

LANDSCAPING THE FRAMEWORK OF BIO RESEARCH PROJECT

Generation of the 3D Atlas for Drug Target Discovery

Byung-Cheol Kim¹ and Sunghoon Kim^{1,2}

¹Medicinal Bioconvergence Research Center, 8th floor, B-dong, Advanced Institutes of Convergence Technology, Iui-dong, Yeongtong-gu, 443-270, Suwon, Gyeonggi-do, Republic of Korea

²College of Pharmacy, Seoul National University, San 56-1, Shillim-dong, Kwanak-gu, 151-742, Seoul, Republic of Korea

Keywords: Scientific information visualization, Project management, Drug target discovery.

Abstract: As the size of pharmacological research organizations is getting bigger, it is ever more critical to grasp the entire R&D scene as a vivid image, because a lack of such an image makes it hard to understand a project as a whole and thus to make high-level decisions. We provide one prototype of such a kind so that all the researchers of a project can intuitively share the current situation, recognize the bottlenecks, and collaborate with one another to resolve the problems. Our approach is to exploit the faculty of human vision, especially spatial perception and memory, by defining the axes of information space to be commensurable and interpolatable, making shapes of the process and data on those axes, and providing referential cues in the space.

1 INTRODUCTION

Over the past decade, companies and laboratories in the pharmaceutical industry are getting bigger ever and the mission to identify and validate a new drug target involves more complicated, different but correlated processes and data. Although an overwhelming number of data are gushing out from laboratories on a daily basis, stakeholders can hardly get the integral information to make high-level decisions like whether to increase or reduce the budget, the next-step R&D direction, even whether or not to kill the project, etc. In other words, a lack of the entire *image* of the situation could impede, defer or dismantle a huge target discovery project which has been funded hundreds of millions of dollars.

Since a human can only deal with just a few chunks of symbolic information at the same time due to the limitation of human memory, the only way to simultaneously deliver a number of data is to make them as images which can be recognized at once. The problem is that, in many cases, the information of scientific projects is heterogeneous and even non-geometric. Naive visualization of such information often leads to just another heavy and hardly-perceptible set of data elements. The data should be well organized in a geometrically informative way to exploit the faculty of human vision optimized for geometric shapes. Therefore, the crux of the problem

is how to define the information space in which the processes and data are visualized so that they can be seen as a kind of geometrically meaningful shapes.

We provide one prototype visualization framework, a 3-dimensional atlas for the drug target discovery, by establishing the criteria for defining such axes of information space and referential cues.

2 RELATED WORK

Traditionally, biomolecular networks and interactions have been visualized in two dimensions. Cytoscape is a representative software to layout and query biomolecular interaction networks extracted from high-throughput expression data and other molecular states (Shannon et al., 2003). It is still popular because of its ease of use to visualize a network of many nodes and links. The Systems Biology Graphical Notation (SBGN) is a kind of 2-dimensional(2D) circuit diagrams to represent networks of biochemical interactions (Novre et al., 2009). The Biological Connection Markup Language (BCML) is defined to describe biological pathways based on the SBGN (Beltrame et al., 2011). KEGG, Kyoto Encyclopedia of Genes and Genomes (<http://www.genome.jp/kegg/>), is a well-known database including pathway data and circuit-like diagrams. These approaches are very useful to find out specific mechanisms and related data.

Table 1: An example target inventory in a tabular form. There are too many data items to examine in order to comprehend the status of the entire drug target inventory. In addition, the number of targets and their properties can rapidly grow even to hundreds and thousands. (Quantification of non-digit information should be defined carefully in accordance with specific categories of data).

No	TargetID	Disease Indication	Mode of Action	Structure	Assay	Biomarker	Lead Availability	Clinical Validation
1	AIMP2	NSCLC	Yes(1.0)	No(0)	N/A	0.6	No(0)	Yes(1.0)
2	CDH6	Renal Cancer	Yes(1.0)	0.5	Yes(1.0)	0.1	0.9	0.5
3	AIMP3	Lymphoma	0.7	Yes(1.0)	0.3	No(0)	0.4	0.2
4	CD23	RA	Yes(1.0)	No(0)	Yes(1.0)	No(0)	0.1	No(0)
5	Gp96	Lupus	Yes(1.0)	No(0)	0.8	Yes(1.0)	0.9	Yes(1.0)

But they always provide all the detailed information or the entire data at the same time, which cannot be comprehended figuratively when the number of data becomes bigger than even hundreds.

Thus, for the large number of data, systems have been developed beyond 2D planar visualization to fully exploit the 3-dimensional(3D) human perceptual system. Arena3D was suggested as layered visualization of biological networks in 3D (Pavlopoulos et al., 2008). The layers make a group of nodes distinctive to the other groups. FORG3D provides a force-directed 3D graph editing capability for visualization of genome scale data (Jussi Paananen, 2009). Nodes and their links are automatically positioned by simulating them as a force-directed mesh grid. Interactorium(Skyrails) is a software platform to visualize protein-protein interaction and to navigate the interaction link in 3D (Widjaja et al., 2009). It is based on a scriptable command system so that the user can program his/her own menus and visualization methods. Systems and methods above provide 3D visualization frameworks but still draw all the raw data which can be too big to take as informative images.

Out of nodes and links, there has been another approach to provide a kind of prospect on the information as a terrain map or landscape. A gene expression map for *C. elegans* was created by drawing it as terrain in 2.5D (xy-plane and z-elevation) (Kim et al., 2001). GeneTerrain is a kind of 2.5D terrain map which focuses on visualizing the difference map of expression data (You et al., 2008). Recently, Pfizer's research centre developed a drug target landscape, a visualization system to convey the overall meaning of an integrated set (Campbell et al., 2010). It produces a concept of zone in which the plotted data are grouped geometrically and meaningfully. This kind of approach can convert data into figuratively informative things. Thus we try to elaborate it further in the next section by establishing some principles for such visualization.

3 PRINCIPLES

There are two kinds of consideration to visualize non-geometric information. One is how to draw the information and the other is how to watch it. More specifically on the former, it is about where to position each data in space. That leads to the definition of the axes of information space on which data are plotted to be viewed with perceptual ease.

The axes of information space for perceptual visualization should be

Commensurable. The data item visualized on the same axis should be the same *kind*, conceptually at least. For example, a mark pointing to an item on the left side must be able to be said less or more '...' in a certain quality than one on the right side. The bounded length or elongation of plotted items can express the level, state, and/or quality of the axial information. Then, this might help the user's recognition of the information as a single factor instead of a mere collection of items.

Interpolatable. Each item on the commensurable axis must be able to be tightly coupled with its neighbours to the extent that any point between two items can be meaningfully drawn by mixing the values of the two. For instance, a point value between two given point values, p_1 and p_2 , can be positioned with the mixing parameter, t , as follows:

$$p(t) = (1-t)p_1 + tp_2 \quad (0 \leq t \leq 1)$$

Consequently, the point, $p(t)$, describes a line which is a *geometric object out of non-geometric data*.

Therefore, the plotted data in the 2D or 3D information space based on such axes can be seen as a *shape*, which is easy to remember and even get some intuition about future R&D directions, because the user can think of it not as a set of data but as a sort of shape. For example, growing or shrinking of the shape can convey the feeling of aliveness of a project. Hence the user can consider how to rescue it by observing and analyzing the evolution of the shape. The user can also see the evolving history by visualizing

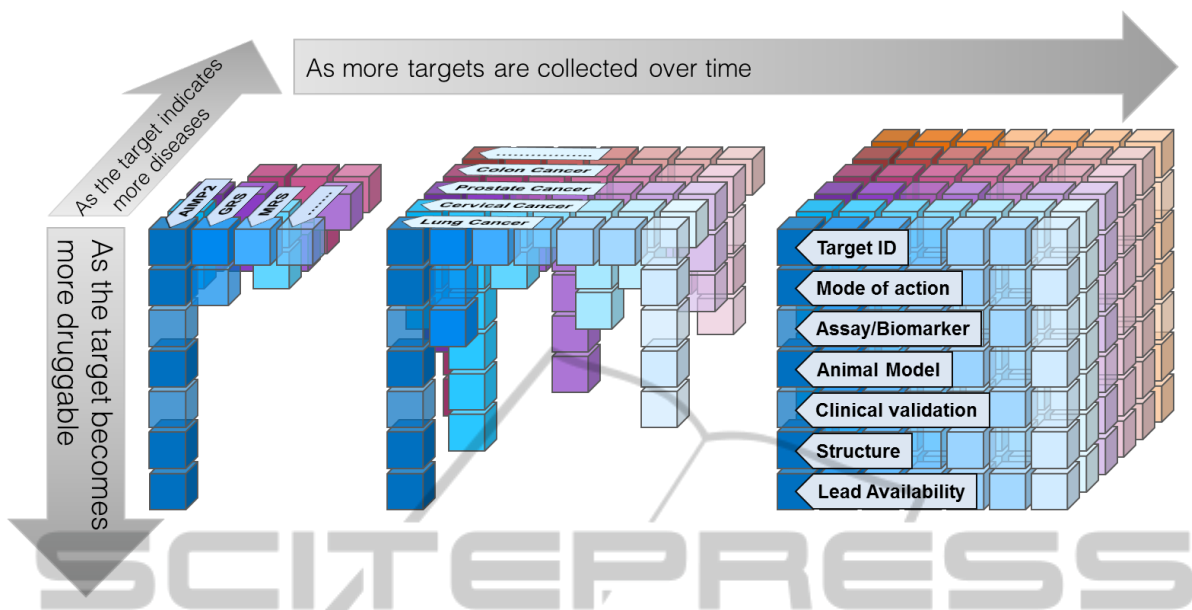


Figure 1: An example 3D atlas for identification and validation of druggable targets on the three axes, i.e., targets, druggability, and disease indication. Each data item is drawn as a cube which has non-geometric properties such as brightness, hue, and transparency for the sort of targets, the kind of diseases, and the degree of progress, respectively.

the data over time.

The latter is about how to facilitate spatial perception. The essence of spatial perception comes from movement, i.e., translation and rotation. One can feel the position by translating and the orientation by rotating. Perception of the position and orientation of an object is generally relative to something else, more formally, fiducial points or referential objects. Hence navigating the information space without any reference would give the user a chance to fall into a pitfall of *lost-in-space*. There can be two kinds of scheme to provide spatial reference, intrinsic and extrinsic. The intrinsic reference is a unique geometric and/or colored parts of the entire shape. So the user can catch up his or her relative position by remembering the (pattern of) position of those unique parts. The extrinsic reference is a kind of compass positioned in some corner of the space or the user's screen. The user can instantly recognize the current position and orientation in the space. The former is good to apply when there are few objects to inspect and they are simple enough to embed additional unique shapes and/or colors. On the other hand, the latter would be better if there are too many objects to view in the space.

4 CASE STUDY: A TARGET DISCOVERY PROCESS

We have been developing a 3D atlas to visualize the target inventory. For instance, a target inventory can be described in a tabular form such as Table 1. The number of targets listed up will be grow rapidly and there can be more properties like *in vivo* efficacy of lead, disease model, patent, and so on. All the items in this tabular form, however, can hardly be retained together in the human memory in an organized form. Thus, in order to make them as images to be perceived at a moment, we defined three axes of information space, targets (x-axis), druggability (y-axis), and indication of disease (z-axis).

The x-axis is for targets, i.e., items of the same x-coordinate indicate properties of the same drug target. Neighboring targets are not randomly placed but arranged by their functional proximity. So the targets on the x-axis evolve from a target niche of primary interest to other targets that show functional proximity to the initial targets. The y-axis represents the maturity of each target with respect to druggability. In other words, the longer the elongation on the y-axis is, the more druggable the target becomes. The axial information covers seven categories such as a target ID, mode of action, assay/biomarker, animal model, structure, availability of lead, and clinical validation. The z-axis is for diseases for which a target can be a

new drug candidate. Then, a target's indication of disease will marshal on the z-axis from diseases of initial concern to other pathologically highly-connected diseases.

Therefore, these axes create a unique space for the evolving target inventory. Each data item is drawn as a cube on those axes so that a set of cubes can be perceived as a big cube being built up like Lego blocks. The solidness and volume of the big cube can convey the feeling of completeness and richness of validated druggable targets. A cube has three non-geometric properties, transparency, hue, and brightness which represent the degree of progress, the kind of diseases, and the sort of targets, respectively. These hue and brightness can function as the referential cue when translating and rotating the entire cube. Figure 1 depicts an example 3D atlas for drug target discovery.

5 CONCLUSIONS

The problem itself is simple. The entire image of the R&D scene is needed and should be generated based on real data. The constraint is that the image ought to be geometrically informative to exploit human spatial perception. Thus the properties of the information axes are defined so that a collection of data items can appear as a geometric shape. To appreciate and remember the shape concretely, a referential cue is suggested to be an anchor point in the information space in which the user can often be confused or lost.

The suggested prototype has partly shown the proof of concept for the drug target discovery process in 3D space. But the information axes of the 3D atlas should be elaborated to be more commensurable and interpolatable. Finally, to more facilitate the user's spatial perception, motion guidance for inspecting the 3D atlas such as navigational paths along the pathological connections will be developed.

ACKNOWLEDGEMENTS

This research was supported by Global Frontier Research Grant NRF-M1AXA002-2010-0029785 of the National Research Foundation funded by the Ministry of Education, Science and Technology of Korea.

REFERENCES

Beltrame, L., Calura, E., Popovici, R. R., Rizzetto, L., Guedez, D. R., Donato, M., Romualdi, C., Draghici,

- S., and Cavaliere, D. (2011). The biological connection markup language: a sbgn-compliant format for visualization, filtering and analysis of biological pathways. *Bioinformatics*, 27(15):2127–2133.
- Campbell, S. J., Gaulton, A., Marshall, J., Bichko, D., Martin, S., Brouwer, C., and Harland, L. (2010). Visualizing the drug target landscape. *Drug Discovery Today*, 15(1/2):3–15.
- Jussi Paananen, G. W. (2009). Forg3d: Force-directed 3d graph editor for visualization of integrated genome scale data. *BMC Systems Biology*, 3:26.
- Kim, S. K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J. M., Eizinger, A., Wylie, B. N., and Davidson, G. S. (2001). A gene expression map for caenorhabditis elegans. *Science*, 293:2087–2092.
- Novre, N. L., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., Bergman, F. T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villiger, A., Boyd, S. E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T. C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D. B., Sander, C., Sauro, H., Snoep, J. L., Kohn, K., and Kitano, H. (2009). The systems biology graphical notation. *Nature Biotechnology*, 27:735–741.
- Pavlopoulos, G. A., O'Donoghue, S. I., Satagopam, V. P., Soldatos, T. G., Pafilis, E., and Schneider, R. (2008). Arena3d: visualization of biological networks in 3d. *BMC Systems Biology*, 2:104.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13:2498–2504.
- Widjaja, Y. Y., Pang, C. N. I., Li, S. S., Wilkins, M. R., and Lambert, T. D. (2009). The interactorium: visualising proteins, complexes and interaction networks in a virtual 3d cell. *Proteomics*, 9(23):5309–5315.
- You, Q., Fang, S., and Chen, J. Y. (2008). Geneterrain: visual exploration of differential gene expression profiles organized in native biomolecular interaction networks. *Information Visualization*, 9:1–12.