(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 January 2020 (16.01.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/014280 A1

(51) International Patent Classification:
G16B 40/20 (2019.01) *G06N 3/04* (2006.01)
G16B 20/20 (2019.01)

(21) International Application Number:
PCT/US2019/041078

(22) International Filing Date:
09 July 2019 (09.07.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/696,699	11 July 2018 (11.07.2018)	US
2021473	16 August 2018 (16.08.2018)	NL
16/505,100	08 July 2019 (08.07.2019)	US

(71) Applicant: **ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).

(72) Inventors: **KASHEFHAGHIGHI, Dorna**; 5200 Illumina Way, San Diego, California 92122 (US). **KIA, Amirali**; 5200 Illumina Way, San Diego, California 92122 (US). **FARH, Kai-How**; 5200 Illumina Way, San Diego, California 92122 (US).

(74) Agent: **DURDIK, Paul A.** et al.; 637 Main St., Half Moon Bay, California 94019 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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).

(54) Title: DEEP LEARNING-BASED FRAMEWORK FOR IDENTIFYING SEQUENCE PATTERNS THAT CAUSE SEQUENCE-SPECIFIC ERRORS (SSEs)

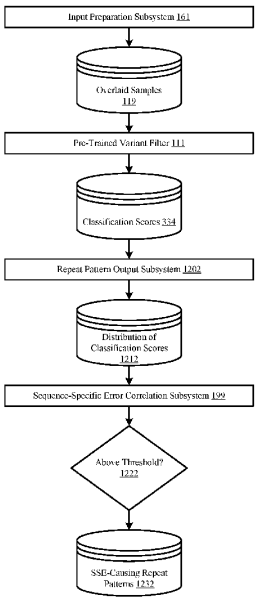


FIG. 12

(57) Abstract: The technology disclosed presents a deep learning-based framework, which identifies sequence patterns that cause sequence-specific errors (SSEs). Systems and methods train a variant filter on large-scale variant data to learn causal dependencies between sequence patterns and false variant calls. The variant filter has a hierarchical structure built on deep neural networks such as convolutional neural networks and fully-connected neural networks. Systems and methods implement a simulation that uses the variant filter to test known sequence patterns for their effect on variant filtering. The premise of the simulation is as follows: when a pair of a repeat pattern under test and a called variant is fed to the variant filter as part of a simulated input sequence and the variant filter classifies the called variant as a false variant call, then the repeat pattern is considered to have caused the false variant call and identified as SSE-causing.

WO 2020/014280 A1

[Continued on next page]

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*