



- (51) **International Patent Classification:**
G16B 30/00 (2019.01) *C12Q 1/6869* (2018.01)
- (21) **International Application Number:**
PCT/US2022/075287
- (22) **International Filing Date:**
22 August 2022 (22.08.2022)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/245,639 17 September 2021 (17.09.2021) US
- (71) **Applicants:** **ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US). **ILLUMINA SOFTWARE, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).
- (72) **Inventors:** **GROS, Thomas**; 5200 Illumina Way, San Diego, California 92122 (US). **CHESNY, Zoey Wei**; 5200 Illumina Way, San Diego, California 92122 (US).
- (74) **Agent:** **FUNK, Christopher**; 1010 North 500 East, Suite 210, North Salt Lake, Utah 84054 (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, 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))

- (54) **Title:** AUTOMATICALLY IDENTIFYING FAILURE SOURCES IN NUCLEOTIDE SEQUENCING FROM BASE-CALL-ERROR PATTERNS

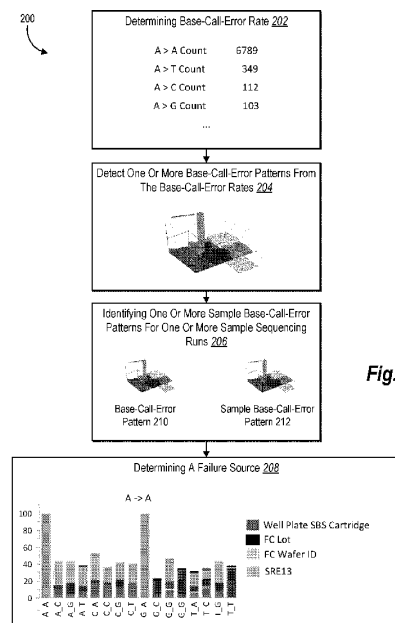


Fig. 2

- (57) **Abstract:** Methods, systems, and non-transitory computer readable media are disclosed for accurately and efficiently identifying base-call-error scars or patterns from sequencing data to determine failure sources that contribute to the base-call-error scars or patterns. For example, the disclosed system can utilize a reference genome to determine nucleotide-specific errors within a run of a sequencing pipeline. Based on the co-occurrence of different nucleotide-specific errors, the disclosed system can determine a base-call-error scar. The disclosed system can further determine one or more sample error scars from sample sequencing runs that correlate to the base-call-error scar. Based on the correlation and by utilizing a statistical model, the disclosed system can identify failure sources contributing to the nucleotide-specific errors within the base-call-error scar.