Novel regulators of endosome dynamics, MHCII antigen presentation and chemosensitivity

Wijdeven, R.H.M.

Chapter 6:
Exploring genome-wide datasets of MHC class II antigen presentation

Ruud H Wijdeven[1,2], Jeroen M Bakker[1,2], Petra Paul[1], Jacques Neefjes[1]

[1] Division of Cell Biology, the Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, the Netherlands
[2] These authors contributed equally to this work

Molecular Immunology 55 (2013), pp. 172-174

Abstract
MHC class II molecules (MHCII) are critical for presenting antigens to CD4+ T-cells. They control ignition of CD4+ T cells and are as such involved in most auto-immune diseases. To define proteins and pathways controlling MHCII antigen presentation and expression, we performed a genome-wide flow cytometry based RNAi screen. Hits were subsequently classified by two screens that monitored the intracellular distribution and transcription of MHCII. This multi-dimensional approach allowed sub-classification of hits into functional groups as a first step to defining new pathways controlling MHCII antigen presentation. The datasets from this screen are used as a template for several follow-up studies. This review focuses on how data from genome-wide screens can be used for target-lead identification, data mining, systems biology and systematic cell biology.

Introduction
Antigen presenting cells (APCs) activate CD4+ helper T-cells by presenting antigens on their Major Histocompatibility Complex class II molecules (MHCII). Many auto-immune diseases are caused by improper antigen presentation or undesired stimulation of autoreactive CD4+ T-cells and MHCII is the strongest genetic link to most auto-immune diseases (1). Expression of MHCII is limited to professional APCs like dendritic cells, macrophages and B-cells. Other cell types only express MHCII after cytokine stimulation, hereby contributing to local immune responses.
Transcription of MHCII is controlled by a transcriptional 'master regulator' called CIITA (2). After translation and synthesis in the ER, MHCII associates with the invariant chain (Ii) for transport to late endosomes, called MHCII compartments (MIIC). Here, the chaperone protein HLA-DM enhances the actual peptide loading. Resident proteases degrade the MHCII-associated invariant chain, leaving only a small Ii fragment termed CLIP in the peptide binding groove of MHC II (3). Endocytosed antigens are also digested and trimmed by proteases in the MIIC and the resulting fragments compete with CLIP to fill the peptide binding groove. Ultimately MHCII is transported to the plasma membrane for antigen presentation. This transport involves motor proteins and is under control of pathways that are poorly understood (4).
Many direct players in MHCII antigen presentation are defined but how these fit into pathways is unclear. For example it is known that CIITA is expressed in a tissue specific manner, but it remains unclear how the transcription factors regulating this expression are activated. This addresses a general biological question: how tissue selective expression is imposed. In principle, genome-wide data sets should contain the information on these (and other) issues, but it is unclear how to extract them. We discuss findings of how MHC class II expression and transport are regulated and ways to systematically use large datasets for solving such biological questions.

Current status
To identify genes controlling MHCII expression, a genome wide siRNA screen was performed. After knockdown of every annotated gene, a two color high throughput flow cytometry analysis was performed to examine effects on MHCII expression. Two monoclonal antibodies were used to visualize correctly loaded MHCII on the

plasma membrane (L243) and MHCII bound to the CLIP fragment (CerCLIP), which represents inefficient loading of MHCII with correct antigenic peptides. After multiple rounds of screening, 276 genes were identified that affected MHCII expression (L243) and/or peptide loading (CerCLIP). The set of genes included most already known factors in the MHCII antigen presenting pathway, such as CIITA, DM and the HLA-DRA and -DRB chains. By using two antibodies probing different conformations of one protein, genes could be divided as being involved in controlling expression or peptide loading of MHCII molecules using flow cytometry.
To extract genes controlling MHCII expression at the transcriptional level, all hits were analyzed for their effect on MHCII transcription. mRNA levels of MHCII, the invariant chain and CIITA were determined by qPCR to select the genes involved in controlling MHCII and/or CIITA transcription. Nine candidates were identified including CIITA. From these data,we were able to define a transcriptional network that controls MHCII expression.
Still, the majority of the hits were not functionally assigned and possibly regulated intracellular trafficking of MHCII. To study this intracellular localization of GFP-tagged MHCII was visualized using confocal microscopy after knockdown of the hits. Using Cellprofiler software the phenotypes were clustered into different groups. These effects included vesicle enlargement, clustering or dispersion of vesicles and a redistribution of MHCII to the cell surface. This information was combined with the flow cytometry and qPCR data and used for a data-based integrative bioinformatic approach to predict functional relationships of the candidates and build pathways and networks of genes controlling MHCII functioning.
Next we tried to use these data to extract new biology. Several hits from the screen gave a redistribution of MHCII to the cell surface after knockdown, resembling a mature dendritic cell phenotype where the intracellular pool of MHCII is exported to the cell surface. The biology of this massive but regulated export is poorly understood. By analyzing microarray expression data of immature and mature dendritic cells it was found that 11 of the genes that gave redistribution of MHCII to the cell surface were actually downregulated upon maturation. These genes were silenced in dendritic cells using shRNAs and MHCII expression and distribution were analyzed by flow cytometry and confocal microscopy. Silencing of six hits generated immature dendritic cells with a mature phenotype in terms of MHCII expression and distribution. Without the use of unbiased screening these genes would not have been linked to MHCII transport, illustrating the power of RNAi screening (5).

Future prospects

RNAi screening data can be used as a data mining tool. The advantage of siRNA screening is that every gene is a datapoint and can be treated as such. This makes it possible to compare different screens and to integrate different types of large scale analyses like expression data and mass-spec studies. This is very important as siRNA screens miss hits due to ineffective silencing or redundancy among genes. Yet, they represent a major step forward by adding a functional component that allows for identification of new proteins and pathways. Furthermore siRNA datasets can be used for drug discovery and the identification of new inhibitors for known genes. Here we discuss the use of our dataset for these purposes and speculate how screens in

general can be used for new data-based Systems Biology.

Data-based Systems Biology

siRNA-based screens provide a powerful tool for studying cell biological processes. Potential interacting proteins can be selected based on overlapping phenotypes, which represents shared pathways. This is a first step in the definition of a scaffold of pathways, which then has to be experimentally tested.

We exploited this route to identify proteins involved in controlling transport of MHCII containing vesicles to the plasma membrane in dendritic cells following maturation signals. Amongst the selected genes that control MHCII transport in immature dendritic cells is a small GTPase called Arl14 or Arf7. Arl14/ARF7 is a newly identified protein and is located on MIIC. Silencing Arl14/ARF7 in immature dendritic cells redistributes MHCII to the plasma membrane, a phenotype that is typically observed in mature dendritic cells. These observations were used as a starting point for building a pathway explaining the control of MHCII transport in dendritic cells.

Again, various sets of information were used to predict interacting proteins with Arl14/ARF7. A GEF, which activates Arl14/ARF7, was identified as these have so-called SEC7 domains. We screened our siRNA data sets for proteins with this domain and the same phenotype as Arl14/ARF7 silencing (ie upregulation of MHCII) and identified two candidates. In vitro testing revealed that only one, PSD4, activated Arl14/ARF7. The effector of Arl14/ARF7 was identified by performing a yeast two-hybrid. This provided one unknown protein, now called ARF7EP, which was then used as a bait to identify interaction proteins by mass spec. This yielded a myosin motor called Myo1e which is involved in the anchoring of vesicles to the actin cytoskeleton. Eliminating this pathway then allows the MIIC (and thus MHCII) to leave the intracellular location for surface deposition. This exemplifies how large data sets can be used to build new networks involved in biologically relevant processes, like MHCII transport in dendritic cells. This can be further expanded as the GAP and the proteins controlling the activity of the GEF are not identified. They may hide in the siRNA generated data sets and have to be found.

The siRNA screen yielded a wide variety of proteins; kinases, ring- and zinc finger proteins, proteins involved in the ubiquitination pathway, controllers of transcription etc. However, almost half of the hits represented new and unknown proteins. The latter are possibly the most interesting ones and yet the most difficult one to place in functional networks. The approach discussed above represents an unbiased selection allowing placing unknown proteins in networks. Using data-based Systems Biology approaches as described here enables the construction of new pathways based around single hits thus constructing new cell biology.

Data mining to feed Systems Biology

Another relevant use for screens involves data mining. Many labs check new screening datasets for the phenotype of their proteins of interest. This is helpful to place proteins in pathways and emphasizes the importance of a uniform publication format. It calls for an online database for uploading screening data (including the primary data) to allow easy and open access. In addition, it will be very important to arrive at proper statistical methods to assess the quality of screens (such as z-factors). Such open

networks would allow broad use of data sets and accelerate cell biology and can be set up like the GEO database at the NCBI website for microarray information.
A common data format would also allow data-based Systems Biology at a higher level. The results from different screens can be compared and used to construct pathways by combining the information. By for example comparing data sets where effects on cell surface receptors were measured, genes involved in different aspects of endocytosis can be clustered and used to extract pathways. By including negative data (a gene that does not give a phenotype) branches in pathways may be determined. For example when protein A and B belong to the same pathway in one screen while only protein A shows up in another screen, A and B should have different functions and pathways. By using a scoring system and comparing multiple screens, this may result in a more functional approach to elucidate branching of pathways in the cell. Obviously, such approaches will only allow prediction of pathways and these have to be experimentally tested.
Systems Biology usually integrates literature and microarray datasets. The advantage of including siRNA data is its functional and unbiased approach. With the publication of more siRNA datasets in the public domain, Systems Biology will convert into a data-based Systems Biology which improves the quality of predicted pathways.

Applying data sets to define new drug targets and new drug off-target effects

New proteins regulating MHCII expression provide new drug targets to manipulate MHCII antigen presentation. This is of interest for vaccines and treatment of auto-immune diseases that are related to MHCII. Usually enzymes and receptors are considered druggable targets. Some of our hits have defined inhibitors such as the kinase DYRK1a. DYRK1a is a kinase important in Down's syndrome and has never been linked to antigen presentation. DYRK1a depletion with siRNAs results in lower MHCII surface expression and this is expected for chemical inhibition of DYRK1a as well. By using genome wide screening data as a source for new targets, we are selecting compounds to manipulate MHCII antigen presentation.
Besides predicting inhibitors on the basis of the hit list, the genome wide screen can also be used as a template to identify new targets for drugs already used in the clinic. The effect of drugs can be placed in the siRNA data sets after phenotypic comparison of drug and siRNA in a multidimensional manner (combining flow cytometry, qPCR and microscopy). Drugs affecting MHCII expression were identified by screening a library containing 1500 FDA-approved drugs in a setting similar as the RNAi screen. Using this approach, several clinically used compounds were identified that affect MHCII expression and that have never been linked to manipulation of antigen presentation. By comparing the phenotypes of the drugs to these of the siRNAs, only a limited number of siRNA hits give a corresponding phenotype. These are the potential targets of the drugs. With this approach, we identified an inhibitor of CIITA. By using target prediction programs, target-lead combinations were made and current work focuses on validating these predictions to identify a new target for this drug.
These two approaches illustrate how genetic screening data can be used to improve drug discovery without prior definition of the target, only based on the biological

effect.

Conclusions

Genome-wide RNAi screening data can provide important insights into cell biological pathways and processes and identify genes that would otherwise not have been considered. A functional dataset can be used for many other purposes and should be accessible to the scientific community. This requires a common format and platform at which datasets can be uploaded and analyzed. Only then the data will be used at its maximum potential and serve system biologists, drug finders, immunologists and cell biologists to accelerate their research.

References

1. D. D. Chaplin, M. E. Kemp, The major histocompatibility complex and autoimmunity. Year Immunol 3, 179-198 (1988).
2. C. H. Chang, S. Guerder, S. C. Hong, W. van Ewijk, R. A. Flavell, Mice lacking the MHC class II transactivator (CIITA) show tissue-specific impairment of MHC class II expression. Immunity 4, 167-178 (1996).
3. J. M. Riberdy, J. R. Newcomb, M. J. Surman, J. A. Barbosa, P. Cresswell, HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. Nature 360, 474-477 (1992).
4. J. Neefjes, M. L. Jongsma, P. Paul, O. Bakke, Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11, 823-836 (2011).
5. P. Paul et al., A Genome-wide multidimensional RNAi screen reveals pathways controlling MHC class II antigen presentation. Cell 145, 268-283 (2011).