Knowledge graph analytics platform combining LINCS and IDG for Parkinson's disease drug target illumination

Jeremy Yang124, Chris Gessner23, Joel Duerksen3, Daniel Biber4, Jessica Binder5, Murat Ozturk2, Brian Foot5, Robin McIntire6, Kyle Stirling2, Ying Ding5, and David Wild12

1School of Informatics, Computing and Engineering, Indiana University, Bloomington, IN, USA; 2Data2Discovery, Inc., Bloomington, IN, USA; 3Katana Graph, Austin, TX, USA; 4School of Medicine, Dept. of Internal Medicine, Translational Informatics Division, University of New Mexico, Albuquerque, NM, USA; 5School of Information, Dell Medical School, University of Texas, Austin, TX, USA

Combining LINCS + IDG for a powerful new approach to identify novel drug targets for complex diseases

LINCS, “Library of Integrated Network-based Cellular Signatures”, and IDG, “Illuminating the Druggable Genome”, are both NIH projects and consortia that have generated rich datasets for the study of the molecular basis of human health and disease. LINCS L1000 expression signatures provide unbiased systems/omics experimental evidence. IDG provides compiled and curated knowledge for Illumination and prioritization of novel drug target hypotheses. Together, these resources can support a powerful new approach to identifying novel drug targets for complex diseases, such as Parkinson’s disease (PD), which continues to inflict severe harm on human health, and resist traditional research approaches.

**KGAP algorithm (for target illumination)**
1. Select disease query as drug indication defining drugset.
2. (Optional) Filter drugset by ATC for biomedical coherence (prefer disease-modifying over symptom-relieving).
3. Query KG for genes associated with drug perturbagens, scoring and ranking by graph analytics for evidence quantification.
4. Validate ranking against MoA targets.

**Validation results for Parkinson’s Disease (PD)**

These results show enrichment between the aggregated expression profiles from LINCS and the independent knowledge sources of the literature corpus. This agreement supports the inference that since our method can independently “re-discover” known targets, it is capable of discovering new targets.

**Gene illumination case study: Synaptogyrin-3 (SYNGR3)**

PD is complex in its polygenic etiology, and in its clinical presentation and progression. IDG resource TIN-X (Target Importance and Novelty Explorer) was used to rank and filter KGAP results for PD. TIN-X defines “importance” and “novelty” from mentions and co-mentions of diseases and genes in the PubMed literature. One high ranking but understudied gene, Synaptogyrin-3 (SYNGR3), was manually investigated as a plausible new drug target, based on published experimental and theoretical links with statin drugs.

**Opportunities, next steps, future directions**
Future directions include production quality knowledge graph, a completed LIDIA UI, KGAP-based research in Parkinson's disease and other selected disease areas. This work was inspired by the Common Fund Data Ecosystem (CFDE), designed to foster, facilitate, and further empower integrative data analysis such as this. We are eager to collaborate with experimentalists and medical scientists to evaluate results and extend this work. Plans for community development and engagement include involving data science students from IU and UNM, and facilitating open science through an open platform for development and application science. In addition, complementary research efforts are in progress at IU, UNM and D2D with relevance and potential synergies with this project.

**For more information**
For more information, refer to our paper and GitHub repositories. LINCS and IDG are large international consortia, both from the NIH Common Fund, with diverse contributors and extensive public resources to offer.

**Conclusions**
Combining LINCS and IDG via KGAP has empowered the investigation of the molecular basis of complex diseases, and specifically identification and prioritization of novel drug targets. The generality of the approach indicates that KGAP is applicable to many disease areas, in addition to Parkinson’s Disease. The KGAP team is comprised of scientists and engineers with a range of skills and knowledge who combined to make this project possible.

**IDG DrugCentraldrugsets for KGAP disease queries**
Approved drugs represent strong knowledge to high regulatory standards (FDA, EMA, PMDA) for specific indications related to disease conditions. DrugCentral provides: (1) Indications for all approved drugs, defining disease specific drugsets for KGAP queries, (2) Mechanism-of-Action (MoA) target associations, for validation of KGAP results.

**KGAP Neo4j Knowledge Graph**

**DrugCentral**

**DrugCentral**

**KGAP Prototype Webapp**

**KGAP Team**
1. David Wild (PI), Professor, IU; President, D2D
2. Daniel Biber, Data Scientist, D2D
3. Jessica Binder, Post-doctoral fellow, UNM
4. Ying Ding, Professor, U'Texas; co-Founder, D2D
5. Joel Duerksen, Sr Developer, D2D
6. Brian Poole, Sr Developer, D2D
7. Christopher Gessner, PhD student, IU
8. Robin McIntire, EVP, D2D
9. Murat Ozturk, Developer, D2D
10. Kyle Stirling, Lecturer, IU; VP Dev, D2D
11. Jeremy Yang, Research Scientist, UNM & D2D