

# XIAOTONG YAO

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

Homepage: [xtyao.github.io](https://xtyao.github.io) ◇ ORCID: [0000-0002-5140-6639](https://orcid.org/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 through Genome Graphs*
- Courses: Optimization Methods, Data Structure and Algorithms, Population 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: *Optimization 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.

(In review) 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

(In review) Carrot-Zhang J, **Yao X**, Devarakonda S, et al. *Whole-genome Characterization of lung Adenocarcinomas lacking Alterations in RTK/RAS/RAF/MAPK Pathway*. Cancer Res. 2020;80: 5895–5895.

Wala JA, ..., Zhang C, Imielinski M, Beroukhi 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

## SKILLS

---

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

## 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

## WORK EXPERIENCE

---

### Graduate Research Assistant

*Weill Cornell Medicine*

Jul 2015 - Present

*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

*New York University*

December 2013 - May 2015

*New York, NY*

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

### Teaching Adjunct

*New York University*

Feb 2014 - May 2015

*New York, NY*

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

### Bioinformatics Intern

*3E Bioenergy*

Jun 2014 - May 2015

*New Brunswick, NJ*

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

## 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
- Pyrigo are superenhancer-associated duplication “towers” in breast and ovarian cancer
- Tyfonas are “typhoons” of amplified fold-back inversions in acral melanoma

### **Illuminating Rearranged Cancer 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
- Systematically captured complex coding and non-coding SV driver events in pan-cancer genomes

### **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 - Sep 2020

*Senior leader: Meyerson M, Govindan R, Imielinski M*

*New York, NY*

- Part of 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

## 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 protein 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](#)