



(51) International Patent Classification:

G16B 40/30 (2019.01) G06N 3/04 (2006.01)
C12Q 1/6869 (2018.01) G06Q 10/04 (2012.01)
G06F 17/16 (2006.01) G16B 50/50 (2019.01)

(21) International Application Number:

PCT/EP2021/075306

(22) International Filing Date:

15 September 2021 (15.09.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2051077-2 15 September 2020 (15.09.2020) SE

(72) Inventors; and

(71) Applicants: DHAPOLA, Parashar [IN/SE]; 28 D Sankt
Hans gränd, Lgh 1102, 226 42 Lund (SE). KARLSSON,
Göran [SE/SE]; Ekelundsvägen 14, 22472 Lund (SE).

(74) Agent: HØIBERG P/S; Adelgade 12, 1304 Copenhagen K
(DK).

(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, IT, 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).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

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

(54) Title: METHOD AND SYSTEM FOR SUBSAMPLING OF CELLS FROM SINGLE-CELL GENOMICS DATASET

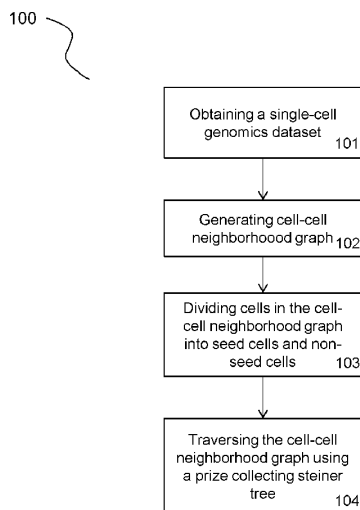


FIG. 2

(57) Abstract: The present disclosure relates to a computer-implemented method of extracting a subsample of cells from a plurality of cells in a single-cell genomics dataset, the method comprising the steps of: obtaining a single-cell genomics dataset represented in at least two dimensions, wherein information about each cell is represented in a first dimension and information about genomic features is represented in a second dimension; generating a cell-cell neighborhood graph from the single-cell genomics dataset, the cell-cell neighborhood graph providing information about similarities of the genomic features of the cells, wherein the cells are represented as vertices in the cell-cell neighborhood graph; dividing the cells in the cell-cell neighborhood graph into seed cells and non-seed cells; assigning at least one first prize to the seed cells and at least one second prize to the non-seed cells in the cell-cell neighborhood graph; and traversing the cell-cell neighborhood graph using a prize collecting steiner tree algorithm to obtain a subsample of cells. The present disclosure further relates to a computer system for extracting a subsample of cells from a plurality of cells in a single-cell genomics dataset.