Chemical Phenomics Initiative to Drive Therapeutic Target Discovery

Charles H. Williams & Charles C. Hong

Division of Cardiovascular Medicine, Departments of Biochemistry & Molecular Biology, and Physiology, University of Maryland School of Medicine, Baltimore Maryland

Abstract

Chemical genetics involves the discovery, development and use of chemical probes for interrogation of biological processes and for translational discoveries. In a manner analogous to classic forward mutagenesis screens, the Hong lab has conducted an unbiased, high-throughput chemical screen for small molecules that specifically modulate early embryonic development in zebrafish, and has carried out the follow-up task of identifying the pharmacological targets of a number of developmental modulators. Using this target-agnostic, “high content” phenotypic screening platform, we have discovered novel BMP, Wnt and hedgehog inhibitors, as well as first-in-class modulators of cell signaling components. Moreover, since disturbances in developmental pathways play central role in the pathogenesis of many human illnesses, small molecules that selectively target them have significant translational potential. Yet the complexity and unbiased nature of phenotypic screens make the crucial follow-up task of identifying the biologically relevant target of each hit very challenging. Leveraging the available molecular genetic information on early zebrafish embryogenesis, we have developed ZePASS (Zebrafish Phenotypic Anatomical Similarity System), an unbiased deep learning method to map the actions of small molecules and accelerate target identification. Finally, we are leveraging UK Biobank phenotype-genotype database to identify the clinical phenotypes associated with naturally occurring human genetic variations in the target genes, and then utilize these associations to ultimately guide therapeutic development for important unmet clinical needs. Finally, digitally annotated results of chemical screens and target deconvolution will be made available to the broader scientific community via a searchable, online database named Chemical Phenomics Initiative.

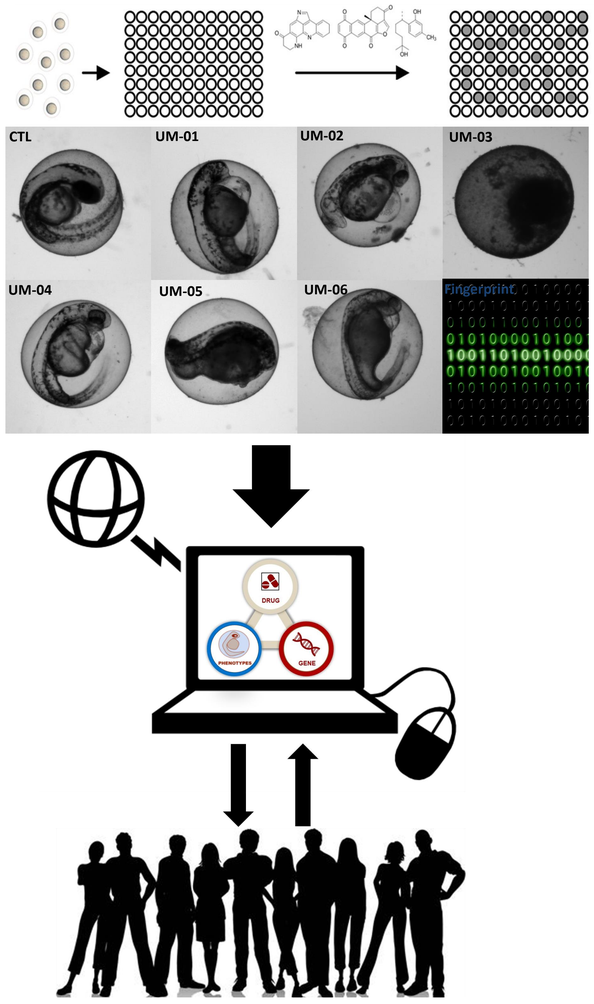


Fig 1. Zebrafish embryos were exposed to a diverse chemical library at gastrulation and allowed to develop. UM01-06 display a variety of phenotypes. Phenotypes are annotated, given a phenotype fingerprint, and deposited into a web accessible database. Once targets are identified this is linked to therapeutic predictions powered by human genetic data from UKBiobank. This web portal will be a launching point for tool and therapeutic development.

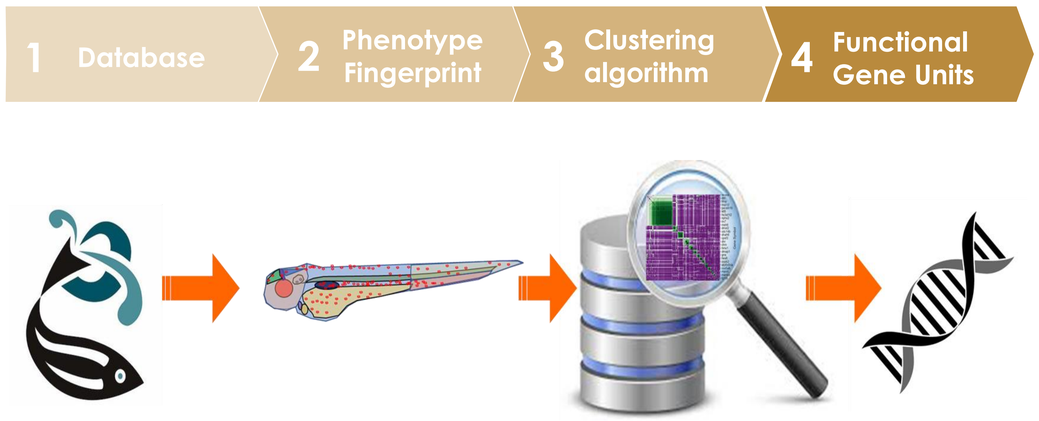
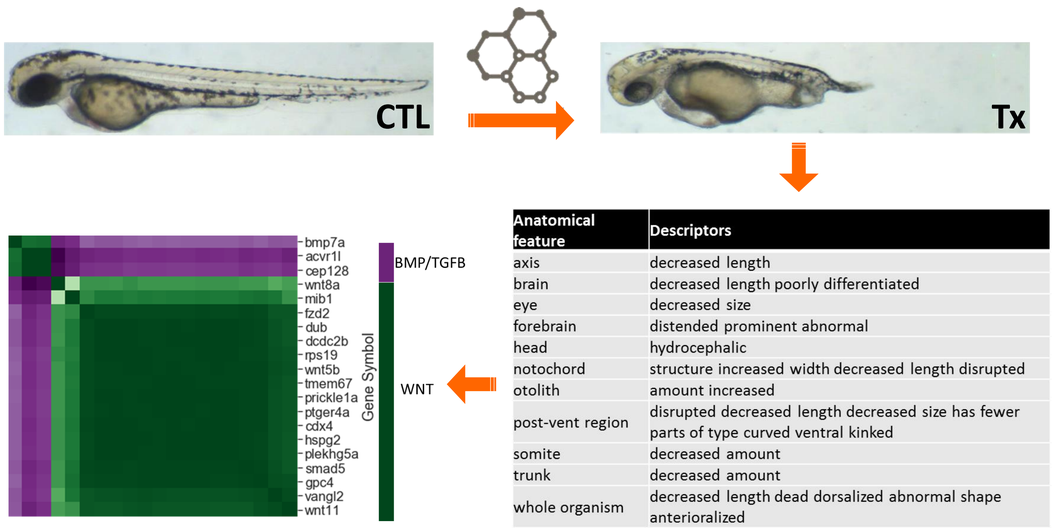


Fig 2. Gene (mutant/morphant) – phenotype data from ZFIN was segmented and 29 anatomical features were selected to phenotype between fertilization and 48 hours post fertilization. The resulting 2909 gene associated phenotype were converted from text descriptors to digital fingerprints. These fingerprints can be clustered and queried to reveal functional genetic units.



Anatomical feature	Descriptors
axis	decreased length
brain	decreased length poorly differentiated
eye	decreased size
forebrain	distended prominent abnormal
head	hydrocephalic
notochord	structure increased width decreased length disrupted
otolith	amount increased
post-vent region	disrupted decreased length decreased size has fewer parts of type curved ventral kinked
somite	decreased amount
trunk	decreased amount
whole organism	decreased length dead dorsalized abnormal shape anteriorialized

Fig 3. Identified hits have been subjected to a deep phenotypic analysis of the 29 anatomical features. Pharamacophores. Example: Qualitative text descriptors from compound treated embryo. This class of compounds perturbed 11 different anatomical features. Similarity query of the fingerprint database generated from ZFIN identified potential activity in Wnt and potentially TGFB.

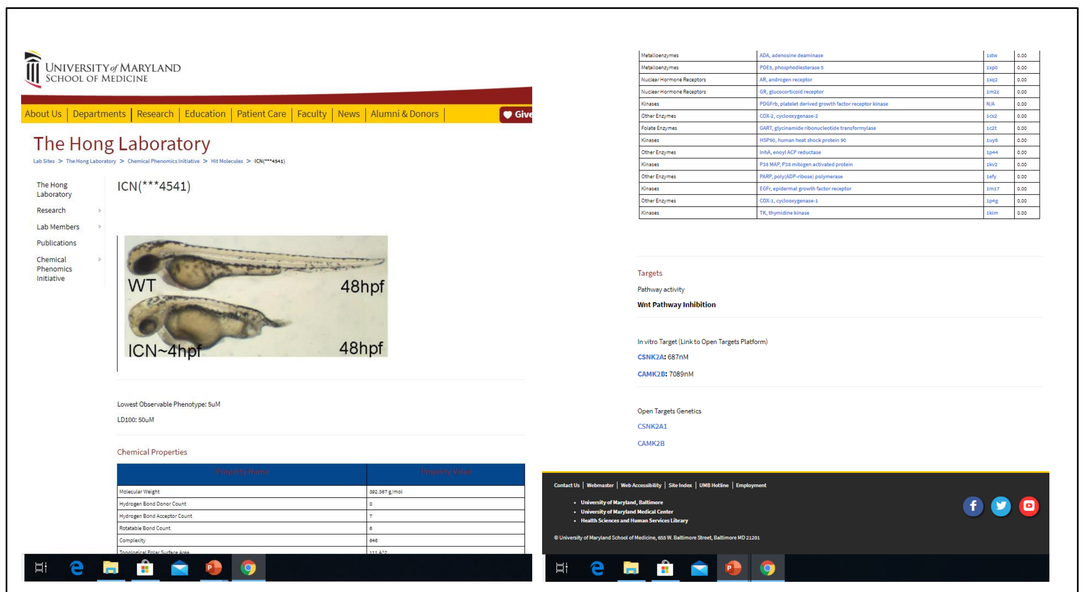
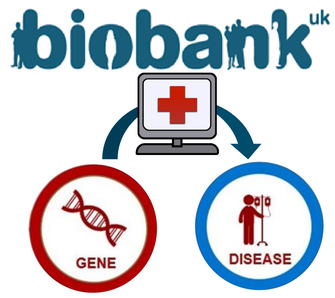


Fig 4. Prototype website under development: Example of ICN4541 compound with chemical properties and validated chemical target. Links have been developed to link out to other resources about CSNK2a1 and CSNK2a2. Further developments coming including further population with identified hits, query-able content, and incorporation of UKBiobank based PheWAS therapeutic target prediction.

Translation beyond discovery



UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government and the Northwest Regional Development Agency. In addition to information collected during the baseline assessment, 100,000 UK Biobank participants have worn a 24-hour activity monitor for a week, and 20,000 have undertaken repeat measures. A programme of online questionnaires is being rolled out (diet, cognitive function, work history and digestive health) and UK Biobank has embarked on a major study to scan (image) 100,000 participants (brain, heart, abdomen, bones & carotid artery). UK Biobank is linking to a wide range of electronic health records (cancer, death, hospital episodes, general practice), and is developing algorithms to accurately identify diseases and their sub-sets. Blood biochemistry is being analysed (such as hormones & cholesterol). Genotyping has been undertaken on all 500,000 participants and these data are being used in health research.

The depth of the data generated by the size of Uk Biobank allows for powering of statistical tests for even rare variants. Phenome wide association studies (PheWAS) of a drug target can identify known consequences of chronic activation or inactivation of a gene in the context of a human at a population level. This allows for identification of both therapeutic avenues and potential pitfalls caused by undesired on-target effects.

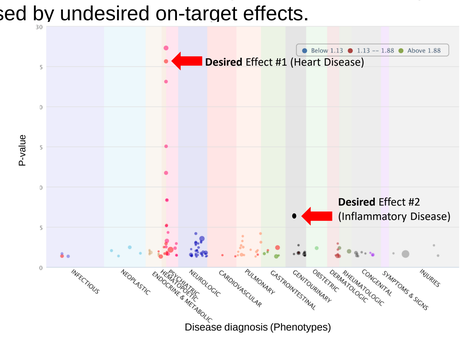


Fig 5. Example of a Uk Biobank powered Phenome wide association study identifies new potential therapeutic indications for novel small molecules.

Next steps

Manual screening and phenotype annotation introduces observer bias, recording bias, and significant investment in human capital. We are developing computer vision analysis techniques to automate the discovery and annotation pipeline.

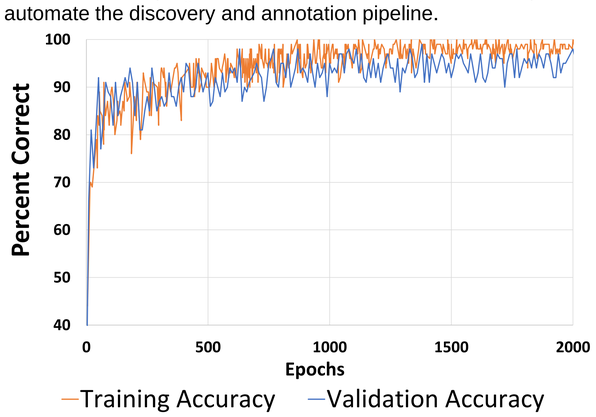


Fig 6. After 2000 epochs the model predicted >95% accuracy in predicting 1 of 8 defined phenotypes.

Research funded by NIGMS R01GM118557