile.

We perform a robustness analysis to evaluate the stability of the identified driver nodes to potential errors in the molecular network reconstruction. We simulate attacks with increasing intensity, that is, up to 20% of the nodes or edges in the network. When removing nodes, we consider the following cases: (a) random deletion, (b) preferential removal of high-degree nodes, and (c) preferential removal of low-degree nodes. Preferential attacks are performed by selecting nodes with a probability p proportional to their degree k, that is, $p \propto k$ for high-degree nodes and $p \propto -k$ for low-degree nodes. When perturbing edges, we test: (a) random addition, (b) random deletion, and (c) random rewiring. For each case, we simulated $1,000$ repetitions and we computed the target control centrality τ for the driver nodes identified in the original network. Then, we report the percentage of nodes that cease to be drivers (i.e., $\tau = 0$), that is, the percentage of nodes that are drivers in our analysis, but are no longer able to control any target in the perturbed case.

ACKNOWLEDGMENTS

SUPPORTING INFORMATION

Supporting information for this article is available at https://doi.org/10.1162/netn_a_00180.

AUTHOR CONTRIBUTIONS

Giulia Bassignana: Conceptualization; Data curation; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing. Jennifer Fransson: Data curation; Methodology; Writing – original draft; Writing – review & editing. Vincent Henry: Data curation; Formal analysis; Methodology; Software; Writing – original draft; Writing – review & editing. Olivier Colliot: Conceptualization; Funding acquisition; Project administration; Supervision; Writing – original draft; Writing – review & editing. Violetta Zujovic: Conceptualization; Data curation; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing. Fabrizio De Vico Fallani: Conceptualization; Investigation; Methodology; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

FUNDING INFORMATION

REFERENCES

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., . . . Zbicz, K. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. **DOI:** https://doi.org/10.1093/nar/gkx1095, **PMID:** 29140470, **PMCID:** PMC5753372

Airas, L. (2015). Hormonal and gender-related immune changes in multiple sclerosis. *Acta Neurologica Scandinavica*, *132*(S199), 62–70. **DOI:** https://doi.org/10.1111/ane.12433, **PMID:** 26046561

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data.*

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. **DOI:** https://doi.org/10.1038/nrg1272, **PMID:** 14735121

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., . . . Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53–59. **DOI:** https://doi.org/10.1038/nature07517, **PMID:** 18987734, **PMCID:** PMC2581791

Betzel, R. F., Gu, S., Medaglia, J. D., Pasqualetti, F., & Bassett, D. S. (2016). Optimally controlling the human connectome: The role of network topology. *Scientific Reports*, *6*. **DOI:** https://doi.org/10.1038/srep30770, **PMID:** 27468904, **PMCID:** PMC4965758

Biggs, N. (1974). *Algebraic graph theory*. Cambridge University Press. **DOI:** https://10.1017/CBO9780511608704

Bitsch, A., Kuhlmann, T., Costa, C. D., Bunkowski, S., Polak, T., & Brück, W. (2000). Tumour necrosis factor alpha mRNA expression in early multiple sclerosis lesions: Correlation with demyelinating activity and oligodendrocyte pathology. *Glia*, *29*(4), 366–375. **DOI:** https://doi.org/10.1002/(SICI)1098-1136(20000215)29:4<366::AID-GLIA7>3.0.CO;2-Y, **PMID:** 10652446

Bonifazi, P., Goldin, M., Picardo, M. A., Jorquera, I., Cattani, A., Bianconi, G., . . . Cossart, R. (2009). GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. *Science*, *326*(5958), 1419–1424. **DOI:** https://doi.org/10.1126/science.1175509, **PMID:** 19965761

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. **DOI:** https://doi.org/10.1038/nrn2575, **PMID:** 19190637

Bullmore, E. T., & Bassett, D. S. (2011). Brain graphs: Graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, *7*(1), 113–140. **DOI:** https://doi.org/10.1146/annurev-clinpsy-040510-143934, **PMID:** 21128784

Choi, J. K., Yu, U., Yoo, O. J., & Kim, S. (2005). Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics*, *21*(24), 4348–4355. **DOI:** https://doi.org/10.1093/bioinformatics/bti722, **PMID:** 16234317

Chu, F., Shi, M., Zheng, C., Shen, D., Zhu, J., Zheng, X., & Cui, L. (2018). The roles of macrophages and microglia in multiple sclerosis and experimental autoimmune encephalomyelitis. *Journal of Neuroimmunology*, *318*, 1–7. **DOI:** https://doi.org/10.1016/j.jneuroim.2018.02.015, **PMID:** 29606295

Commault, C., Van der Woude, J., & Frasca, P. (2019). Functional target controllability of networks: Structural properties and efficient algorithms. *IEEE Transactions on Network Science and Engineering*. **DOI:** https://doi.org/10.1109/TNSE.2019.2937404

De Vico Fallani, F. , Latora, V., & Chavez, M. (2017). A Topological criterion for filtering information in complex brain networks. *PLoS Computational Biology*, *13*(1), e1005305. **DOI:** https://doi.org/10.1371/journal.pcbi.1005305, **PMID:** 28076353, **PMCID:** PMC5268647

De Vico Fallani, F., Richiardi, J., Chavez, M., & Achard, S. (2014). Graph analysis of functional brain networks: Practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1653), 20130521.

DOI: https://doi.org/10.1098/rstb.2013.0521, **PMID:** 25180301, **PMCID:** PMC4150298

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. **DOI:** https://doi.org/10.1093/bioinformatics/bts635, **PMID:** 23104886, **PMCID:** PMC3530905

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., . . . Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078. **DOI:** https://doi.org/10.1073/pnas.0704320104, **PMID:** 17576922, **PMCID:** PMC1904171

Drier, Y., Sheffer, M., & Domany, E. (2013). Pathway-based personalized analysis of cancer. *Proceedings of the National Academy of Sciences*, *110*(16), 6388–6393. **DOI:** https://doi.org/10.1073/pnas.1219651110, **PMID:** 23547110, **PMCID:** PMC3631698

Gao, J., Liu, Y.-Y., D'Souza, R. M., & Barabási, A.-L. (2014). Target control of complex networks. *Nature Communications*, *5*, 5415. **DOI:** https://doi.org/10.1038/ncomms6415, **PMID:** 25388503, **PMCID:** PMC4243219

Gu, S., Betzel, R. F., Mattar, M. G., Cieslak, M., Delio, P. R., Grafton, S. T., . . . Bassett, D. S. (2017). Optimal trajectories of brain state transitions. *NeuroImage*, *148*(Suppl. C), 305–317. **DOI:** https://doi.org/10.1016/j.neuroimage.2017.01.003, **PMID:** 28088484, **PMCID:** PMC5489344

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., . . . Bassett, D. S. (2015). Controllability of structural brain networks. *Nature Communications*, *6*, 8414. **DOI:** https://doi.org/10.1038/ncomms9414, **PMID:** 26423222, **PMCID:** PMC4600713

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature*, *406*(6792), 147–150. **DOI:** https://doi.org/10.1038/35018000, **PMID:** 10910346

Hauser, S. L., Oksenberg, J. R., & Baranzini, S. E. (2015). Multiple Sclerosis. In *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease* (pp. 1001–1014). Elsevier. **DOI:** https://doi.org/10.1016/B978-0-12-410529-4.00085-1

Hautus, M. (1970). Stabilization controllability and observability of linear autonomous systems. *Indagationes Mathematicae (Proceedings)*, *73*, 448–455. **DOI:** https://doi.org/10.1016/S1385-7258(70)80049-X

Henry, V. J., Saïs, F., Marchadier, E., Dibie, J., Goelzer, A., & Fromion, V. (2017). BiPOm: Biological interlocked Process Ontology for metabolism. How to infer molecule knowledge from biological process? In *International Conference on Biomedical Ontology, ICBO 2017*. Newcastle upon Tyne, United Kingdom.

Heuberger, C., & Wagner, S. (2011). The number of maximum matchings in a tree. *Discrete Mathematics*, *311*(21), 2512–2542. **DOI:** https://doi.org/10.1016/j.disc.2011.07.028, **PMID:** 22158732, **PMCID:** PMC3226351

Hopcroft, J. E., & Karp, R. M. (1971). A n5/2 algorithm for maximum matchings in bipartite. In *12th Annual Symposium on Switching and Automata Theory (SWAT 1971)* (pp. 122–125). **DOI:** https://doi.org/10.1109/SWAT.1971.1

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833),

41–42. **DOI:** https://doi.org/10.1038/35075138, **PMID:** 11333967

Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics: Series A, Control*, 1(2), 152–192. **DOI:** https://doi.org/10.1137/0301010

Klickstein, I., Shirin, A., & Sorrentino, F. (2017). Energy scaling of targeted optimal control of complex networks. *Nature Communications*, 8(1), 15145. **DOI:** https://doi.org/10.1038/ncomms15145, **PMID:** 28436417, **PMCID:** PMC5413984

Koper, O. M., Kamińska, J., Sawicki, K., & Kemona, H. (2018). CXCL9, CXCL10, CXCL11, and their receptor (CXCR3) in neuroinflammation and neurodegeneration. *Advances in Clinical and Experimental Medicine: Official Organ Wroclaw Medical University*, 27(6), 849–856. **DOI:** https://doi.org/10.17219/acem/68846, **PMID:** 29893515

Li, A., Cornelius, S. P., Liu, Y.-Y., Wang, L., & Barabási, A.-L. (2016). The fundamental advantages of temporal networks. *arXiv:1607.06168 [nlin]*

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. **DOI:** https://doi.org/10.1093/bioinformatics/btp352, **PMID:** 19505943, **PMCID:** PMC2723002

Li, J., Chen, X., Pequito, S., Pappas, G. J., & Preciado, V. M. (2018). Structural target controllability of undirected networks. In *2018 IEEE Conference on Decision and Control (CDC)* (pp. 6656–6661). Miami Beach, FL: IEEE. **DOI:** https://doi.org/10.1109/CDC.2018.8619399

Lin, C.-T. (1974). Structural controllability. *IEEE Transactions on Automatic Control*, 19(3), 201–208. **DOI:** https://doi.org/10.1109/TAC.1974.1100557

Liseron-Monfils, C., Olson, A., & Ware, D. (2018). NECorr, a Tool to rank gene importance in biological processes using molecular networks and transcriptome data. *bioRxiv: 326868*. **DOI:** https://doi.org/10.1101/326868, **PMID:** 30356219

Liu, J., Guan, X., & Ma, X. (2005). Interferon regulatory factor 1 is an essential and direct transcriptional activator for interferon {gamma}-induced RANTES/CCl5 expression in macrophages. *The Journal of Biological Chemistry*, 280(26), 24347–24355. **DOI:** https://doi.org/10.1074/jbc.M500973200, **PMID:** 15860458

Liu, Y., & Liu, G. (2004). Number of maximum matchings of bipartite graphs with positive surplus. *Discrete Mathematics*, 274(1), 311–318. **DOI:** https://doi.org/10.1016/S0012-365X(03)00204-8

Liu, Y.-Y., & Barabási, A.-L. (2016). Control principles of complex networks. *Reviews of Modern Physics*, 88(3). **DOI:** https://doi.org/10.1103/RevModPhys.88.035006

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature*, 473(7346), 167–173. **DOI:** https://doi.org/10.1038/nature10011, **PMID:** 21562557

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2012). Control centrality and hierarchical structure in complex networks. *PLoS ONE*, 7(9), e44459. **DOI:** https://doi.org/10.1371/journal.pone.0044459, **PMID:** 23028542, **PMCID:** PMC3459977

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Gene replacement and transgenic animals. In *Molecular cell biology*. (4th ed.). W. H. Freeman.

Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., . . . Turner, R. (2010). Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PLoS ONE*, 5(4), e10232. **DOI:** https://doi.org/10.1371/journal.pone.0010232, **PMID:** 20436911, **PMCID:** PMC2860504

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). **DOI:** https://doi.org/10.1101/002832

Lugagne, J.-B., Sosa Carrillo, S., Kirch, M., Köhler, A., Batt, G., & Hersen, P. (2017). Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature Communications*, 8. **DOI:** https://doi.org/10.1038/s41467-017-01498-0, **PMID:** 29150615, **PMCID:** PMC5693866

Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, 296(5569), 910–913. **DOI:** https://doi.org/10.1126/science.1065103, **PMID:** 11988575

McCormick, S. M., & Heller, N. M. (2015). Regulation of macrophage, dendritic cell, and microglial phenotype and function by the SOCS proteins. *Frontiers in Immunology*, 6. **DOI:** https://doi.org/10.3389/fimmu.2015.00549

Menara, T., Bassett, D. S., & Pasqualetti, F. (2019). Structural controllability of symmetric networks. *IEEE Transactions on Automatic Control*, 64(9), 3740–3747. **DOI:** https://doi.org/10.1109/TAC.2018.2881112.2881112

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Baraba'si, A.-L. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224), 1257601. **DOI:** https://doi.org/10.1126/science.1257601, **PMID:** 25700523, **PMCID:** PMC4435741

Mosser, D. M., & Edwards, J. P. (2008). Exploring the full spectrum of macrophage activation. *Nature Reviews Immunology*, 8(12), 958–969. **DOI:** https://doi.org/10.1038/nri2448, **PMID:** 19029990, **PMCID:** PMC2724991

Motik, B., Cuenca Grau, B., & Sattler, U. (2008). Structured objects in owl: Representation and reasoning. In *Proceeding of the 17th international conference on World Wide Web - WWW '08* (pp. 555–547). Beijing, China: ACM Press. **DOI:** https://doi.org/10.1145/1367497.1367573

Muldoon, S. F., Pasqualetti, F., Gu, S., Cieslak, M., Grafton, S. T., Vettel, J. M., & Bassett, D. S. (2016). Stimulation-based control of dynamic brain networks. *PLoS Computational Biology*, 12(9), e1005076. **DOI:** https://doi.org/10.1371/journal.pcbi.1005076, **PMID:** 27611328, **PMCID:** PMC5017638

Mullighan, C. G., Goorha, S., Radtke, I., Miller, C. B., Coustan-Smith, E., Dalton, J. D., . . . Downing, J. R. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*, 446(7137), 758–764. **DOI:** https://doi.org/10.1038/nature05690, **PMID:** 17344859

Murota, K., & Poljak, S. (1990). Note on a graph-theoretic criterion for structural output controllability. *IEEE Transactions on Automatic Control*, 35(8), 939–942. **DOI:** https://doi.org/10.1109/9.58507

Musen, M. A. (2015). The protégé project: A look back and a look forward. *AI matters*, *1*(4), 4–12. **DOI:** https://doi.org/10.1145/2757001.2757003, **PMID:** 27239556, **PMCID:** PMC4883684

Newman, M. (2010). *Networks: An introduction*. Oxford, New York: Oxford University Press.

Noli, L., Capalbo, A., Ogilvie, C., Khalaf, Y., & Ilic, D. (2015). Discordant growth of monozygotic twins starts at the blastocyst stage: A case study. *Stem Cell Reports*, *5*(6), 946–953. **DOI:** https://10.1016/j.stemcr.2015.10.006

Pu, C.-L., Pei, W.-J., & Michaelson, A. (2012). Robustness analysis of network controllability. *Physica A: Statistical Mechanics and its Applications*, *391*(18), 4420–4425. **DOI:** https://doi.org/10.1016/j.physa.2012.04.019

Raza, S., Robertson, K. A., Lacaze, P. A., Page, D., Enright, A. J., Ghazal, P., & Freeman, T. C. (2008). A logic-based diagram of signalling pathways central to macrophage activation. *BMC Systems Biology*, *2*, 36. **DOI:** https://doi.org/10.1186/1752-0509-2-36, **PMID:** 18433497, **PMCID:** PMC2383880

R Core Team. (2014). *R: A language and environment for statistical computing*.

Robert, C., Lu, X., Law, A., Freeman, T. C., & Hume, D. A. (2011). Macrophages.com: An on-line community resource for innate immunity research. *Immunobiology*, *216*(11), 1203–1211. **DOI:** https://doi.org/10.1016/j.imbio.2011.07.025, **PMID:** 21924793

Rugh, W. J., & Kailath, T. (1995). *Linear system theory* (2nd ed.). Pearson.

Ruths, J., & Ruths, D. (2014). Control Profiles of Complex Networks. *Science*, *343*(6177), 1373–1376. **DOI:** https://doi.org/10.1126/science.1242063, **PMID:** 24653036

Sen, B., Chu, S.-H., & Parhi, K. K. (2019). Ranking regions, edges and classifying tasks in functional brain graphs by sub-graph entropy. *Scientific Reports*, *9*(1), 1–20. **DOI:** https://doi.org/10.1038/s41598-019-44103-8, **PMID:** 31110317, **PMCID:** PMC6527859

Shields, R., & Pearson, J. (1976). Structural controllability of multiinput linear systems. *IEEE Transactions on Automatic Control*, *21*(2), 203–212. **DOI:** https://doi.org/10.1109/TAC.1976.1101198

Song, L., Langfelder, P., & Horvath, S. (2012). Comparison of co-expression measures: Mutual information, correlation, and model based indices. *BMC Bioinformatics*, *13*(1), 328. **DOI:** https://doi.org/10.1186/1471-2105-13-328, **PMID:** 23217028, **PMCID:** PMC3586947

Sontag, E. D. (1998). *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (2nd ed.). New York: Springer-Verlag. **DOI:** https://doi.org/10.1007/978-1-4612-0577-7

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., . . . Lancet, D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*, *54*(1), 1.30.1–1.30.33. **DOI:** https://doi.org/10.1002/cpbi.5, **PMID:** 27322403

Steuer, R., Kurths, J., Daub, C. O., Weise, J., & Selbig, J. (2002). The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, *18*(Suppl. 2), S231–S240. **DOI:** https://doi.org/10.1093/bioinformatics/18.suppl_2.S231, **PMID:** 12386007

Strauss, O., Dunbar, P. R., Bartlett, A., & Phillips, A. (2015). The immunophenotype of antigen presenting cells of the mononu-clear phagocyte system in normal human liver—A systematic review. *Journal of Hepatology*, *62*(2), 458–468. **DOI:** https://doi.org/10.1016/j.jhep.2014.10.006, **PMID:** 25315649

Sun, X., Hu, F., Wu, S., Qiu, X., Linel, P., & Wu, H. (2016). Controllability and stability analysis of large transcriptomic dynamic systems for host response to influenza infection in human. *Infectious Disease Modelling*, *1*(1), 52–70. **DOI:** https://doi.org/10.1016/j.idm.2016.07.002, **PMID:** 29928721, **PMCID:** PMC5963324

Tang, E., Giusti, C., Baum, G. L., Gu, S., Pollock, E., Kahn, A. E., . . . Bassett, D. S. (2017). Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nature Communications*, *8*(1), 1252. **DOI:** https://doi.org/10.1038/s41467-017-01254-4, **PMID:** 29093441, **PMCID:** PMC5665937

Taylor, I. W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., . . . Wrana, J. L. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology*, *27*(2), 199–204. **DOI:** https://doi.org/10.1038/nbt.1522, **PMID:** 19182785

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., . . . Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet. Neurology*, *17*(2), 162–173. **DOI:** https://doi.org/10.1016/S1474-4422(17)30470-2

Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., & Rueckert, D. (2017). A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, *64*(1), 155–165. **DOI:** https://doi.org/10.1109/TBME.2016.2549363, **PMID:** 27046891

UniProt Consortium (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. **DOI:** https://doi.org/10.1093/nar/gky1049, **PMID:** 30395287, **PMCID:** PMC6323992

Uno, T. (1997). Algorithms for enumerating all perfect, maximum and maximal matchings in bipartite graphs. In H. W. Leong, H. Imai, & S. Jain (Eds.), *Algorithms and Computation* (pp. 92–101). Springer. **DOI:** https://doi.org/10.1007/3-540-63890-3_11

Uygun, S., Peng, C., Lehti-Shiu, M. D., Last, R. L., & Shiu, S.-H. (2016). Utility and limitations of using gene expression data to identify functional associations. *PLoS Computational Biology*, *12*(12). **DOI:** https://doi.org/10.1371/journal.pcbi.1005244

Wagner, S. G. (2007). On the number of matchings of a tree. *European Journal of Combinatorics*, *28*(4), 1322–1330. **DOI:** https://doi.org/10.1016/j.ejc.2006.01.014

Wagner, T., Valero-Cabre, A., & Pascual-Leone, A. (2007). Noninvasive Human Brain Stimulation. *Annual Review of Biomedical Engineering*, *9*, 527–65. **DOI:** https://doi.org/10.1146/annurev.bioeng.9.061206.133100, **PMID:** 17444810

Wang, L., Wang, S., & Li, W. (2012). RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*, *28*(16), 2184–2185. **DOI:** https://doi.org/10.1093/bioinformatics/bts356, **PMID:** 22743226

Wilson, H. M. (2014). SOCS proteins in macrophage polarization and function. *Frontiers in Immunology*, *5*. **DOI:** https://doi.org/10.3389/fimmu.2014.00357

Wuchty, S. (2014). Controllability in protein interaction networks. *Proceedings of the National Academy of Sciences*, *111*(19), 7156–7160. **DOI:** https://doi.org/10.1073/pnas.1311231111, **PMID:** 24778220, **PMCID:** PMC4024882

Zañudo, J. G. T., Yang, G., & Albert, R. (2017). Structure-based control of complex networks with nonlinear dynamics. *Proceedings of the National Academy of Sciences*. **DOI:** https://doi.org/10.1073/pnas.1617387114, **PMID:** 28655847, **PMCID:** PMC5514702

Zdeborová, L., & Mézard, M. (2006). The number of matchings in random graphs. *Journal of Statistical Mechanics: Theory and Experiment*, *2006*(05), P05003. **DOI:** https://doi.org/10.1088/1742-5468/2006/05/P05003

Zhang, Y., Garas, A., & Scholtes, I. (2017). Controllability of temporal networks: An analysis using higher-order networks. *arXiv:1701.06331 [physics]*

Zhao, J., Yang, T.-H., Huang, Y., & Holme, P. (2011). Ranking candidate disease genes from gene expression and protein interaction: A Katz-centrality based approach. *PLoS ONE*, *6*(9), e24306. **DOI:** https://doi.org/10.1371/journal.pone.0024306, **PMID:** 21912686, **PMCID:** PMC3166320