



- (51) **International Patent Classification:**
G16H 50/20 (2018.01) *G16H 50/50* (2018.01)
- (21) **International Application Number:**
PCT/US2022/075006
- (22) **International Filing Date:**
16 August 2022 (16.08.2022)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/233,707 16 August 2021 (16.08.2021) US
- (71) **Applicant: INSITRO, INC.** [US/US]; 279 East Grand Avenue, Suite 200, South San Francisco, California 94080 (US).
- (72) **Inventors: CASALE, Francesco, Paolo;** c/o Insitro, Inc., 279 East Grand Avenue, Suite 200, South San Francisco, California 94080 (US). **BEREKET, Michael;** c/o Insitro, Inc., 279 East Grand Avenue, Suite 200, South San Francisco, California 94080 (US). **ALBERT, Matthew;** c/o Insitro, Inc., 279 East Grand Avenue, Suite 200, South San Francisco, California 94080 (US).
- (74) **Agent: XIAO, Jian et al.;** Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, California 94304-1018 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA,

(54) **Title:** DISCOVERY PLATFORM

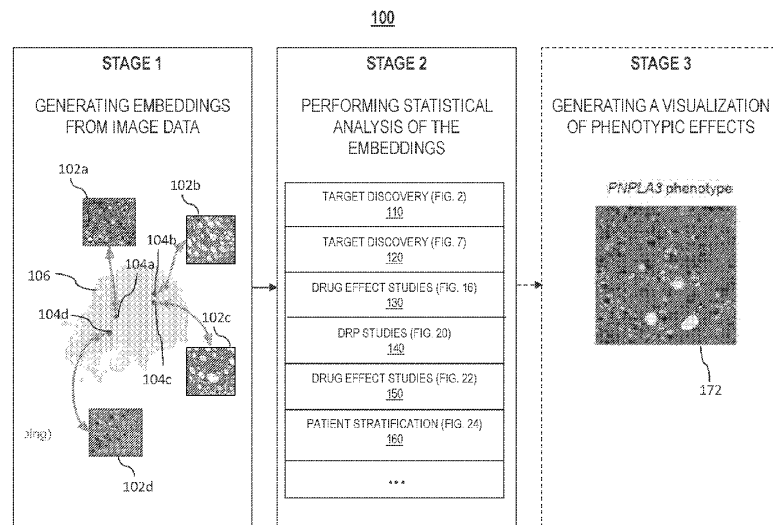


FIG. 1

(57) **Abstract:** The present disclosure relates to a discovery platform including machine-learning techniques for using medical imaging data to study a phenotype of interest, such as complex diseases with weak or unknown genetic drivers. An exemplary method of identifying a covariant of interest with respect to a phenotype comprises: receiving covariant information of a covariate class and corresponding phenotypic image data related to the phenotype obtained from a group of clinical subjects; inputting the phenotypic image data into a trained unsupervised machine-learning model to obtain a plurality of embeddings in a latent space, each embedding corresponding to a phenotypic state reflected in the phenotypic image data; and determining, based on the covariant information for the group of clinical subjects, the plurality of embeddings, and, one or more linear regression models, an association between each candidate covariant of a plurality of candidate covariants and the phenotype state to identify the covariant of interest.

[Continued on next page]

NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
 - before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
-