Integrated Computational Analysis of Genes Associated with Human Hereditary Insensitivity to Pain. A Drug Repurposing Perspective

Jörn Lötsch 1,2*, Catharina Lippmann 2, Dario Kringel 1 and Alfred Ultsch 3

1 Institute of Clinical Pharmacology, Goethe-University, Frankfurt am Main, Germany, 2 Fraunhofer Institute of Molecular Biology and Applied Ecology-Project Group, Translational Medicine and Pharmacology (IM-E-TMP), Frankfurt am Main, Germany, 3 DataBionics Research Group, University of Marburg, Marburg, Germany

Genes causally involved in human insensitivity to pain provide a unique molecular source of studying the pathophysiology of pain and the development of novel analgesic drugs. The increasing availability of “big data” enables novel research approaches to chronic pain while also requiring novel techniques for data mining and knowledge discovery. We used machine learning to combine the knowledge about \( n = 20 \) genes causally involved in human hereditary insensitivity to pain with the knowledge about the functions of thousands of genes. An integrated computational analysis proposed that among the functions of this set of genes, the processes related to nervous system development and to ceramide and sphingosine signaling pathways are particularly important. This is in line with earlier suggestions to use these pathways as therapeutic target in pain. Following identification of the biological processes characterizing hereditary insensitivity to pain, the biological processes were used for a similarity analysis with the functions of \( n = 4,834 \) database-queried drugs. Using emergent self-organizing maps, a cluster of \( n = 22 \) drugs was identified sharing important functional features with hereditary insensitivity to pain. Several members of this cluster had been implicated in pain in preclinical experiments. Thus, the present concept of machine-learned knowledge discovery for pain research provides biologically plausible results and seems to be suitable for drug discovery by identifying a narrow choice of repurposing candidates, demonstrating that contemporary machine-learned methods offer innovative approaches to knowledge discovery from available evidence.

Keywords: data science, computational biology, pain, humans, genetic variation, machine learning, perception, big data

INTRODUCTION

Persistent pain is a major healthcare issue, as defined by WHO, affecting about a fifth of the European population increasing to a third in the over 70-year old (Elliott et al., 1999; Breivik et al., 2006). It has a highly complex pathophysiology (Julius and Basbaum, 2001; Schaible, 2007) and it is triggered by several different causes, such as cancer (Portenoy, 1992) and surgery (Kehlet et al., 2006). Therefore, the search for novel analgesic strategies is an active research topic receiving public funding (Kringel and Lötsch, 2015). In addition to the identification of novel drug
targets from molecular research as the main line of research, scanning the existing pharmacopeia for repurposing candidates becomes increasingly successful (Ashburn and Thor, 2004). This is facilitated by developments in computational data science (President’s Information Technology Advisory Committee, 2005).

An accepted source of novel options for the pharmacological treatment of (persistent) pain is the study of genes causally involved in hereditary syndromes with insensitivity to pain (Goldberg et al., 2012; Table 1). This set of genes has provided the targets of novel analgesics (Table 2). However, this unique set of genotypes also enables the study of key biological processes of pain from the perspective of the functions of the involved genes. As shown recently, similarities between the profiles of biological processes in which the genes coding for the targets of available drugs with the profile of the biological processes in which a given set of genes is involved can be employed for drug classification or repurposing (Lötsch and Ultsch, 2016a,b). Hence, picturing the genetic background of human insensitivity to pain could be explored for drug repurposing based on functional similarity in addition to the development of novel substances targeting some of the respective gene products (Table 2). The present analysis made extensive use of computational biology, knowledge discovery methods, publicly available databases and data mining tools (Table 3) to merge results from pain, genetics and pharmacological research.

| Gene   | NCBI   | Gene product       | Syndrome                                                                 | Syndrome code | OMIM   | References       |
|--------|--------|--------------------|--------------------------------------------------------------------------|----------------|--------|------------------|
| ATL1   | 51062  | Atlastin GTPase 1  | Neuropathy, hereditary sensory, type ID                                  | HSN1D          | 613708 | Guelly et al., 2011 |
| ATL3   | 25923  | Atlastin GTPase 3  | Neuropathy, hereditary sensory, type IF                                  | HSN1F          | 609369 | Kornak et al., 2014 |
| CLTCL1 | 8218   | Clathrin heavy chain like 1 | Insensitivity to Pain with Preserved Temperature sensation | EMPF1          | 614388 | Sheffer et al., 2016 |
| DNML1  | 10059  | Dynamin 1 like     | Encephalopathy, lethal, due to defective mitochondrial peroxisomal fission 1 | HADDS          | 617330 | Chao et al., 2017   |
| EBF3   | 253738 | Early B-cell factor 3 | Hypotonia, ataxia, and delayed development syndrome | HADDS          | 617330 | Chao et al., 2017   |
| FAM134B| 54463  | Family with sequence similarity 134, member B | Neuropathy, hereditary sensory and autonomic, type IIIB | HSN2B          | 613115 | Kurth et al., 2009  |
| IKBKAP | 8518   | Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein | Dysautonomia, familial (Riley-Day syndrome) | HSN III        | 223900 | Anderson et al., 2001; Slaugenhaupt et al., 2001 |
| KIF1A  | 547    | Kinesin family member 1A | Hereditary sensory neuropathy type IIC | HSN2C          | 614213 | Riviere et al., 2011 |
| LIFR   | 3977   | Leukemia inhibitory factor receptor alpha | Congenital pain insensitivity phenotype with progressive vertebral destruction | HADDS          | 617330 | Chao et al., 2017   |
| MPV17  | 4358   | Mpv17 mitochondrial inner membrane protein | Navajo neurohepatopathy | NNH            | 256810 | Karadimas et al., 2006 |
| NGF    | 4803   | Nerve growth factor (beta polypeptide) | Neuropathy, hereditary sensory and autonomic, type V | HSN V          | 608654 | Einarsdottir et al., 2004 |
| NTRK1  | 4914   | Neurotrophic tyrosine kinase, receptor, type 1 | Sensitivity to pain, congenital, with anhidrosis | HSN IV         | 256800 | Indo et al., 1996    |
| PRDM12 | 59335  | PR/SET domain 12 | Insensitivity to Pain with hypohidrosis | HSN VIII       | 616488 | Chen et al., 2015  |
| RAB7A  | 7879   | RAB7A, member RAS oncogene family | Charcot-Marie-Tooth type 2B neuropathy | CMT2B          | 600882 | Verhoeven et al., 2003; Janssens et al., 2014 |
| SCN1A  | 11280  | Nav1.9 (sodium voltage-gated channel alpha subunit 11) | Neuropathy, hereditary sensory and autonomic, type VII | HSN VII        | 615548 | Leipold et al., 2013; Woods et al., 2015 |
| SCN9A  | 6335   | Nav1.7 (sodium channel, voltage-gated, type II, alpha subunit 9) | Insensitivity to pain, chanelopathy-associated | HSN VII        | 243000 | Cox et al., 2006  |
| SPTLC1 | 10558  | Serine palmitoyltransferase, long chain base subunit 1 | Neuropathy, hereditary sensory and autonomic, type IA | HSN1A          | 162400 | Dawkins et al., 2001 |
| SPTLC2 | 9517   | Serine palmitoyltransferase, long chain base subunit 2 | Neuropathy, hereditary sensory and autonomic, type IC | HSN1C          | 613640 | Rotthier et al., 2009 |
| TTR    | 7276   | Transthyretin | Carpal tunnel syndrome, familial | HADDS          | 617330 | Chao et al., 2017   |
| WNK1   | 65125  | WNK lysine deficient protein kinase 1 | Neuropathy, hereditary sensory and autonomic, type II (Morvan disease) | HSN2A          | 201300 | Lafreniere et al., 2004 |

The genes were queried on March 13, 2017 from the “Online Mendelian Inheritance in Man” (OMIM) database at https://omim.org/ and the GeneCards database at http://www.genecards.org (Pitha et al., 1997).

HSAN, Hereditary sensory and autonomic neuropathy; HSN, Hereditary sensory neuropathy; OMIM, Online Mendelian Inheritance in Man database.
Table 2 | Novel analgesic drugs developed with the purpose to antagonistically target genes associated with human hereditary insensitivity to pain, i.e., to mimic the pain-insensitivity phenotype observed in carriers of loss-of-function mutations in these genes, being currently in a clinical phase of development according to publicly available sources of information (Table 3).

| Gene          | Pharmacological target                                                                 | Drug               | Action       | Company                          |
|---------------|----------------------------------------------------------------------------------------|--------------------|--------------|----------------------------------|
| NGF           | Nerve growth factor (beta polypeptide)                                                 | Tanezumab          | Antibody    | Pfizer                           |
| NTRK1         | Neurotrophic tyrosine kinase, receptor, type 1                                           | ASP-7962           | Blocker      | Astellas Pharma                  |
| SCCN9A        | Na⁺v,1.7 (sodium channel, voltage-gated, type IX, alpha subunit 9)                      | Funapide           | Blocker      | Xenon Pharmaceuticals             |
|               |                                                                                        | CC-8464            | Blocker      | Chromocell                       |
|               |                                                                                        | DSP-2230           | Blocker      | Daiippon Pharma                  |
|               |                                                                                        | GDC-0310           | Blocker      | Genentech/Xenon Pharmaceuticals   |

The information was queried on March 31, 2017 from the Thomson Reuters Integrity database at https://integrity.thomson-pharma.com.

**METHODS**

Data were analyzed using the R software package (version 3.3.3 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008). Following querying the relevant sets of genes by mining publicly available databases, the analyses aimed at (i) identifying the systems biology of hereditary insensitivity to pain as conferred by the functions of the causally involved genes and (ii) to explore whether this knowledge can be employed for drug repurposing approaches aimed at identifying potential candidates for the treatment of pain. The analytical steps are summarized in Figure 1 and described in full detail in the following paragraphs.

**Data Mining**

Genes involved in several different syndromes sharing the common phenotype of insensitivity to pain were queried on March 27, 2017 from the “Online Mendelian Inheritance in Man” (OMIM) database at (Online Mendelian Inheritance in Man, OMIM®, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA) at https://omim.org/ and the GeneCards database at http://www.genecards.org (Rebhan et al., 1997) for “insensitivity to pain.” In addition, the medical literature was searched in the PubMed database at https://www.ncbi.nlm.nih.gov/pubmed. This provided a set of $n = 20$ genes (Table 1) that included genes causing human hereditary sensory and autonomic neuropathies (HSAN) and further genes for which published evidence is available for a causal implication with the common phenotype of insensitivity to pain.

To assess whether the known functions of the genes associated with human hereditary insensitivity to pain could be used for drug repurposing, genes coding for known molecular targets of known drugs were taken from the DrugBank database (version 5.0; http://www.drugbank.ca; Wishart et al., 2006, 2008). Specifically, a query of the DrugBank database on March 13, 2017, identified 4,834 drug entries, including 4,630 FDA-approved small molecule drugs, interacting with 2,215 unique targets. This provided the 4,834 × 2,215 “drug vs. gene” matrix.

**Assessment of the Functions of Genes Associated with Insensitivity to Pain**

The biological functions in which the products of genes associated with insensitivity to pain are involved were queried from the Gene Ontology knowledge base (GO; http://www. geneontology.org/; Ashburner et al., 2000). In this database, knowledge about the functions of genes is stored using a controlled and clearly defined vocabulary of GO terms (Camon et al., 2003) allocated to each gene (Camon et al., 2004). The GO database is searchable for three major categories, consisting of biological process, cellular component and molecular function. As the best representation of processes affected in hereditary insensitivity to pain as a potential source of therapeutic approaches, the GO category biological process was selected. According to the GO database, this category contains one or more ordered collections of molecular functions involving chemical or physical transformations such as cell growth and maintenance or signal transduction (Ashburner et al., 2000).

To capture biological processes that are particularly relevant to the present gene set, while eliminating data noise from common processes, the set of pain insensitivity genes was submitted to over-representation analysis (ORA) (Backes et al., 2007). This compared the occurrence (as defined by its annotation term) of the particular set of genes covered by a GO term with the number of genes expected to be defined by this term. The significance of the association of a GO term with the expected list of genes was determined by means of a Fisher’s exact test (Fisher, 1922). The ORA attributed $p$-values to the resulting GO terms. A $p$-value threshold, $t_p$, of 0.05 was applied and corrected for multiple testing using the false discovery rate (Benjamini and Hochberg, 1995).

The result was a representation of the complete knowledge about the biological roles of genes associated with insensitivity to pain in the form of a directed acyclic graph (DAG; Thulasiraman and Swamy, 1992). This depicted the biological processes as a higher level of organization of genes and signaling pathways described in a polyhierarchical structure where the processes are connected to each other by “is-a,” “part-of,” and “regulates” relationships (Ashburner et al., 2000). This result was used as the basis for a description of the functions of genes associated with insensitivity to pain. However, the complete DAG usually contained 136 GO terms that eluded intuitive interpretation (Miller, 1956). Therefore, the obtained results were transformed into a more intelligible form using the method...
TABLE 3 | Overview on data sources and computational tools used for the present data science approach to drug repurposing from knowledge about the functions of genes related to insensitivity to pain in humans.

| Site name | URL | References |
|-----------|-----|------------|
| Gene names and functions | AmiGO (search utility for GO) | http://amigo.geneontology.org/ | Carbon et al., 2009 |
| Gene Ontology (GO) | http://www.geneontology.org/ | Ashburner et al., 2000 |
| HUGO Gene Nomenclature Committee | http://www.genenames.org/ | Seal et al., 2011 |
| NCBI gene index database | http://www.ncbi.nlm.nih.gov/gene/ | |
| GeneCards | http://www.genecards.org | Reban et al., 1997 |
| Human diseases | Online Mendelian Inheritance in Man (OMIM®) database | http://www.ncbi.nlm.nih.gov/omim | |
| Drugs | DrugBank database | http://www.drugbank.ca | Wishart et al., 2006, 2008 |
| Thomson Reuters Integrity database | https://integrity.thomson-pharma.com | Nonfree |
| Software | R software | http://CRAN.R-project.org/ | R Development Core Team, 2008 |

All resources are publicly available, most of them free of charge.

FIGURE 1 | Scheme of the data analysis workflow. The analyses had two major aims, i.e., (i) assessing the biological functions of genes reportedly associated with insensitivity to pain (upper line) and (ii) using the biological processes in which these genes are involved to find repurposing candidates among DrugBank listed drugs (lower line). To this end, genes were identified in databases and their biological functions were associated based on the annotations in the Gene Ontology database; for the drug target coding genes (>4,000) with a filter for too many irrelevant terms implemented as an overrepresentation analysis. Such filter was not necessary for the only 20 genes associated with insensitivity to pain. However, for the latter, overrepresentation analysis (ORA) followed by functional abstraction was performed to obtain a comprehensible set of >10 biological functions which summarize the biological roles of these genes in an interpretable manner. The information obtained in the ORA of the 20 pain insensitivity genes was used to generate a virtual “PainInsensitivity drug” that was introduced into the “drug vs. biological processes” matrix of all drugs. Subsequently, unsupervised machine learning was used to find data structures among all drugs. Those drugs that, in the high dimensional vector space of associations with GO terms (biological processes), laid near the virtual “PainInsensitivity drug” were the repurposing candidates.

of “functional abstraction” (Ultsch and Lötsch, 2014). This aims to reduce the numbers of GO terms using a heuristic search algorithm that identifies so-called “functional areas” (Ultsch and Lötsch, 2014), which are GO terms that qualify by their informational importance as headlines representing specific aspects (taxonomies) of the complete DAG with maximal coverage, precision, informational value and conciseness (Ultsch and Lötsch, 2014).

Assessment of Drug Repurposing Based on Computational Analysis

Following analysis of the biological processes in which the genes associated with human hereditary insensitivity to pain, the possibility was explored whether the discovered knowledge could be employed for a drug repurposing approach that uses functional similarity between drugs and key functions of a trait-relevant gene set (Lötsch and Ultsch, 2016a,b). This was based on (i) the biological processes identified as characterizing the genes associated with insensitivity to pain obtained in above analysis, and required in addition (ii) associating drugs, via the genes coding for their targets, with biological processes followed by (iii) the identification of those drugs that are associated with similar biological processes that characterize the genes associated with insensitivity to pain phenotypes. These analytical steps will be described in as follows.

As described previously (Lötsch and Ultsch, 2016a,b), the association of drugs with biological processes was obtained by combining drugs which are associated with molecular targets available from the DrugBank database with genes associated with biological processes as queried from the Gene Ontology database (Table 3). Specifically, a set of n = 2,215 genes queried as drug targets from the DrugBank database was submitted to overrepresentation analysis to identify relevant biological processes, expressed as GO terms, which can be addressed
by the available drugs. ORA, performed as described above, however, with the parameters $t_\alpha = 1 \times 10^{-15}$ and $\alpha$ correction according to Bonferroni (1936), provided 830 GO terms that can be considered as specifically describing the biological processes in which 2,215 targets of the drugs known to the DrugBank database are involved. This resulted in a $2,215 \times 830$ “gene vs. biological process” matrix. Here, in contrast to its use for functional interpretation of a gene set applied onto the pain insensitivity genes as described above, the ORA served merely as a filter for relevant GO terms, which was set at a conservative $p$-value threshold to eliminate too many generic terms from the matrix; as a functional interpretation of the biological roles of all drug targets was not intended, functional abstraction was not applied. From the matrix product of the “drug vs. genes” matrix and the “genes vs. processes” matrix, a $4,834 \times 380$ “drug vs. biological process” matrix was provided.

Subsequently, a virtual drug “PainInsensitivity” was added to the “drug vs. biological process” matrix. This “drug” carried a single vector composed of the numbers that indicate how often particular biological processes have been associated with a member of the $n = 20$ sized gene set causally involved in hereditary insensitivity to pain. Using the biological processes that were addressed by both, DrugBank queried drugs and the “PainInsensitivity” drug, a $4,512 \times 380$ “drug vs. biological process” matrix was obtained. Within this matrix, the Euclidian distances of each of the 4,511 DrugBank annotated drugs to the “PainInsensitivity” drug were calculated.

To eliminate drugs dissimilar to the “PainInsensitivity” drug, the reciprocals of the Euclidian distances were submitted to a machine learning (ULutsch and Lötsch, 2014) was used. The neural network consisted of a two-dimensional toroid grid (ULutsch, 2003) of $50 \times 80$ neurons ($n = 4,000$ units). Each neuron holds, in addition to the input vector from the high-dimensional space of processes associated to each drug, a further vector carrying “weights,” which were initially randomly drawn from the range of the data variables and subsequently adapted to the data during the learning phase that used 50 epochs.

The trained emergent self-organizing map (ESOM) represents the drugs on a two-dimensional toroid map as the localizations of the “best matching units” (BMU). On top of this grid the distance structures in the high-dimensional feature space of biological processes is depicted in form of a so-called U-Matrix (ULutsch and Sieman, 1990; Lötsch and ULutsch, 2014). The machine learning was performed using the R library “UMatrix” (M. Thrun et al., Marburg, Germany, publicly available at http://www.uni-marburg.de/fb12/datenbionik/softwareweb/packages/ ABCanalysis/index.html). Only the cluster was further regarded that included the “PainInsensitivity” drug. For the DrugBank annotated members of this cluster, evidence was queried from the literature supporting, or discouraging, a possible involvement in pain.

RESULTS

Computational Analysis of the Genes Involved in Hereditary Insensitivity to Pain

To identify the systems biology of hereditary insensitivity to pain, a set of $n = 20$ unique genes (Table 1) for which published evidence supported an association with the human phenotype of inherited insensitivity to pain was queried form publicly available databases (Table 3). The biological functions associated with the expression of these genes and their respective products were queried from the Gene Ontology knowledge base (GO; http://www.geneontology.org; Ashburner et al., 2000). Over-representation analysis (ORA) against all human genes ($n = 18,750$ in the used GO version) and using an FDR corrected $p$-value threshold of $t_\alpha = 0.05$ resulted in 136 significant GO terms (Supplementary Figure 1).

Subsequent functional abstraction, which is a method developed to reduce a large set of GO terms to a comprehensible small subset of “headline terms” or “functional areas” that represent specific aspects (taxonomies) of the complete polyhierarchy with maximal coverage, precision, informational value and conciseness (ULutsch and Lötsch, 2014), identified 8 GO terms qualifying as headlines to summarize the biological functions that are particularly important addressed by the 20 genes associated with insensitivity to pain, among all human genes (Table 4). A GO annotation with pain was found in two particular taxonomies of the polyhierarchy, i.e., under the headline “multicellular organismal response to stress” (GO:0033555) that exclusively included response to pain (GO:0048265) and its descendants, and again under the headline “Neurological system process” (GO:0050877), which at the chosen $p$-value threshold ended downstream with “sensory perception of pain” (GO:0019233) and “neuronal action potential” (GO:0019228).
A second larger group of biological processes typically annotated with the genes associated with insensitivity to pain comprised developmental and structural aspects of the nervous system, i.e., "nervous system development" (GO:0007097) and "neuron death" (GO:0070997) including "negative regulation of apoptotic process" (GO:0043524) as a downstream terminal of the taxonomy, and also processes of "cellular component organization" (GO:0016043). A third larger group of biological processes typically annotated with the genes associated with insensitivity to pain comprised metabolic processes ("organic substance metabolic process," GO:0071704), which included biosynthetic processes related to sphingolipids ("sphingolipid biosynthetic process," GO:0046512) and ceramides ("ceramide biosynthetic process," GO:0046513), "nerve growth factor processing" (GO:0032455), "positive regulation of lipophagy" (GO:1904504), and "positive regulation of histone H3-K9 dimethylation" (GO:1900111). The remaining processes could only be summarized in large sets of more heterogeneous biological functions such as "biological regulation."

Candidate Drugs for Repurposing

To explore whether the results of the computational analysis of genes associated with human insensitivity to pain could be employed for drug repurposing, a virtual drug "PainInsensitivity" was created that carried a single vector composed of the numbers of how often particular biological processes have been associated with a member of the \( n = 20 \) genes causally involved in hereditary insensitivity to pain. Only the \( n = 38 \) biological processes were included that were also associated with any drug queried from the DrugBank database. Similarly, vectors coding for the associations with biological processes were assigned to each drug. The data space was reduced by eliminating drugs at large Euclidian distances from the "PainInsensitivity" drug, using nested ABC analysis.

Following projection of the resulting \( 414 \times 38 \) sized "drug vs. biological process" data space onto a toroid grid of \( 50 \times 80 \) neurons and training of a self-organizing map, a U-matrix visualization was displayed on top of this SOM (Figure 2). Large U-heights in this visualization indicated a large gap in the data space whereas low U-heights indicated that the points are close to each other in the data space. The distance-based data structure indicated a cluster that comprised the virtual drug "PainInsensitivity" (red dot in Figure 2) together with further \( n = 22 \) DrugBank listed substances (Table 5). These substances can be considered as repurposing candidates for pain, based on their vicinity (yellow dots in Figure 2) in the high-dimensional data space to the biological processes associated with the genes causally involved in insensitivity to pain as represented on the SOM by the "PainInsensitivity" drug.

**DISCUSSION**

The present analysis used empirical evidence for human genes that when nonfunctional may cause the phenotype of insensitivity to pain. The set of \( n = 20 \) genes included those causally involved in hereditary sensory and autonomic neuropathies and for long in the center of pain research, and in addition further genes involved in several different neurological syndromes or, such as SCN9A, genes that when nonfunctional are mainly associated with insensitivity to pain. Indeed, the only additional phenotype in subjects carrying loss-of-function mutations in SCN9A was anosmia (Cox et al., 2010). In the present analysis, the common phenotype of insensitivity to pain associated with the complete gene set was used to address the functional background of pain insensitivity and to explore a computational attempt at using the information for drug repurposing. The computational approach allowed using data on the biological functions of genes acquired in any context, without restriction to pain research (Lötsch et al., 2013).

The computational analysis of genes associated with human insensitivity to pain included, in addition to basic functions as neuronal signaling or the perception of pain, as major components (i) nervous system development and structural component assembly and (ii) lipid-mediator based signaling including sphingolipids and ceramides. Nervous system development, while presently probably also associated with the hereditary developmental disorders of the nervous system, appear to be a consistent part of the genetic background of pain since they also emerged as particularly prominent systemic functions of larger sets of genes related to all aspects of pain (Ultsch et al., 2016). As discussed there, as possible explanation of the consistent appearance of nervous system development among key biological processes exerted by pain-relevant sets of genes involves the concept of chronic pain as a dysregulation in biological processes that describe its systemic features of learning and neuronal plasticity (Mansour et al., 2014). Thus, the structural aspect of nervous system development would be compatible with both, hereditary developmental neuropathies and pain.

A role of lipid signaling is compatible with prior knowledge of the pathophysiology of pain. Indeed, the ceramide-to-sphingosine pathway has been proposed already as a therapeutic target in pain (Salvemini et al., 2013). Similarly,
FIGURE 2 | 3D-view of the U-matrix visualization of distance based structures of the 414 × 38 sized “drug vs. biological process” matrix, comprising the 413 drugs annotated with one or more of the n = 38 biological processes assigned to both, the set of 20 genes causally implicated in insensitivity to pain and the drug targets queried from the DrugBank database, which following ABC analysis based item selection were found in at closer Euclidian distances form the virtual “PainInsensitivity” drug (red dot) that carried all of the n = 38 processes. The U-matrix has been obtained using a projection of the data points onto a toroid grid of 4,000 neurons where opposite edges are connected. The U-Matrix was colored as a geographical map with brown (up to snow-covered) heights and green valleys with blue lakes. Valleys indicate clusters and watersheds indicate borderlines between different clusters. The dots indicate the so-called “best matching units” (BMUs) of the self-organizing map (SOM), which are those neurons whose weight vector is most similar to the input. A single neuron can be the BMU for more than one data point; hence, the number of BMUs may not be equal to the number of drugs. In the vicinity of the red dot, i.e., the virtual “PainInsensitivity” drug, a mount ridge surrounded valley was observed that represented a cluster of drugs (yellow dots) most similar to the virtual “PainInsensitivity” drug. However, the latter was located eccentrically in the cluster indicating that the similarity to any of the DrugBank queried repurposing candidates for pain therapy is incomplete. The other drugs (gray dots) lay outside the cluster of the “PainInsensitivity” drug and could therefore be rejected as repurposing candidates based on the present approach to search for drugs that with respect to the addressed biological processes are most similar to the pattern of biological processes in which the genes associated with insensitivity to pain are involved. The figure has been created using the R software package (version 3.3.3 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008) using our R library “Umatrix” (M. Thrun, F. Lerch, Marburg, Germany, http://www.uni-marburg.de/fb12/datenbionik/software, file http://www.uni-marburg.de/fb12/datenbionik/Umatrix.tar.gz).

sphingosine-1-phosphate induced nociceptor excitation and ongoing pain behavior in mice and humans (Camprubi-Robles et al., 2013) and therefore, sphingosine-1-phosphate receptors have been proposed as novel targets for the treatment of pain (Welch et al., 2012). While these findings have emerged from basic research, the present prominent role of this pathway highlights the suitability of the computational approach analyzing available information about a particular subset of pain-related genes. Again, the biological plausibility applies to both, developmental neuropathies and pain.

The idea behind using the knowledge about the biological processes characterizing insensitivity to pain for drug repurposing is to use the similarity measure in the high-dimensional vector space of the drug’s interactions with biological processes for the identification of substances qualifying for the treatment of traits defined on the basis of biological processes (Lötsch and Ultsch, 2016b). A disease-relevant gene set is functionally analyzed and compared with the biological processes in which the targets of available drugs are involved. This approach presently resulted in the selection of a subset of n = 22 substances, which is the size of the cluster that surrounded the functional information of the present trait of interest in the high-dimensional data space (Figure 2). Results obtained with the present computational approach require basic research verification. However, the set included substances for which such evidence could be found. For example, myristic acid is contained at 18.64% in coconut oil, which when orally administered to rats produced moderate anti-inflammatory and anti-nociceptive effects (Intahphuak et al., 2010). Imatinib, a tyrosine kinase inhibitor, had no antinociceptive effects in a nerve injury model in rats when administered alone; however, when combined with a previously ineffective dose of morphine, complete pain relief was obtained (Donica et al., 2014), which was attributed to its platelet-derived growth factor inhibiting actions (McGary et al., 2004).

Similarly, flavopiridol or alvocidib, an inhibitor of cyclin-dependent kinases, was shown to facilitate the recovery from tactile allodynia when in a rat nerve injury model, which was attributed to its Janus kinase pathway inhibiting actions (Tsuda et al., 2011). Ellagic acid, a polyphenolic compound from plants such as raspberries, eucalyptus, and nuts (Clifford and Scalbert, 2000), showed antinociceptive effects and potentiated...
the effects of morphine in different rat models of pain (Mansouri et al., 2014). Furthermore, staurosporine, an alkaloid initially isolated from the bacterium *Streptomyces staurosporces* (Omura et al., 1977) and found to inhibit protein kinases, inhibited Angiotensin II mediated sensitization of post mortem analyzed human nerves (Anand et al., 2015), which is in line with the targeting of angiotensin 2 type II receptors as a novel treatment for neuropathic pain (Rice et al., 2014).

However, the list of candidate drugs for repurposing also included the multitargeted receptor tyrosine kinase inhibitors sorafenib and sunitinib, for which evidence suggests hyperalgesic actions. This emphasizes that, as discussed previously (Lötsch and Ultsch, 2016b), a limitation of the present implementation of this computational approach to drug repurposing consist of the unsigned inclusion of drug vs. target interactions, i.e., without distinction of agonistic from antagonistic actions. Therefore, topical experts’ knowledge is required to amend this limitation. Of further note, the present computational approach at drug repurposing based on computational analysis of gene functions and similarity measures in the high-dimensional data space still depends on the accuracy and completeness of the information in the queried databases. This makes it vulnerable to research bias and erroneous information in the databases, however, the increasing trend toward “big data” supports the expectation of a continuously broadening knowledge base.

The present computational analysis of the genes involved in human hereditary insensitivity to pain in a drug repurposing context extends previous applications of the recently proposed concept of “process pharmacology” (Lötsch and Ultsch, 2016b) by (i) the inclusion of a topically selected set of genes in addition to a previously used set of genes derived from microarray analysis, and (ii) the introduction of the concept of a “virtual drug,” i.e., a vector of biological processes representing the key functions of a gene set that can be entered as a further drug into the drug vs. biological process matrix, thereby facilitating the search for repurposing candidates by using similarity measures in the high dimensional vector space. As described previously (Lötsch and Ultsch, 2016a,b), in the present approach diseases are regarded as resulting from the activity of pathophysiological processes captured in the GO database via the category “biological processes,” i.e., series of events or molecular functions with a defined beginning and end (Ashburner et al., 2000). Such a focus on disease relevant biological processes (Lötsch and Ultsch, 2016b) has been shown to provide a robust basis for drug classification agreeing with classical approaches (Lötsch and Ultsch, 2016a) and may provide a phenotypic approach to drug discovery and repurposing.

The present ORAs were performed with different *p*-value thresholds. The conservative *p*-value threshold of *tp* = 1·10−15 used for the analysis of the drug targets was heuristically chosen to accommodate the intention to only include highly relevant terms and to obtain a set size of 500–1,000 terms that had proven suitable in previous similar analyses (Lötsch and Ultsch, 2016a,b). By contrast, the *p*-value threshold of 0.05 chosen for the ORA of the *n* = 20 genes associated with human insensitivity to pain was chosen for practical reasons. That is, it was the strictest criterion providing the generally accepted significance level of *p* = 0.05 and a correction for multiple testing, a stricter criterion or the use of more conservative α-correction according to Bonferroni (1936) led to a nearly empty ORA result for the present set of *n* = 20 genes involved in human insensitivity to pain and could therefore not applied. However, a systematic test of the optimum *p*-value threshold for the present computational drug repurposing approach remains an active research topic to be addressed in future work.

By placing the modulation of biological processes into the focus with molecular targets merely serving as intermediates, the present concept may exceed the particular molecular mechanism by which a drug acts making it particularly suitable for drug repurposing by eliminating the restriction to a specific molecular target. In this respect, the present concept

### Table 5

| Drug name                  | DrugBank ID | Evidence supporting involvement in pain                      |
|----------------------------|-------------|---------------------------------------------------------------|
| “PainInsensitivity” drug   | none        | Not applicable (virtual drug)                                 |
| Acetylsalicylic acid       | 945         | Approved analgesic                                            |
| Myristic acid              | 8231        | Antinociceptive effects (Intahphuak et al., 2010)             |
| Phosphonoserine            | 4522        | –                                                             |
| Tamoxifen                  | 675         | –                                                             |
| Adenosine monophosphate    | 131         | Its inhibition attenuated pain (Liu et al., 2007)             |
| Phosphonothreonyne         | 2482        | –                                                             |
| Imatinib                   | 619         | Restored morphine analgesic potency (Donica et al., 2014)     |
| Sorafenib                  | 398         | Might induce pain (Di Cesare Mannelli et al., 2015)           |
| Flavopiridol               | 3496        | Recovery from tactile allodynia (Tsuada et al., 2011)         |
| Nintedanib                 | 9079        | –                                                             |
| Ellagic acid               | 8846        | Antinociceptive effects (Mansouri et al., 2014)               |
| Dasatinib                  | 1254        | –                                                             |
| Sunitinib                  | 1268        | Hyperalgesic effects (Bullon et al., 2016)                    |
| Staurosporine              | 2010        | Counteracting capsaicin sensitization (Anand et al., 2015)    |
| Lenvatinib                 | 9078        | –                                                             |
| Pazopanib                  | 6589        | –                                                             |
| MP470                      | 5216        | –                                                             |
| ABT-869                    | 6080        | –                                                             |
| XL999                      | 5014        | –                                                             |
| Yohimbine                  | 1392        | Antinociceptive effects (Shannon and Lutz, 2000)              |
| Ponatinib                  | 8901        | –                                                             |
| Caffeine                   | 201         | Analgesic effects shown and discussed (Baratloo et al., 2016) |

The 22 drugs are the members of the cluster in the high dimensional space to which the virtual “PainInsensitivity” drug belonged (Figure 2).
complements non-redundantly alternative implementations of computational science in drug research and development. For example, the analysis of molecular interaction networks (e.g., protein-protein interaction networks or gene-gene co-expression networks) may be employed to study the possible consequences of target removals, considering a target as effective when its removal modifies the network in an essential way (Penrod et al., 2011). Other approaches proposed to use supervised machine learning, i.e., support vector machines, to analyze druggable protein-protein interaction on the basis of the number of shared GO terms indicating similarities in biological function between two interacting proteins (Sugaya et al., 2007; Sugaya and Ikeda, 2009). Further approaches include pathway based analysis where pathway annotations, such as provided by the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database are “translated into a two-dimensional statistical test problem that involves Wilcoxon’s signed rank sum test in order to compute a Z-score for each pathway that quantifies the degree of alteration across the different experimental conditions” (Herwig and Lehrach, 2006). These approaches have also shown to provide clinically useful approaches to drug development such as the identification of anticancer drug combinations (Azmi et al., 2010), which complements a similar success of the present method (Lötsch and Ultsch, 2016b). This emphasizes the increasing utility of a variety of computational approaches to gene functions in drug research, for which the present analysis adds further support.

Most contemporary machine learning techniques for the analysis of genomic data are of the supervised learning type (for a recent review, see Libbrecht and Noble, 2015). These methods aim at the diagnosis (classification) of cases. By contrast, in the present work unsupervised methods were used, which mainly aim at structure detection in high dimensional (genetic) data. Typical methods for structure detection in high dimensional data are (i) nonlinear projections such as multidimensional scaling (MDS) (Tzeng et al., 2008), t-SNE (van der Maaten and Hinton, 2008; Bushati et al., 2011) or curvilinear component analysis (Alanis-Lobato et al., 2015) and (ii) clustering methods summarized in (Pirim et al., 2012). The ESOM/Umatrix method used here can be regarded as a combination of a disentangling and neighborhood preserving projection method combined with a clustering algorithm. This method has been demonstrated to be superior to other approaches in the sense, that complex and intertwined clusters can be identified, however, no spurious structures are artificially introduced by the clustering method itself (Ultsch and Lötsch, 2017) for which a superiority to many other projection methods has been shown (Tasdemir and Merényi, 2012).

**CONCLUSIONS**

Genes causally involved in human insensitivity to pain provide a unique source of studying the pathophysiology of pain and the development of novel analgesic drugs. In keeping with the contemporary trend toward "big data" analysis in biomedical research, an integrated computational analysis was performed to study the set of genes for emergent, principal pathophysiological processes that characterize insensitivity to pain. As a result, a particular importance for pain perception was observed for processes related to structural changes in the nervous system and to ceramide and sphingosine signaling pathways, which is compatible with suggestions of using these pathways as therapeutic target in pain. Using the biological processes characterizing hereditary insensitivity to pain for drug repurposing, a clear cluster of \( n = 22 \) substances emerged that comprised several drugs for which implications in pain have been shown occasionally in preclinical experiments. Thus, the present concept provides biologically plausible results and seems to be suitable for drug discovery by identifying a narrow choice of repurposing candidates, demonstrating that contemporary machine-learned methods offer innovative approaches to knowledge discovery from previous evidence.

**AUTHOR CONTRIBUTIONS**

Wrote the paper, conception or design of the work: JL. Analyzed the data: JL and CL. Data collection: DK and JL. Provided advice for data analysis: AU.

**FUNDING**

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

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnmol.2017.00252/full#supplementary-material

**Supplementary Figure 1** | This figure shows the top-down representation of the GO terms representing the biological processes in which the set of \( n = 20 \) genes causally involved in hereditary absence of pain perception (Table 1) is involved (for an enlarged version, see the Supplementary Figure). The graphical representation follows the standard of the GO knowledgebase, where GO terms are related to each other by “is-a,” “part-of,” and “regulates” relationships forming a polyhierarchy organized in a directed acyclic graph (DAG, Thulasiraman and Swamy, 1992). The figure represents the results of an over-representation analysis with parameters for the \( p \)-value threshold, \( t_p = 0.05 \) and FDR \( \alpha \)-correction. (Top) Significant terms are shown as colored circles with the number of member genes, the number of expected genes by chance and the significance of the deviation in the observed from the expected number of genes indicated (yellow = headline, red = significant term located in the polyhierarchy below a functional area). Blue vertices or blue labels, are the most specific terms (leaves of the DAG) at the end of a taxonomy in the polyhierarchy. The biological processes in which these genes are involved can be summarized by seven primary “functional areas” representing the most remarkable nodes with respect to their localization in the polyhierarchy. (Bottom) Zoomed parts of particular functional areas, recreated in a slightly different arrangement to enhance visibility.
A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity. *Pain* 157, 2747–2757. doi: 10.1097/j.pain.0000000000000944

Functional abstraction as a method to discover knowledge in gene ontologies. *PLoS ONE* 9:e90191. doi: 10.1371/journal.pone.0090191

Computed ABC analysis for rational selection of most informative variables in multivariate data. *PLoS ONE* 10:e0129767. doi: 10.1371/journal.pone.0129767

Machine-learned cluster identification in high-dimensional data. *J. Biomed. Inform.* 66, 95–104. doi: 10.1016/j.jbi.2016.12.011

Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Mutations in the small GTP-ase late endosomal protein RAB7 cause Charcot-Marie-Tooth type 2B neuropathy. *Am. J. Hum. Genet.* 72, 722–727. doi: 10.1086/367847

Sphingosine-1-phosphate receptors as emerging targets for treatment of pain. *Biochem. Pharmacol.* 84, 1551–1562. doi: 10.1016/j.bcp.2012.08.010

DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36, D901–D906. doi: 10.1093/nar/gkm958

The phenotype of congenital insensitivity to pain due to the NaV1.9 variant p.L811P. *Eur. J. Hum. Genet.* 23, 561–563. doi: 10.1038/ejhg.2014.166

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.