XIAOTONG YAO

+1(917)833-5729 ⋄ xiy2006@med.cornell.edu ⋄ xiaotong.yao23@gmail.com

Homepage: xtyao.github.io ⋄ ORCID: 0000-0002-5140-6639

1300 York Ave ⋄ New York, NY 10065

EDUCATION

Ph.D., Weill Cornell Medicine, Computational Biology Jul 2015 - expected **Apr 2021**

- Mentored by Imielinski, M: *Illuminating Rearranged Cancer Genome Structures throguh Genome Graphs*
- Courses: Optimization Methods, Data Structure and Algorithms, Poplulation Genetics, Statistical Genetics and Linear Models, Biomedical Machine Learning

M.S., New York University, Bioinformatics Aug 2013-May 2015
Overall GPA: 3.9/4.0 *New York, NY*

- Mentored by Vogel, C: *iSUMO - Integrative Prediction of Functionally Relevant SUMOylated Proteins*
- Courses: Linear Regression and Multivariate Analysis, Statistics in Biology, Bioinformatics and Genomes, Biological Databases and Data Mining, Proteomics Informatics

B.S., Zhejiang University, Biotechnology Aug 2009-May 2013
Overall GPA: 3.4/4.0 *New York, NY*

- Mentored by Chen, M: *Optimizaiton of Streptomycin Production in S. avermitilis by Metabolic Network Analysis*

PUBLICATIONS

Hadi K, **Yao X**, Behr JM, ..., Imielinski M. *Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs.* Cell. 2020;183: 197–210.e32.

(Accepted at Nature Communications) Dewhurst SM, **Yao X**, ..., de Lange T, Imielinski M. *Structural Variant Evolution after Telomere Crisis.* BioRxiv. 2020. p. 2020.09.29.318436. doi:10.1101/2020.09.29.318436

Carrot-Zhang J., **Yao X**, Devarakonda S., ..., TCGA Reasearch Network, Meyerson M, Govindan R, Imielinski M (2021). *Whole-genome Characterization of Lung Adenocarcinomas Lacking Alterations in the RTK/RAS/RAF Pathway.* Cell Rep. 34, 108707.

Wala JA, ..., Zhang C, Imielinski M, Beroukhim R. *SvABA: Genome-wide Detection of Structural Variants and Indels by Local Assembly.* Genome Res. 2018. doi:10.1101/gr.221028.117

Gerstung M, Jolly C, Leshchiner I, ..., PCAWG Consortium. *The Evolutionary History of 2,658 Cancers.* Nature. 2020;578: 122–128.

Rheinbay E, Nielsen MM, ..., PCAWG Consortium. *Analyses of Non-coding Somatic Drivers in 2,658 Cancer Whole Genomes.* Nature. 2020;578: 102–111.

Yao X, ..., Vogel C. *iSUMO - Integrative Prediction of Functionally Relevant SUMOylation Events.* bioRxiv. 2017. p. 056564. doi:10.1101/056564

(Submitted) Jungmin Choi, Aranzazu Manzano, Weilai Dong, ..., Richard P. Lifton, Joseph Schlessinger, and Alessandro D. Santin. Integrated mutational landscape analysis of uterine leiomyosarcomas.

RESEARCH PROJECTS

Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs May 2016 - Oct 2020
Senior leader: Imielinski M *New York, NY*

- Topology of junction copy number reveals novel classes of complex structural variants
- Rigma are deletion "chasms" at fragile sites arising early in GI tumor evolution
- Pyrgo are superenhancer-associated duplication "towers" in breast and ovarian cancer
- Tyfonas are "typhoons" of amplified fold-back inversions in acral melanoma

Illuminating Rearranged Genome Structure through gGnome Dec 2017 - Now
Senior leader: Imielinski M *New York, NY*

- Developed R API to genome graph data structures with a series of algorithms
- Designed an interactive web app to visualize genome graphs on arbitrary genomic regions
- Quick exploration of genome graphs built from 1000 Genome Project

Structural Variant Evolution after Telomere Crisis Dec 2019 - Sep 2020
Senior leader: de Lange T & Imielinski M *New York, NY*

- Screened more than a hundred shallow WGS for prevalent SV regions in clones of telomere crisis-surviving cells
- Reconstruct the exact linear allele after rearrangement with deep WGS
- Built consistent phylogeny using both SVs and SNVs
- Proved a single parental allele of chr12 to be the origin of SVs during crisis using allelic imbalances

Whole-genome Characterization of Lung Adenocarcinomas Lacking Alterations in RTK / RAS / RAF Pathways Dec 2017 - Nov 2020
Senior leader: Meyerson M, Govindan R, Imielinski M, TCGA genomic data analysis network

- Discovered *KRAS* or *RTK/RAS/RAF* alterations (RPA) from WGS previously missed by whole exome sequencing
- Found higher *TP53* loss of function frequency in RPA- cancers
- Delineated diverse complex structural variation mechanisms creating amplification of oncogenes

Integrated mutational landscape analysis of uterine leiomyosarcomas Dec 2018 - Dec 2020
Senior leader: Richard Lifton, Joseph Schlessinger, Alessandro Santin

- Identified the spectrum of complex SV events in uterine leiomyosarcomas whole genomes and found prevalent chromothripsis and chromoplexy, some disrupting tumor suppresors

Somatic Whole Genome Dynamics of Precancer in Barrett's Esophagus Mar 2018 - now
Senior leader: Patricia Galipeau, Brian Reid, Nicolas Robine, Marcin Imielinski

- Breakage-fusion-bridge cycles enrich in patient's Barrett's esophagus (BE) who eventually developed esophageal adenocarcinoma
- Rigma is an early, prevalent SV event type in BE regardless of cancer outcome

Structural Variation Signatures of Testicular Germ Cell Tumors Mar 2020 - now
Senior leader: Victoria Cortessis, Katherine Hoadley, TCGA genomic data analysis network

- Classified mechanisms amplifying *KRAS* in seminoma
- Discovered elevated burden of deletions and rigma in various non-seminoma

OPEN SOURCE SOFTWARE

Main author

- [JaBbA](#) - junction balance analysis
- [gGnome](#) - an R API to genome graphs

Contributor

- [gGnome.js](#) - an interactive web-based genome browser for genome graphs
- [gUtils](#) - elegant and fast genomic interval operations
- [gTrack](#) - static genome browser style plots
- [fishHook](#) - Gamma-Poisson regression for count data on genomic intervals
- [GxG](#) - interaction matrices between genomic bins

SKILLS

Biology	Cancer Genomics, Computational Biology, Systems Biology
Sequencing Informatics	Illumina, Oxford Nanopore WGS
Statistics & Machine Learning	Generalized Linear Models, Random Forests
Computer Languages	R, Shell, Python, Java
Optimization	CPLEX, Gurobi
Data Visualization	ggplot, shiny, plotly
Scientific Communication	LaTeX, Adobe Illustrator, (R)markdown
Databases	SQLite, MySQL
Other Tools	Git, Emacs, Docker, Nextflow, Hugo

WORK EXPERIENCE

Graduate Research Assistant   Jul 2015 - Present
Weill Cornell Medicine   *New York, NY*

- Discovered three new patterns of complex structural variations in cancer genomes
- Designed, implemented, benchmarked *Junction Balance Analysis* to reconstruct genome graphs
- Developed *gGnome* for genome graph data structure
- Captured ongoing SV evolution in post-telomere crisis cell lines
- Characterized whole genome landscape in lung adenocarcinomas without RTK/RAS/RAF pathways alterations
- Contributed to large cancer sequencing consortiums including The Cancer Genome Atlas (TCGA) and the Pan-Cancer Analysis of Whole Genomes (PCAWG)

Graduate Research Assistant   December 2013 - May 2015
New York University   *New York, NY*

- Trained predictive models for post-translational modification from protein function databases

Teaching Adjunct   Feb 2014 - May 2015
New York University   *New York, NY*

- Taught R statistical programming for Introduction to Biostatistics
- Tutored techniques for Biological Databases and Data Mining

Bioinformatics Intern   Jun 2014 - May 2015
3E Bioenergy   *New Brunswick, NJ*

- Compared crop genomes for candidate genes linked to drought resistance in sweet sorghum

AWARDS

Asia Regional Winner, World 2nd Runner Up & Best New Application   Nov 2011
2011 International Genetically Engineered Machines   *Hong Kong, China; Cambridge, MA*

- Team member of ZJU-China
- Designed multicolor fluorescent expression system in biofilm responsive to gradients of oxygen level

Master's Student Research Grant   2014 & 2015
NYU Biology Master's Program   *New York, NY*

- Funding for protien sumoylation prediction by mining public protein databases

Broad Institute Workshop Travel Grant   2015
Broad Institute   *Cambridge, MA*

- Travel grant for single cell genomics workshop

MEDIA

New Algorithm Reveals Patterns In How Tumors Extensively Rewrite Their DNA. Available: Weill Cornell Medicine News

Anderson A. *Algorithm Uncovers New Complex Rearrangement Classes in Cancer Genomes.* 1 Oct 2020. Available: Genomeweb.com