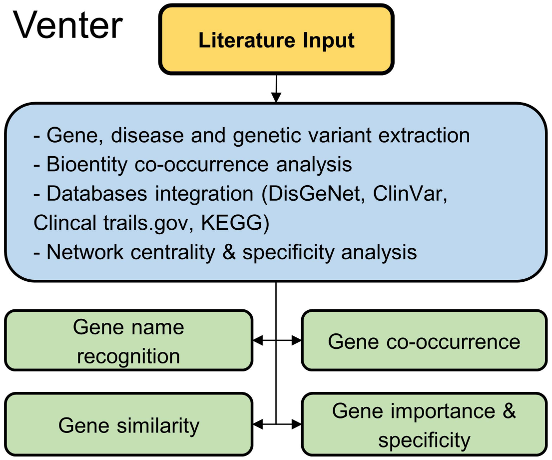
ARC LP - Transfer Learning for Genome Analysis and Personalised Recommendation

Dist Prof Jie Lu, A/Prof Guangquan Zhang, Dr Yi Zhang, Dr Qian Zhang, Mr Mengjia Wu

Dr Hua Lin, Mr Mark Grosser, Mr Steven Tipper, Prof Deon Venter

Objective 1: Analysing genome and gene expression (GGE) data associations through uncertainty learning

Our team has developed BiblioGene Miner software to extract gene entities from free scientific text, integrate gene similarity information from biomedical databases, and analyse gene importance and specificity via network analytics. Those analysing results will act as gene features to facilitate heterogeneous bioentity graph representation.



Literature Input

- Gene, disease and genetic variant extraction
- Bioentity co-occurrence analysis
- Databases integration (DisGeNet, ClinVar, Clincal trails.gov, KEGG)
- Network centrality & specificity analysis

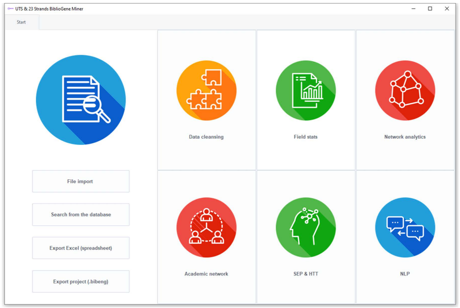
Gene name recognition

Gene co-occurrence

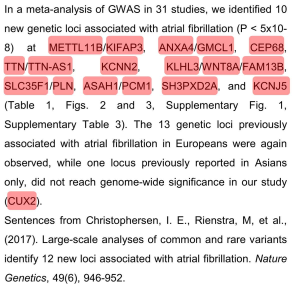
Gene similarity

Gene importance & specificity

BiblioGene Miner architecture



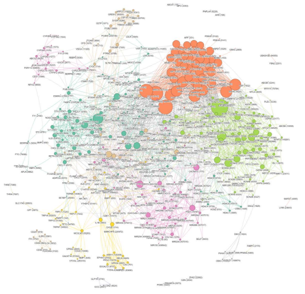
BiblioGene Miner main interface



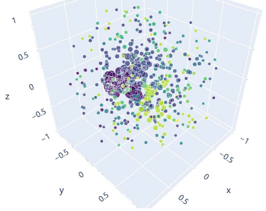
In a meta-analysis of GWAS in 31 studies, we identified 10 new genetic loci associated with atrial fibrillation (P < 5x10-8) at METTL11B/KIFAP3, ANXA4/GMCL1, CEP68, TTN/TTN-AS1, KCNN2, KLHL3/WNT8A/FAM13B, SLC35F1/PLN, ASAH1/PCM1, SH3PXD2A, and KCNJ5 (Table 1, Figs. 2 and 3, Supplementary Fig. 1, Supplementary Table 3). The 13 genetic loci previously associated with atrial fibrillation in Europeans were again observed, while one locus previously reported in Asians only, did not reach genome-wide significance in our study (CUX2).

Sentences from Christophersen, I. E., Rienstra, M, et al., (2017). Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nature Genetics*, 49(6), 946-952.

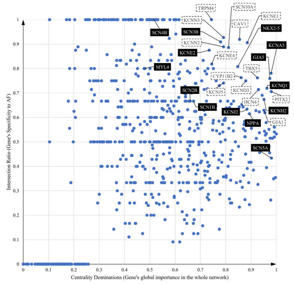
Gene recognition and textual features



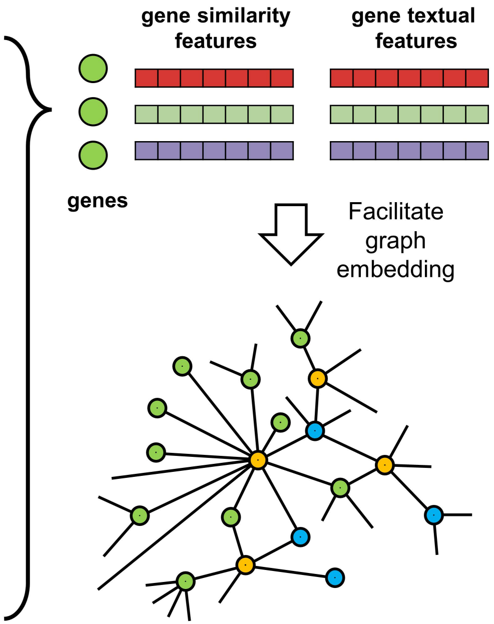
Gene co-occurrence map



Gene similarity feature space



Gene importance & specificity



Facilitate graph embedding