Mapping Cancer Markers, June 2021 update

Code update for the Mapping Cancer Markers (MCM) application

While planning the MCM transition to sarcoma, we needed to expand the number of comparisons we could explore with our current code. For example, exploring biomarkers that distinguish leimyosarcoma subtypes (i.e., soft-tissue and uterine); unfortunately, such flexibility in work unit design was not possible with the existing MCM application. Thus, we made changes to the code to expand the MCM application's ability to deal with multi-label data sets, and allow it to search for more specific signatures within them.

Before the sarcoma data set, MCM analyzed lung and ovarian cancer data sets, both of which have binary labels. Our lung cancer data set labelled samples as either “cancer” or “no cancer” – as we were searching for diagnostic signature. Our ovarian data set labelled samples as short or long survival – searching for prognostic signature. The sarcoma data set is multi-label, and labels samples with seven different subtypes of sarcoma. Thus, the update would enable us to find signatures that differentiate the histological subtypes and identifies specific patient subgroups.

When a dataset has binary labels, MCM will find signatures that can predict that binary label. With a multi-label data set, we can direct MCM to search for either binary or multiclass signatures. Currently, MCM is searching for both in the sarcoma data set. A multiclass sarcoma signature distinguishes every subtype from every other; given any sarcoma sample, it will diagnose the specific cancer subtype. A binary sarcoma signature distinguishes one group of subtypes from the rest, but does not distinguish specific subtypes. For example, among the seven sarcoma subtypes are two leimyosarcoma subtypes, soft-tissue and uterine. MCM is presently searching for binary signatures that distinguish leimyosarcoma from the rest.

This small extension to the application adds the needed capabilities for sarcoma project, and provide additional flexibility for future projects, while preserving backwards compatibility. In the recent months, World Community Grid members have processed thousands of work units to beta-test the new code. These changes give MCM the power to make fine-grained adjustments that tailor the dataset to the precise question each work unit will explore.

Below are some examples to illustrate the point. Figure 1A represents the sarcoma dataset, with seven subtypes of samples. MCM can use the multi-labelled dataset as-is and search for multiclass (one-against-all) signatures. (MCM can do this now.)

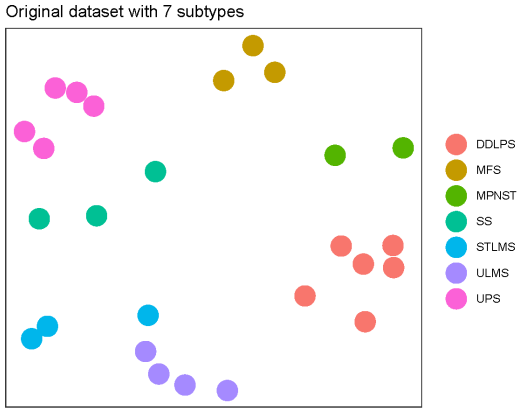
Figure 1B shows how MCM can reduce sarcoma to a binary dataset, by splitting the subtypes into two groups and searching for binary signatures. (MCM can do this now.)

Figure 1C shows an alternative reduction to a binary dataset. (Ditto.)

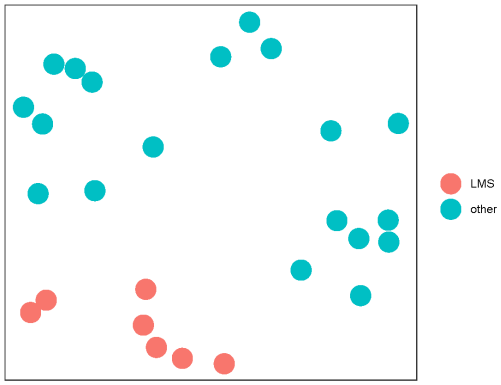
Figure 1D shows how MCM's new capabilities will allow a work unit to focus on specific subtypes. Excluded samples are grey and crossed-out. (New code.)

Figure 1E shows the MCM's new capabilities in full, excluding individual samples and changing their labels. Relabeled samples are outlined in black. (New code.)

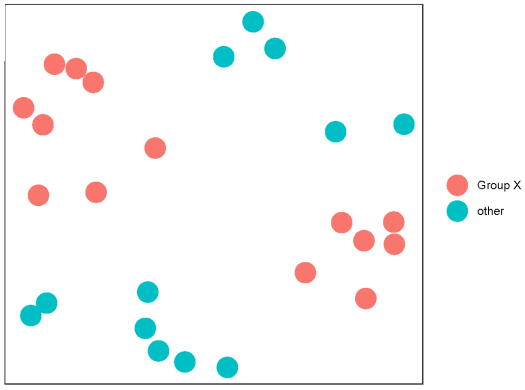
1A


Original dataset with 7 subtypes
DDLPS
MFS
MPNST
SS
STLMS
ULMS
UPS

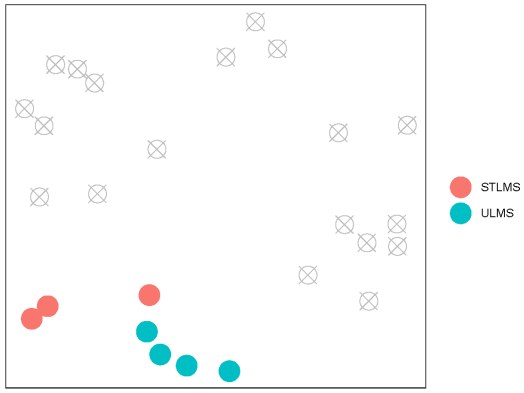
1B


Binary split of dataset
LMS
other

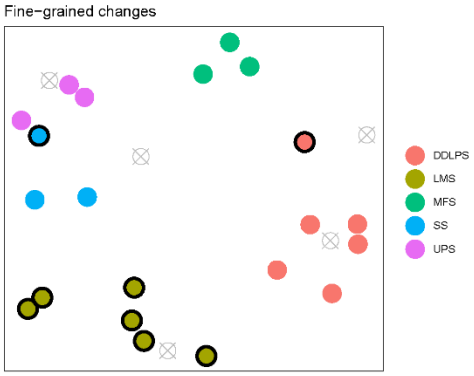
1C


Alternative binary split
Group X
other

1D


Focus on specific subtypes
STLMS
ULMS

1E


Fine−grained changes
DDLPS
LMS
MFS
SS
UPS

Additional Information

Publications

Since the last update, we have published multiple manuscripts and developed additional algorithms and databases to benefit the comprehensive MCM results analyses, but several also, not surprisingly, took a diversion towards exploring important details about COVID-19:

- Rahmati S, O'Rielly DD, Li Q, Codner D, Dohey A, Jenkins K, Jurisica I, Gladman DD, Chandran V, Rahman P. Rho-GTPase pathways may differentiate treatment response to TNF-alpha and IL-17A inhibitors in psoriatic arthritis. Sci Rep. 10(1):21703, 2020.Porras P, Barrera E, Bridge A, Del-Toro N, Cesareni G, Duesbury M, Hermjakob H, Iannuccelli M, Jurisica I, Kotlyar M, Licata L, Lovering RC, Lynn DJ, Meldal B, Nanduri B, Paneerselvam K, Panni S, Pastrello C, Pellegrini M, Perfetto L, Rahimzadeh N, Ratan P, Ricard-Blum S, Salwinski L, Shirodkar G, Shrivastava A, Orchard S. Towards a unified open access dataset of molecular interactions. Nat Commun. 11(1):6144, 2020.
- Glaus Garzon JF, Pastrello C, Jurisica I, Hottiger MO, Wenger RH, Borsig L. Tumor cell endogenous HIF-1α activity induces aberrant angiogenesis and interacts with TRAF6 pathway required for colorectal cancer development. Neoplasia. 22(12):745-758, 2020.
- Perfetto L, Pastrello C, Del-Toro N, Duesbury M, Iannuccelli M, Kotlyar M, Licata L, Meldal B, Panneerselvam K, Panni S, Rahimzadeh N, Ricard-Blum S, Salwinski L, Shrivastava A, Cesareni G, Pellegrini M, Orchard S, Jurisica I, Hermjakob HH, Porras P. The IMEx Coronavirus interactome: an evolving map of Coronaviridae-Host molecular interactions, Database, 2020. baaa096. doi: 10.1093/database/baaa096
- Ali SA, Gandhi R, Potla P, Keshavarzi S, Espin-Garcia O, Shestopaloff K, Pastrello C, Bethune-Waddell D, Lively S, Perruccio AV, Rampersaud YR, Veillette C, Rockel JS, Jurisica I, Appleton CT, Kapoor M. Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis Cartilage, 28(11):1471-1481, 2020.
- Clotet-Freixas S, McEvoy CM, Batruch I, Pastrello C, Kotlyar M, Van JAD, Arambewela M, Boshart A, Farkona S, Niu Y, Li Y, Famure O, Bozovic A, Kulasingam V, Chen P, Kim SJ, Chan E, Moshkelgosha S, Rahman SA, Das J, Martinu T, Juvet S, Jurisica I, Chruscinski A, John R, Konvalinka A. Extracellular Matrix Injury of Kidney Allografts in Antibody-Mediated Rejection: A Proteomics Study, J Am Soc Nephrol, 31(11):2705-2724, 2020.
- Yao Z, Aboualizadeh F, Akula I, Snider J, Tang P, Kotlyar M, Jurisica I, Stagljar I. Split Intein-Mediated Protein Ligation, a Novel Method for Detecting Protein-Protein Interactions and Their Inhibition, Nat Commun, 11(1):2440, 2020.
- Van JAD, Clotet-Freixas S, Hauschild AC, Batruch I, Jurisica I, Elia Y, Mahmud FH, Sochett E, Diamandis EP, Scholey JW, Konvalinka A. Urinary proteomics links keratan sulfate degradation and lysosomal enzymes to early type 1 diabetes, PLoS One, 15(5):e0233639, 2020. doi: 10.1371/journal.pone.0233639.
- Sinsky J, Majerova P, Kovac A, Kotlyar M, Jurisica I, Hanes J. Physiological Tau Interactome in Brain and Its Link to Tauopathies, J Proteome Res, 19(6):2429-2442, 2020. doi: 10.1021/acs.jproteome.0c00137.
- Agapito G, Pastrello C, Guzzi PH, Jurisica I, Cannataro M. BioPAX-Parser: parsing and enrichment analysis of BioPAX pathways, Bioinformatics, 36(15):4377-4378, 2020. doi: 10.1093/bioinformatics/btaa529.
- Pinheiro M, Lupinacci FCS, Santiago KM, Drigo SA, Marchi FA, Fonseca-Alves CE, Andrade SCDS, Aagaard MM, Basso TR, Dos Reis MB, Villacis RAR, Roffé M, Hajj GNM, Jurisica I, Kowalski LP, Achatz MI, Rogatto SR. Cancers (Basel), 12(5):E1289, 2020. doi: 10.3390/cancers12051289.
- Reis PP, Tokar T, Goswami RS, Xuan Y, Sukhai M, Seneda AL, Moz LES, Perez-Ordonez B, Simpson C, Goldstein D, Brown D, Gilbert R, Gullane P, Irish J, Jurisica I, Kamel-Reid S. A 4-gene

signature from histologically normal surgical margins predicts local recurrence in patients with oral carcinoma: clinical validation, Scientific Reports, 10(1):1713, 2020. doi: 10.1038/s41598-020-58688-y.

- Tokar, T., Pastrello, C., Jurisica, I. GSOAP: A tool for visualization of gene set over-representation analysis, Bioinformatics, Bioinformatics, 36(9):2923-2925, 2020. doi: 10.1093/bioinformatics/btaa001
- Rahmati R, Abovsky M, Pastrello C, Kotlyar M, Lu R, Cumbaa CA, Rahman P, Chandran V, Jurisica I, pathDIP 4: an extended pathway annotations and enrichment analysis resource for human, model organisms and domesticated species, Nucl Acids Res, 48(D1): D479–D488, 2020, https://doi.org/10.1093/nar/gkz989
- Kennedy, S., Jarboui, M-A, Srihari, S, Raso, C, Bryan, K, Dernayka, L, Charitou. T, Bernal-Llinares, M, Herrera-Montavez, C, Krstic, A, Matallanas, D, Kotlyar, M, Jurisica, I, Curak, J, Wong, V, Stagljar, I, LeBihan, T, Imrie, L, Pillai, P, Lynn, M, Fasterius, E, Szigyarto, C. A-K, Breen, J, Kiel, C, Serrano, L, Rauch, N, Rukhlenko, O, Kholodenko, B, Iglesias-Martinez, L, Ryan, C, Pilkington, R, Cammareri, P, Sansom, O, Shave, S, Auer, M, Horn, N, Klose, F, Ueffing, M, Boldt, K, Lynn, D, Kolch, W, Extensive Rewiring of the EGFR Network in Colorectal Cancer Cells Expressing Transforming Levels of KRASG13D, Nat Commun, 11(1):499, 2020. doi: 10.1038/s41467-019-14224-9
- Cohn DE, Barros-Filho MC, Minatel BC, Pewarchuk ME, Marshall EA, Vucic EA, Sage AP, Telkar N, Stewart GL, Jurisica I, Reis PP, Robinson WP, Lam WL. Share, Reactivation of Multiple Fetal miRNAs in Lung Adenocarcinoma, Cancers, 13(11):2686, 2021.
- McEvoy CM, Clotet-Freixas S, Tokar T, Pastrello C, Reid S, Batruch I, RaoPeters AAE, Kaths JM, Urbanellis P, Farkona S, Van JAD, Urquhart BL, John R, Jurisica I, Robinson LA, Selzner M, Konvalinka A., Normothermic Ex-vivo Kidney Perfusion in a Porcine Auto-Transplantation Model Preserves the Expression of Key Mitochondrial Proteins: An Unbiased Proteomics Analysis, Mol Cell Proteomics, 100101, 2021. doi: 10.1016/j.mcpro.2021.100101.
- Freire PP, Marques AH, Baiocchi GC, Schimke LF, Fonseca DL, Salgado RC, Filgueiras IS, Napoleao SM, Plaça DR, Akashi KT, Hirata TDC, El Khawanky N, Giil LM, Cabral-Miranda G, Carvalho RF, Ferreira LCS, Condino-Neto A, Nakaya HI, Jurisica I, Ochs HD, Camara NOS, Calich VLG, Cabral-Marques O., The relationship between cytokine and neutrophil gene network distinguishes SARS-CoV-2-infected patients by sex and age, JCI Insight 6(10):147535, 2021.
- Berra G, Farkona S, Mohammed-Ali Z, Kotlyar M, Levy L, Clotet-Freixas S, Ly P, Renaud-Picard B, Zehong G, Daigneault T, Duong A, Batruch I, Jurisica I, Konvalinka A, Martinu T. Association between renin-angiotensin system and chronic lung allograft dysfunction, Eur Respir J, 2002975, 2021. doi: 10.1183/13993003.02975-2020.
- Agapito, G., Pastrello, C., Jurisica, I. Comprehensive pathway enrichment analysis workflows: Covid-19 case study, Briefings in Bioinformatics, In press.
- Chakraborty M, Chu K, Shrestha A, Revelo XS, Zhang X, Gold MJ, Khan S, Lee M, Huang C, Akbari M, Barrow F, Chan YT, Lei H, Kotoulas NK, Jovel J, Pastrello C, Kotlyar M, Goh C, Michelakis E, Clemente-Casares X, Ohashi PS, Engleman EG, Winer S, Jurisica I, Tsai S, Winer DA., Mechanical Stiffness Controls Dendritic Cell Metabolism and Function, Cell Reports, 34(2):108609, 2021.
- Bhat, M. Elisa Pasini, Chiara Pastrello, Sarah Rahmati, Marc Angeli, Max Kotlyar, Anand Ghanekar and Igor Jurisica, Integrative Analysis of Layers of Data in Hepatocellular Carcinoma Reveals Pathway Dependencies, World J Hepatology, 13(1):94-108, 2021.
- Nakamura A, Rampersaud YR, Sundararajan K, Nakamura S, Wu B, Matip E, Haroon N, Krawetz RJ, Rossomacha E, Gandhi R, Kotlyar M, Rockel JS, Jurisica I, Kapoor M. Zinc finger protein-440 promotes cartilage degenerative mechanisms in human facet and knee osteoarthritis chondrocytes. Osteoarthritis Cartilage. 29(3):372-379, 2021. doi: 10.1016/j.joca.2020.12.004.
- Endisha H, Datta P, Sharma A, Nakamura S, Rossomacha E, Younan C, Ali SA, Tavallaee G, Lively S, Potla P, Shestopaloff K, Rockel JS, Krawetz R, Mahomed NN, Jurisica I, Gandhi R, Kapoor M. MicroRNA-34a-5p Promotes Joint Destruction during Osteoarthritis, Arthritis Rheumatol, 73(3):426-

439, 2021

Grants

We have been able to secure several grants to enable funding for the project, in the cancer direction, but also into neurodegenerative diseases, using tools and approaches we built in cancer informatics, including:

- Identification and characterization of microRNAs essential for the development and fitness of chemoresistant AML from Canadian Institute for Health Research
- Identifying Metabolomic Markers of Disease Activity in Psoriatic Disease using Solid Phase Microextraction - Liquid Chromatography - High Resolution Mass Spectrometry from Canadian Institute for Health Research
- Understanding who does and does not respond to an interprofessional education and self-management model of care for chronic low back pain from Canadian Institute for Health Research
- Novel methods for integrative computational biology from Natural Sciences and Engineering Council of Canada
- Interactome mapping of disease-related proteins using split intein-mediated protein ligation (SIMPL) from Genome Canada,
- The Next Generation Signalling Biology Platform from Ontario Research Funds
- Multi-level integrative ‘omics to identify biomarkers in a Schizophrenia and other major psychoses from Canadian Institutes of Health Research in collaboration with European funding agencies
- Integrative structural biology of pathological tau protein, an appealing therapeutic target for Alzheimer´s disease modifying drugs from Horison2020 European funding.

Thank you for all the contributed computing power that makes this research possible.

MCM Team