BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Benjamin J. Raphael

eRA COMMONS USER NAME (credential, e.g., agency login): braphael

POSITION TITLE: Professor of Computer Science

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Massachusetts Institute of Technology	B.S.	1996	Mathematics
University of California, San Diego	Ph.D.	2002	Mathematics
University of California, San Diego	Postdoctoral	2006	Computer Science (Bioinformatics)

A. Personal Statement

I have the expertise, training, leadership and motivation necessary to successfully carry out the proposed research project. I have a broad background in computational biology, and specific training and expertise in computer science, mathematics, and biology. My research group focuses on the development and application of novel algorithms for spatial transcriptomics, single-cell DNA/RNA sequencing; tumor heterogeneity and evolution; pathway/network analysis of germline and somatic variants; and structural variation analysis. I currently lead a Genome Data Analysis Center (GDAC) from the NCI and I was part of multiple analysis working groups within The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC). The results of our analyzes appear in dozens of publications from these consortia. I co-led the TCGA Pancreatic Adenocarcinoma project and the Networks and Pathways analysis for the ICGC Pan-Cancer Analysis of Whole Genomes (PCAWG) project. Specific to the proposed project, I have an extensive publication record on methods for spatial transcriptomics and single-cell DNA/RNA sequencing. My group has also released dozens of open-source software packages implementing our algorithms which we release and support on github. I have graduated 7 Ph.D. students, mentored 9 completed postdoctoral fellows, and currently advise 9 Ph.D. students and 4 postdoctoral fellows. In summary, I have a demonstrated record of successful and productive research projects and mentoring in single-cell and spatial sequencing, cancer genomics, and network analysis.

a. C. Ma, U. Chitra, S. Zhang, **B.J. Raphael**. (2022) Belayer: Modeling discrete and continuous spatial variation in gene expression from spatially resolved transcriptomics. *Cell Systems*. (In press).
b. R. Zeira, M. Land, A. Strzalkowski, **B.J Raphael**. (2022) Alignment and integration of spatial transcriptomics data. *Nature Methods* 19 (5), 567-575. PMCID: In process
c. S. Zaccaria and **B.J. Raphael**. (2021) Characterizing the allele- and haplotype-specific copy number landscape of cancer genomes at single-cell resolution with CHISEL. *Nature Biotechnology* Feb;39(2):207-214. PMCID: PMC9876616.
d. R. Elyanow, B. Dumitrascu, B.E. Engelhardt, **B.J. Raphael** (2020) netNMF-sc: Leveraging gene-gene interactions for imputation and dimensionality reduction in single-cell expression analysis. *Genome Research* 30(2):195-204. PMCID: PMC7050525

B. Positions and Honors

Positions and Employment

2018-	Affiliate Member, New York Genome Center
2017-	Member, Rutgers Cancer Institute of New Jersey
2016-	Professor, Department of Computer Science, Princeton University
2013-2016	Director, Center for Computational and Molecular Biology, Brown University
2011-2016	Associate Professor, Department of Computer Science & Center for Computational Molecular Biology, Brown University
2006-2011	Assistant Professor, Department of Computer Science & Center for Computational Molecular Biology, Brown University

Other Experience and Professional Memberships

2018-	Member, Human Cell Atlas consortium
2014-2020	Co-leader, Networks and Pathways Analysis Group, ICGC
2014-2017	Co-leader, TCGA Pancreatic Cancer Analysis Working Group
2010-2013	Leader, Structural Aberration Analysis Group, International Cancer Genome Consortium (ICGC)
2009-	Member, The Cancer Genome Atlas, Analysis Working Group
2002-	Member, International Society for Computational Biology
1996-2002	Member, American Mathematical Society

Honors

2021	ISCB Innovator Award
2020	Elected Fellow, International Society of Computational Biology (ISCB)
2020	AACR Team Science Award
2011	National Science Foundation (NSF) CAREER Award
2010-2012	Sloan Research Fellowship, Molecular Biology
2007	Brown Center for Computational Molecular Biology, CCMB Innovator Award
2006	Named one “Tomorrow’s PI’s” by Genome Technology magazine
2005-2010	Burroughs Welcome Career Award at the Scientific Interface
2002-2004	Alfred P. Sloan Foundation / DOE Fellowship in Computational Biology
1999	U.S. Department of Education, GAANN Fellowship in Mathematics (declined)

Grant Review Panels

2018	National Science Foundation
2016	NIH Study Section Member (ad hoc). MABS.
2012-2013	NIH Study Section Member (ad hoc). BDMA and GCAT.
2011	National Cancer Institute Special Panel, ENCODE Panel
2009-2010	Ministry of Education, Singapore
2008	National Cancer Institute of Canada
2008	National Institute of General Medical Science
2008, 2010, 2017	National Science Foundation
2007	NSF/NIGMS Mathematical Biology Study Section, member

C. Contributions to Science

1. Single-cell and spatial DNA and RNA sequencing. Single-cell DNA/RNA sequencing data and spatial transcriptomics data have high rates of errors and missing data that complicate its analysis. We have developed algorithms to address these complications by combining information across individual cells and/or neighboring genomic loci. We recently developed PASTE to align and integrate spatial trancriptomics data from multiple tissue slices (Ziera, et al. 2022). For single-cell DNA sequencing data, we developed CHISEL, the first method to identify allele-specific copy number aberrations in low-coverage single-cell sequencing data (Zaccaria and Raphael 2018), and designed SCARLET, an algorithm to infer phylogenetic trees from targeted single-cell DNA sequencing data using both single-nucletotide variants and copy number aberrations (Satas et al. 2020). We have also used prior knowledge of gene coexpression to aid in dimensionality reduction and imputation in single-cell RNA sequencing (Elyanow et al. 2020).

a. R. Zeira, M. Land, A. Strzalkowski, **B.J Raphael**. (2022) Alignment and integration of spatial transcriptomics data. *Nature Methods* 19 (5), 567-575. PMCID: In process
b. C. Ma, U. Chitra, S. Zhang, **B.J. Raphael**. (2022) Belayer: Modeling discrete and continuous spatial variation in gene expression from spatially resolved transcriptomics. *Cell Systems*. (In press).
c. S. Zaccaria and **B.J. Raphael**. (2021) Characterizing the allele- and haplotype-specific copy number landscape of cancer genomes at single-cell resolution with CHISEL. *Nature Biotechnology* Feb;39(2):207-214. PMCID: PMCID: PMC9876616.
d. R. Elyanow, B. Dumitrascu, B.E. Engelhardt, **B.J. Raphael** (2020) netNMF-sc: Leveraging gene-gene interactions for imputation and dimensionality reduction in single-cell expression analysis. *Genome Research* 30(2):195-204. PMCID: PMC7050525.

2. Tumor heterogeneity and evolution. The cells in a tumor typically differ in their complement of somatic mutations, reflecting the ancestral relationships between cells. This intra-tumor heterogeneity complicates cancer genome studies, particularly because most cancer sequencing data is from a single sample from a tumor, containing thousands-millions of cells. However, careful analysis of intra-tumor heterogeneity can also reveal information about the evolutionary history of the tumor. High-coverage cancer genome sequencing provides an unprecedented opportunity to quantify this heterogeneity and improve our understanding of the underlying mutational process. We have developed several algorithms to identify tumor clones and to reconstruct the phylogenetic tree that relates these clones from bulk sequencing data of multiple tumors. These algorithms are based on a rigorous mathematical formulation of the problem as a matrix factorization problem with explicit evolutionary constraints (El-Kebir et al 2015). We have extended this approach to model additional mutation types (El-Kebir et al. 2016; Satas et al. 2017) and to infer patterns of metastatic spread (El-Kebir et al. 2018).

a. M. El-Kebir*, L. Oesper*, H. Acheson-Field, **B.J. Raphael**. (2015) Reconstruction of clonal trees and tumor composition from multi-sample cancer sequencing data. *Bioinformatics [Proceedings of ISMB 2015]*. 31(12):i62-70. PMCID: PMC4542783.
b. M. El-Kebir M, G. Satas G, L. Oesper L, **B.J. Raphael**. (2016) Inferring the Mutational History of a Tumor Using Multi-state Perfect Phylogeny Mixtures. *Cell Systems*. 3(1):43-53. PMCID: in progress (PMID: 27467246).
c. M. El-Kebir, G. Satas, **B.J. Raphael**. (2018) Inferring parsimonious migration histories for metastatic cancers. *Nature Genetics* 50(5):718-726. PMCID: PMC6103651.
d. G. Satas, S. Zaccaria, G. Mon, and **B.J. Raphael**. (2020) SCARLET: Single-Cell Tumor Phylogeny Inference with Copy-Number Constrained Mutation Losses. *Cell Systems* 10 (4), 323-332. e8. PMCID: PMC7451135.

3. Copy number aberrations and structural variation. Copy number aberrations (CNAs) are common somatic mutations cancer, particularly in solid tumors, but are often challenging to detect in short-read DNA sequencing data of bulk tumors. We have developed several approaches to detect CNAs, identify tumor clones distinguished by these aberrations, and infer the proportions of distinct tumor clones and normal admixture in a bulk tumor sample. Our THetA algorithm (Oesper et al. 2013) was one of the *first* approaches for the later problem and has become one of the standard approaches in the field. We also recently developed HATCHet (Zaccaria and Raphael 2020) to infer allele-specific copy number aberrations simultaneously across multiple bulk tumor samples and RCK (Aganezov and Raphael 2020) to reconstruct allele-specific genome graphs from

bulk tumor samples. We also developed a method to identify copy number aberrations from spatial transcriptomics data (Elyanow et al. 2021).

a. L. Oesper, A. Mahmoody, **B.J. Raphael**. (2013) THetA: Inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. *Genome Biology* 14:R80. PMCID: PMC4054893.
b. S. Aganezov and B.J Raphael. (2020) Reconstruction of clone-and haplotype-specific cancer genome karyotypes from bulk tumor samples. *Genome Research* Sep;30(9):1274-1290. PMCID: In process.
c. S. Zaccaria, **B.J. Raphael.** (2020) Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nature Communications* Sep 2;11(1):4301. PMCID: PMC7468132.
d. R. Elyanow, R. Zeira, M. Land, **B.J. Raphael** (2021) STARCH: Copy number and clone inference from spatial transcriptomics data. *Physical Biology*, Volume 18, Number. PMCID: PMC9876615.

4. Network analysis of genetic variants. There are numerous methods for testing whether pre-defined gene sets (pathways) are enriched for DNA sequence variants. However, such methods do not allow one to identify novel pathways or interactions/crosstalk between pathways. Through our involvement in The Cancer Genome Atlas (TCGA), we realized that the availability of sequencing datasets from many samples allowed the possibility of *discovering* novel gene sets by identifying subnetworks of a genome-scale interaction network that are enriched for genetic variants across multiple samples. Finding such subnetworks is a challenging problem because significant subnetworks are determined by *both* the frequency (or association score) of variants within genes in the subnetwork and the local topology of the subnetwork. We developed the *HotNet* algorithm (Vandin et al. 2012) to address this problem. We further improved our approach for network analysis of variants, introducing *HotNet2* algorithm (Leiserson et al. 2015) and recently the Hierarchical-HotNet algorithm (Reyna et al. 2018). The *HotNet*/*HotNet2* algorithms have been used in multiple publications from TCGA. We used *HotNet2* to perform a Pan-Cancer study of >3000 samples from 12 TCGA cancer types, identifying a number of novel genes and subnetworks that are mutated across cancer types, but with rare mutations that were distinguishable from noise when considered individually.

a. F. Vandin, E. Upfal, **B.J. Raphael**. (2011) Algorithms for detecting significantly mutated pathways in cancer. *Journal of Computational Biology*. 18(3):507-22 [Journal version of paper accepted at *Proceedings of the 14th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2010)*]. PMCID: In process
b. M. Leiserson F. Vandin, …(16 additional authors)…, **B.J. Raphael**. (2015) Pan-Cancer Network Analysis Identifies Combinations of Rare Somatic Mutations across Pathways and Protein Complexes. *Nature Genetics*. 47(2):106-114. PMCID: PMC4444046.
c. M.R. Reyna, M.D.M. Leiserson, **B.J. Raphael.** (2018) Hierarchical HotNet: identifying hierarchies of altered subnetworks. *Bioinformatics* [*ECCB 2018 Proceedings*] 34(17): i972-i980. PMCID: PMC6129270
d. Reyna MA, Haan D, Paczkowska M, Verbeke LPC, Vazquez M, Kahraman A, Pulido-Tamayo S, Barenboim J, Wadi L, Dhingra P, Shrestha R, Getz G, Lawrence MS, Pedersen JS, Rubin MA, Wheeler DA, Brunak S, Izarzugaza JMG, Khurana E, Marchal K, von Mering C, Sahinalp SC, Valencia A; PCAWG Drivers and Functional Interpretation Working Group, Reimand J, Stuart JM, **Raphael BJ**; PCAWG Consortium (2020) Pathway and network analysis of more than 2500 whole cancer genomes. *Nature Communications*. 11(1):729. PMCID: PMC7002574

5. Combinations of variants through mutual exclusivity. Early cancer sequencing studies demonstrated that in many cases somatic mutation of a *single* gene in a pathway (e.g. a pathway controlling cell growth) is sufficient to perturb this pathway. This implies that when examining somatic mutations from many patients, mutations in the same pathway will exhibit a pattern of *mutual exclusivity*. Mutual exclusivity between mutations in pairs of genes has been widely reported, but we were the first to systematically predict pathways by identifying *sets* of genes that exhibit mutual exclusivity; i.e. we *do not use any prior knowledge of interactions* between genes. Our algorithms have had a positive impact in cancer genomics and were used in TCGA study of acute myeloid leukemia and TCGA Pan-Cancer project.

a. F. Vandin, E. Upfal, **B.J. Raphael**. (2012) *De novo* Discovery of Mutated Driver Pathways in Cancer. *Genome Research* 22(2):375-85. [Preliminary version appeared in *Proceedings of the 15th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2011)*]. PMCID: PMC3266044.

b. M.D.M. Leiserson, H-T. Wu, F. Vandin, **B.J. Raphael**. (2015) CoMEt: A Statistical Approach to Identify Combinations of Mutually Exclusive Alterations in Cancer. *Genome Biology*, 8;16:160. PMCID: PMC4531541.
c. M.D.M. Leiserson, M.A.Reyna, **B.J. Raphael**. (2016) A Weighted Exact Test for Mutually Exclusive Mutations in Cancer. Bioinformatics, 32(17):i736-i745. PMCID: PMC5013919
d. T.Y.Park, M.D.M. Leiserson, G.W. Klau, **B.J. Raphael**. (2022) SuperDendrix algorithm integrates genetic dependencies and genomic alterations across pathways and cancer types. *Cell Genomics* Feb 9;2(2):1000992. PMCID: PMC8979493

Complete List of Published Work in MyBibliography
http://www.ncbi.nlm.nih.gov/myncbi/benjamin.raphael.1/bibliography/41418013/public/?sort=date&direction=ascending