Use of Computational Functional Genomics in Drug Discovery and Repurposing for Analgesic Indications

Jörn Lötsch[1,2] and Dario Kringel[1]

The novel research area of functional genomics investigates biochemical, cellular, or physiological properties of gene products with the goal of understanding the relationship between the genome and the phenotype. These developments have made analgesic drug research a data-rich discipline mastered only by making use of parallel developments in computer science, including the establishment of knowledge bases, mining methods for big data, machine-learning, and artificial intelligence, (Table 1) which will be exemplarily introduced in the following.

CURRENT EVIDENCE OF FUNCTIONAL GENOMICS APPROACHES IN ANALGESIC DRUG RESEARCH

The study of variants that modulate the perception of pain or the response to analgesics, and in particular experiments in transgenic mice, have so far identified more than 500 genes as being implicated in the modulation of pain-related phenotypes.[1] To approach pain at a genome-wide level, functional genomics is used to combine data derived from various processes related to DNA sequence, gene expression, and protein function, such as coding and noncoding transcription, protein translation, protein-DNA, protein-RNA, and protein-protein interactions. A literature search in the PubMed database (**Table 1**) on October 5, 2017 for "(functional AND (genomic OR genomics) AND (pain OR *nocicepti* OR hyperalgesi* OR allodyni*) AND pharmacol* AND (genome OR transcriptome OR proteome OR metabolome OR interactome) NOT review" obtained 112 hits. Most referred to genome-wide expression analyses or genetic association studies of small groups of genes, while the use of functional genomics for the explicit purpose of drug discovery and repurposing for analgesic indications counted in only a few major approaches, including a phenotypic approach that will be highlighted in the following, while further approaches included the use of next-generation sequencing and functional proteomics for analgesic target identification.

FUNCTIONAL GENOMICS-BASED PHENOTYPIC APPROACHES TO ANALGESIC DRUG RESEARCH

Phenotypic drug discovery approaches try to address incompletely understood complexities of diseases[2] but do not rely on knowledge about specific drug targets or hypotheses about their particular roles in pain, for example. This contrasts with the widely used target-based strategies of drug discovery and repurposing but has received increasing interest in recent years.[2] As pain has been generally accepted as a very complex trait, functional genomics-based phenotypic approaches qualify for drug research and repurposing in the field. Combining several lines of research (**Table 1**) with machine-learning methods obtained the recently introduced framework of "process pharmacology."[3] This can be regarded as a phenotypic concept that puts the disease rather than the molecular drug target in the focus of drug research and therapy. Adopting the definitions of the GeneOntology database (**Table 1**), it regards pain as a result of alterations of the activity in biological processes, defined as collections of molecular functions involving chemical or physical transformations such as cell growth and maintenance or signal transduction.

Functional genomics-based analgesic drug classification

A functional genomics-based criterion of drug classification proposed recently[3] combines several big-data based sources of information. Specifically, the drug targets, respectively their genetic determinants, are accessible in worldwide available databases such as the DrugBank database (**Table 1**). The biological processes in which the drug target coding genes are involved were queried from further knowledge bases with a current gold-standard being the Gene Ontology (GO) database (**Table 1**). A "drug target versus biological process" matrix was constructed comprising $n = 79$ classical analgesics, i.e., opioids and nonopioids such as nonsteroidal antiinflammatory drugs and related classical analgesics as queried from the DrugBank database, in which these drugs were

[1]Institute of Clinical Pharmacology, Goethe University, Frankfurt am Main, Germany; [2]Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine and Pharmacology TMP, Frankfurt am Main, Germany. Correspondence: Jörn Lötsch (j.loetsch@em.uni-frankfurt.de)

Received 9 October 2017; accepted 28 November 2017; advance online publication 19 January 2018. doi:10.1002/cpt.960

Table 1 Overview on data sources and computational tools used for the present data science approach to analgesic drug repurposing from knowledge about the functions of genes related to insensitivity to pain in humans

	Site name	URL
Gene names and functions	AmiGO (search utility for GO)	http://amigo.geneontology.org/
	Human Pain Genes Database	https://humanpaingenetics.org/hpgdb/
	Gene Ontology (GO)	http://www.geneontology.org/
	HUGO Gene Nomenclature Committee	http://www.genenames.org/
	NCBI gene index database	http://www.ncbi.nlm.nih.gov/gene/
	GeneCards	http://www.genecards.org
Diseases genes	Pain Genes Database	http://www.jbldesign.com/jmogil/enter.html
	Online Mendelian Inheritance in Man (OMIM) database	http://www.ncbi.nlm.nih.gov/omim
Drugs	DrugBank database	http://www.drugbank.ca
	Thomson Reuters Integrity database	https://integrity.thomson-pharma.com
	Registry and results database of federally and privately supported clinical trials	http://www.clinicaltrials.gov/
Reported biomedical evidence	PubMed database	https://www.ncbi.nlm.nih.gov/pubmed
Software	Gene Trail	http://genetrail.bioinf.uni-sb.de/
	R software	http://CRAN.R-project.org/

All recourses except one (Thomson Reuters Integrity database) are publicly available, most of them free of charge. They were accessed on October 8, 2017.

associated with $n = 102$ genes, respectively molecular targets. Via querying the GeneOntology database, these genes were associated with $d = 928$ biological processes using overrepresentation analysis, which compared the occurrence of the particular set of genes covered by a GO term with the number of genes expected to be annotated to this term, and uses Fisher's exact statistics to test the statistical significance of the deviation from the expectation, as explained in more detail previously.[4] Unsupervised machine-learning was used to identify structures within this data space. Specifically, each drug was represented as a vector in a $d = 928$ dimensional feature space of impact strengths on processes. To explore this feature space, topographic mapping was used, which provides data projection methods to create low-dimensional images from high-dimensional data. Specifically, the high-dimensional information was projected onto a two-dimensional grid of artificial neurons on a self-organizing map used previously.[3,4] Following calculation of a so-called U-matrix, which visualizes the distances between artificial neurons as a third dimension, two clusters appeared separated by a "mountain ridge" when using a topographical map analogy (**Figure 1**, left). The clusters perfectly coincided with the two major classes of classical analgesics.

To demonstrate this method, in the following experiment an alternative data projection method was used for nonredundancy.[4] Specifically, topographic mapping was implemented as swarm intelligence, i.e., an algorithm guided by the flocking behavior of numerous independent but cooperating so-called "DataBots," which are self-organizing artificial "life forms" identified with single data objects (analgesics). These "DataBots" can move on a two-dimensional grid, and their movements are either random or follow the attractive or repulsive forces proportionally to the (dis-)similarities of neighboring "DataBots." The data space $D=\{x_{i,d},i=1,...n\}\subset\mathbb{R}^d$ comprising $d = 928$ biological processes associated with the $n = 79$ analgesic drugs was explored for distance-based structures. A parameter-free projection method of a polar swarm of "DataBots," *Pswarm*, was used. Following successful swarm learning, "DataBots" carrying items with similar features were located in groups on the projection grid. After calculation of a U-matrix, two clusters of analgesics emerged (**Figure 1**) that correctly reflected the classification of the analgesics into two main classes of opioid and nonopioid analgesics, reproducing the results obtained previously using an emergent self-organizing map.[4] The classification was flawless and corrected the classification provided by a domain expert for a few uncommon opioids such as alvimopan. The calculations were performed using the R library "DatabionicSwarm" (M. Thrun, https://cran.r-project.org/package=DatabionicSwarm). Finally, the projected data clusters were validated using Ward's method (**Figure 1**, right).

Functional genomics-based analgesic drug repurposing

Repurposing screens using novel molecular techniques such as reprogrammed nociceptor neurons currently shift the trend from target-based to pathway-based repurposing, supported by the inclusion of computational techniques and online resources. Within the present phenotypic concept, computational functional genomics approaches for analgesics drug repurposing may employ the association of drugs with biological processes. However, using the complete set of more than 500 pain-related genes[1] appears to be a too heterogeneous basis for such screens, which suggests the analysis of functionally more focused subsets of pain genes such as the currently known

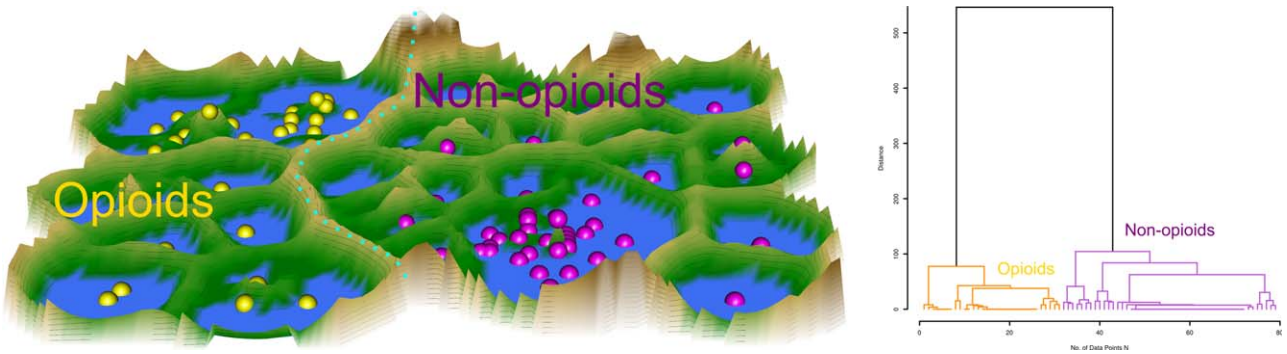

Figure 1 Structure found using unsupervised machine-learning in the high-dimensional data space of the analgesic drug ($n = 79$) vs. computational functional genomics based biological processes ($d = 928$) matrix. Left: The so-called U-matrix displays the result of a projection of the drug vs. biological process interaction matrix onto a toroid neuronal grid where opposite edges are connected. The projection was obtained using a parameter-free polar swarm, *Pswarm*, consisting of so-called DataBots, which are self-organizing artificial "life forms" that carry vectors of the biological processes associated with the drugs via their genetic targets. During the learning phase, the DataBots were allowed to adaptively adjust their location on the grid close to DataBots, according to the Jaccard distance, carrying data with similar features, with a successively decreasing search radius. When the algorithm ends, the DataBots become projected points. To enhance the emergence of data structures on this projection, a generalized U-matrix displaying the distance in the high-dimensional space was added as a third dimension to this visualization. The U-matrix was colored in hypsometric colors making the visualization appear as a geographical map with brown heights and green valleys with blue lakes. Watersheds indicate borderlines between different groups of analgesic drugs. In the present visualization, a curved "mountain range" in the "north–south" direction (marked with a light blue dotted line) separates two main clusters of drugs. These clusters completely coincided with the prior classification of analgesics into opioids and nonopioids subjects according to the pattern of repeated cold pain measurements. The data points are colored according to the emerging two-cluster structure. Right: Ward clustering of the projected data clearly also indicated two clusters, supporting the machine results. The figure was created using the R software package (v. 3.4.2 for Linux; http://CRAN.R-project.org/), in particular the libraries "DatabionicSwarm" (M. Thrun, https://cran.r-project.org/package=DatabionicSwarm). The figure reproduces results of a previous analysis of the same data matrix; however, using a different machine-learning method for nonredundancy.

$n = 22$ genes causally involved in human insensitivity to pain,[5] which are regarded to provide a particularly suitable basis for analgesic drug discovery and repurposing. A computational functional-genomics analysis, performed analogously to the assessments described above, identified processes related to nervous system development and to ceramide and sphingosine signaling pathways as particularly important biological functions of this set of genes.[5] This is in line with suggestions from other approaches to use these pathways as novel therapeutic targets in pain. Following establishment of the functional genomics of hereditary insensitivity to pain, the biological processes were used for a similarity analysis with the functional genomics of database-queried drugs using unsupervised machine-learning, as above. The analysis identified a cluster of $n = 22$ drugs that shared important functional genomic features with hereditary insensitivity to pain. For more than half of the members of this cluster, evidence about an implication in pain could be found in the literature. While it appears unlikely that this will be true for any random set of drugs, a statistically significant difference of the findings with positive hits expected by chance was not tested in that proof-of-concept assessment. By contrast, using the present method to identify pain-relevant genes,[1] using a set of 34 hits, 33 could be supported by empirical evidence, whereas for a random set of 34 genes, only three hits were obtained, suggesting that the method may be suitable for analgesics drug repurposing.

CONCLUSION

Functional genomics enables genome-wide approaches to pain and analgesic drug research. Based on recent technological advances in laboratory data acquisition and data science, the complexity of pain can be approached at an adequately complex research level. Current sparse implementations in analgesic drug discovery and repurposing consist mainly of proteomic, next-generation sequencing and computational phenotypic drug research approaches. In particular, the computational approaches increasingly make use of machine-learning, knowledge discovery in big data, and artificial intelligence. While working in concert with statistics, which can be regarded as a branch of mathematics, these methods have been developed from computer science, which gains increasing importance in pain research. Among limitations is the vulnerability to poor data quality and the dependence on correctness, completeness, and regular maintenance of databases. First results have been presented supporting that computational functional genomics may provide a powerful approach to exploit the biological space of undrugged or unknown targets and poorly understood disease mechanisms and to provide a route to innovative analgesic treatments.

FUNDING

This work was funded by the Landesoffensive zur Entwicklung wissenschaftlich – ökonomischer Exzellenz (LOEWE), LOEWE-Zentrum für Translationale Medizin und Pharmakologie (JL) with the specific project funding under the name "Process pharmacology: A data science based approach to drug repurposing" (JL) and by the European Union Seventh Framework Programme (FP7/2007 - 2013) under grant agreement no. 602919 (JL, GLORIA). The funders had no role in the decision to publish or the preparation of the article.

CONFLICT OF INTEREST

The authors declare there are no competing interests.

1. Ultsch, A., Kringel, D., Kalso, E., Mogil, J.S. & Lötsch, J. A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity. *Pain* **157**, 2747–2757 (2016).
2. Moffat, J.G., Vincent, F., Lee, J.A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
3. Lötsch, J. & Ultsch, A. Process pharmacology: a pharmacological data science approach to drug development and therapy. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 192–200 (2016).
4. Lötsch, J. & Ultsch, A. A machine-learned computational functional genomics-based approach to drug classification. *Eur. J. Clin. Pharmacol.* **72**, 1449–1461 (2016).
5. Lötsch, J., Lippmann, C., Kringel, D. & Ultsch, A. Integrated computational analysis of genes associated with human hereditary insensitivity to pain. A drug repurposing perspective. *Front. Neurosci.* **10**, 252 (2017).