Goo Jun

Ph.D.

☎ +1 (713) 500 9916
✉ Goo.Jun@uth.tmc.edu
Assistant Professor
Department of Epidemiology, Human Genetics and Environmental Sciences
School of Public Health
University of Texas Health Science Center Houston

Education

2010–2014 **Postdoctoral**, *University of Michigan*, Ann Arbor, MI.
Center for Statistical Genetics & Department of Biostatistics

2005–2010 **Ph.D.**, *The University of Texas*, Austin, TX.
Electrical and Computer Engineering
Dissertation: Transfer learning for classification of spatially varying data

1997–1999 **M.S.**, *University of Michigan*, Ann Arbor, MI.
Electrical Engineering:Systems

1993–1997 **B.S.**, *KAIST*, Daejeon, South Korea.
Electrical Engineering/Computer Science

Current and Past Positions

2014 – **Assistant Professor**, *University of Texas Health Science Center*, Houston, TX.
School of Public Health
Human Genetics Center
Department of Epidemiology, Human Genetics and Environmental Sciences

2010 – 2014 **Postdoctoral Research Fellow**, *University of Michigan*, Ann Arbor, MI.
Center for Statistical Genetics & Department of Biostatistics
PIs: Goncalo Abecasis and Michael Boehnke
Analyses of large-scale next-generation sequencing data
- The 1000 Genomes project
  `http://www.1000genomes.org`
- T2D-GENES: Type 2 diabetes genetic exploration by next-generation sequencing in multi-ethnic samples
- GO-ESP: NHLBI grand opportunity exome sequencing project
  `https://esp.gs.washington.edu/drupal/`.

2009 **Adv. Short-Term Research Opportunity(ASTRO)**, *Oak Ridge Nat. Lab.*, Oak Ridge, TN.
Computational Sciences and Engineering Division (CSED)
- Biomass monitoring project

2006 – 2010 **Research Assistant**, *University of Texas*, Austin, TX.
Advanced Data Exploration and Analysis Laboratory (IDEAL)
- NSF IIS-0705815 Advanced learning and integrative knowledge transfer approaches to remote sensing and forecast modeling for understanding land use change
- NSF IIS-0312471 Extraction and Interpretation of Information from Large-Scale Hyperspectral Data for Mapping and Monitoring Wetland Ecosystems

1999–2005 **Senior Engineer**, *Samsung Electronics*, Suwon, South Korea.
Digital Media R&D Center
Multimedia home network middleware, protocol, applications and devices.
DLNA standardization

1998 – 1999 **Research Asssistant**, *University of Michigan*, Ann Arbor, MI.
Biomechanics Laboratory

Research Support

Sep 2017 – Aug 2022 **1R01DK116378-01** , *NIH/NIDDK*.
Microbiome and Worsening Glycemia Among Mexican Americans in Starr County, Texas
Co-PI
The microbiome is a known source of metabolites and is critical in determining responses to diet. The need to understand these relationships is compelling for type 2 diabetes. While accumulating data show associations with diabetes, there is a paucity of data on the progression from normal glycemia to prediabetes and diabetes. The longitudinal study proposed here will allow evaluating relationships of the microbiome with prediabetes development and progression to type 2 diabetes while separating the confounding effects of seasonality and temporality associated with such factors as seasonal variation in food availability.

Sep 2017 – Aug 2019 **1R03DE026509**, *NIH/NIDCR*.
Identification of microRNAs associated with cleft lip
Co-PI
The etiology of cleft lip is complicated, with a variety of genetic and environmental factors involved. This study will identify the contribution of microRNAs (miRs), which are regulated by environmental factors, in lip development. The results of this study will facilitate our understanding of the role of miRs in lip development and will enable us to design future diagnostic and therapeutic approaches to prevent cleft lip.

Sep 2016 – Aug 2018 **1R03DE026208**, *NIH/NIDCR*.
Transcripts and Functions Targeted by Non-coding RNAs in Palate Development
Co-PI
The etiology of cleft palate is complicated by a variety of genetic and environmental factors. This study will identify the distribution and contribution of non-coding RNAs (ncRNAs) in palate development. The results of this study will facilitate our understanding of the role of ncRNAs in palate development and will enable us to design future therapeutic approaches to diagnose and prevent cleft palate.

Jan 2015 – Nov 2019 **UM1 HG008898**, *NIH/NHGRI*.
Genomic Architecture of Common Disease in Diverse Populations
Co-Investigator
This project aim to find genetic architectures of common diseases by performing deep whole-genome sequencing on 50,000 richly phenotyped individuals.

Publications

Journals

Rashkin S, Jun G, Chen S, Genetics, Epidemiology of Colorectal Cancer C, and Abecasis GR, 2017. Optimal sequencing strategies for identifying disease-associated singletons. *PLoS Genet*, 13(6):e1006811.

Manning A, Highland HM, Gasser J, Sim X, Tukiainen T, Fontanillas P, Grarup N, Rivas MA, Mahajan A, Locke AE, Cingolani P, Pers TH, Vinuela A, Brown AA, Wu Y, Flannick J, Fuchsberger C, Gamazon ER, Gaulton KJ, Im HK, Teslovich TM, Blackwell TW, Bork-Jensen J, Burtt NP, Chen Y, Green T, Hartl C, Kang HM, Kumar A, Ladenvall C, Ma C, Moutsianas L, Pearson RD, Perry JRB, Rayner NW, Robertson NR, Scott LJ, van de Bunt M, Eriksson JG, Jula A, Koskinen S, Lehtimaki T, Palotie A, Raitakari OT, Jacobs SBR, Wessel J, Chu AY, Scott RA, Goodarzi MO, Blancher C, Buck G, Buck D, Chines PS, Gabriel S, Gjesing AP, Groves CJ, Hollensted M, Huyghe JR, Jackson AU, Jun G, Justesen JM, Mangino M, Murphy J, Neville M, Onofrio R, Small KS, Stringham HM, Trakalo J, Banks E, Carey J, Carneiro MO,

DePristo M, Farjoun Y, Fennell T, Goldstein JI, Grant G, Hrabe de Angelis M, Maguire J, Neale BM, Poplin R, Purcell S, Schwarzmayr T, Shakir K, Smith JD, Strom TM, Wieland T, Lindstrom J, Brandslund I, Christensen C, Surdulescu GL, Lakka TA, Doney ASF, Nilsson P, Wareham NJ, Langenberg C, Varga TV, Franks PW, Rolandsson O, Rosengren AH, Farook VS, Thameem F, Puppala S, Kumar S, Lehman DM, Jenkinson CP, Curran JE, Hale DE, Fowler SP, Arya R, DeFronzo RA, Abboud HE, Syvanen AC, Hicks PJ, Palmer ND, Ng MCY, Bowden DW, Freedman BI, Esko T, Magi R, Milani L, Mihailov E, Metspalu A, Narisu N, Kinnunen L, Bonnycastle LL, Swift A, Pasko D, Wood AR, Fadista J, Pollin TI, Barzilai N, Atzmon G, Glaser B, Thorand B, Strauch K, Peters A, Roden M, Muller-Nurasyid M, Liang L, Kriebel J, Illig T, Grallert H, Gieger C, Meisinger C, Lannfelt L, Musani SK, Griswold M, Taylor J H. A., Wilson S G., Correa A, Oksa H, Scott WR, Afzal U, Tan ST, Loh M, Chambers JC, Sehmi J, Kooner JS, Lehne B, Cho YS, Lee JY, Han BG, Karajamaki A, Qi Q, Qi L, Huang J, Hu FB, Melander O, Orho-Melander M, Below JE, Aguilar D, Wong TY, Liu J, Khor CC, Chia KS, Lim WY, Cheng CY, Chan E, Tai ES, Aung T, Linneberg A, Isomaa B, Meitinger T, Tuomi T, Hakaste L, Kravic J, Jorgensen ME, Lauritzen T, Deloukas P, Stirrups KE, Owen KR, Farmer AJ, Frayling TM, O'Rahilly SP, Walker M, Levy JC, Hodgkiss D, Hattersley AT, Kuulasmaa T, Stancakova A, Barroso I, Bharadwaj D, Chan J, Chandak GR, Daly MJ, Donnelly PJ, Ebrahim SB, Elliott P, Fingerlin T, Froguel P, Hu C, Jia W, Ma RCW, McVean G, Park T, Prabhakaran D, Sandhu M, Scott J, Sladek R, Tandon N, Teo YY, Zeggini E, Watanabe RM, Koistinen HA, Kesaniemi YA, Uusitupa M, Spector TD, Salomaa V, Rauramaa R, Palmer CNA, Prokopenko I, Morris AD, Bergman RN, Collins FS, Lind L, Ingelsson E, Tuomilehto J, Karpe F, Groop L, Jorgensen T, Hansen T, Pedersen O, Kuusisto J, Abecasis G, Bell GI, Blangero J, Cox NJ, Duggirala R, Seielstad M, Wilson JG, Dupuis J, Ripatti S, Hanis CL, Florez JC, Mohlke KL, Meigs JB, Laakso M, Morris AP, Boehnke M, Altshuler D, McCarthy MI, Gloyn AL, and Lindgren CM, 2017. A low-frequency inactivating akt2 variant enriched in the finnish population is associated with fasting insulin levels and type 2 diabetes risk. *Diabetes*, 66(7):2019–2032.

Jiang ZD, Ajami NJ, Petrosino JF, Jun G, Hanis CL, Shah M, Hochman L, Ankoma-Sey V, DuPont AW, Wong MC, Alexander A, Ke S, and DuPont HL, 2017. Randomised clinical trial: faecal microbiota transplantation for recurrent clostridum difficile infection - fresh, or frozen, or lyophilised microbiota from a small pool of healthy donors delivered by colonoscopy. *Aliment Pharmacol Ther*, 45(7):899–908.

Hixson JE, Jun G, Shimmin LC, Wang Y, Yu G, Mao C, Warren AS, Howard TD, Heide RSV, Van Eyk J, Wang Y, and Herrington DM, 2017. Whole exome sequencing to identify genetic variants associated with raised atherosclerotic lesions in young persons. *Sci Rep*, 7(1):4091.

Metwalli KA, Do MA, Nguyen K, Mallick S, Kin K, Farokhnia N, Jun G, and Fakhouri WD, 2017. Interferon regulatory factor 6 is necessary for salivary glands and pancreas development. *J Dent Res*, :22034517729803.

Almeida M, Blondell L, Peralta JM, Kent J J. W., Jun G, Teslovich TM, Fuchsberger C, Wood AR, Manning AK, Frayling TM, Cingolani PE, Sladek R, Dyer TD, Abecasis G, Duggirala R, and Blangero J, 2016. Independent test assessment using the extreme value distribution theory. *BMC Proc*, 10(Suppl 7):245–249.

Blangero J, Teslovich TM, Sim X, Almeida MA, Jun G, Dyer TD, Johnson M, Peralta JM, Manning A, Wood AR, Fuchsberger C, Kent J J. W., Aguilar DA, Below JE, Farook VS, Arya R, Fowler S, Blackwell TW, Puppala S, Kumar S, Glahn DC, Moses EK, Curran JE, Thameem F, Jenkinson CP, DeFronzo RA, Lehman DM, Hanis C, Abecasis G, Boehnke M, Goring H, Duggirala R, Almasy L, and Consortium TDG, 2016. Omics-squared: human genomic,

transcriptomic and phenotypic data for genetic analysis workshop 19. *BMC Proc*, 10(Suppl 7):71–77.

Nicholson AM, Finch NA, Almeida M, Perkerson RB, van Blitterswijk M, Wojtas A, Cenik B, Rotondo S, Inskeep V, Almasy L, Dyer T, Peralta J, Jun G, Wood AR, Frayling TM, Fuchsberger C, Fowler S, Teslovich TM, Manning AK, Kumar S, Curran J, Lehman D, Abecasis G, Duggirala R, Pottier C, Zahir HA, Crook JE, Karydas A, Mitic L, Sun Y, Dickson DW, Bu G, Herz J, Yu G, Miller BL, Ferguson S, Petersen RC, Graff-Radford N, Blangero J, and Rademakers R, 2016. Prosaposin is a regulator of progranulin levels and oligomerization. *Nature Communications*, 7:11992.

Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, Ma C, Fontanillas P, Moutsianas L, McCarthy DJ, Rivas MA, Perry JRB, Sim X, Blackwell TW, Robertson NR, Rayner NW, Cingolani P, Locke AE, Tajes JF, Highland HM, Dupuis J, Chines PS, Lindgren CM, Hartl C, Jackson AU, Chen H, Huyghe JR, van de Bunt M, Pearson RD, Kumar A, Müller-Nurasyid M, Grarup N, Stringham HM, Gamazon ER, Lee J, Chen Y, Scott RA, Below JE, Chen P, Huang J, Go MJ, Stitzel ML, Pasko D, Parker SCJ, Varga TV, Green T, Beer NL, Day-Williams AG, Ferreira T, Fingerlin T, Horikoshi M, Hu C, Huh I, Ikram MK, Kim BJ, Kim Y, Kim YJ, Kwon MS, Lee J, Lee S, Lin KH, Maxwell TJ, Nagai Y, Wang X, Welch RP, Yoon J, Zhang W, Barzilai N, Voight BF, Han BG, Jenkinson CP, Kuulasmaa T, Kuusisto J, Manning A, Ng MCY, Palmer ND, Balkau B, Stančáková A, Abboud HE, Boeing H, Giedraitis V, Prabhakaran D, Gottesman O, Scott J, Carey J, Kwan P, Grant G, Smith JD, Neale BM, Purcell S, Butterworth AS, Howson JMM, Lee HM, Lu Y, Kwak SH, Zhao W, Danesh J, Lam VKL, Park KS, Saleheen D, So WY, Tam CHT, Afzal U, Aguilar D, Arya R, Aung T, Chan E, Navarro C, Cheng CY, Palli D, Correa A, Curran JE, Rybin D, Farook VS, Fowler SP, Freedman BI, Griswold M, Hale DE, Hicks PJ, Khor CC, Kumar S, Lehne B, Thuillier D, Lim WY, Liu J, van der Schouw YT, Loh M, Musani SK, Puppala S, Scott WR, Yengo L, Tan ST, Taylor Jr HA, Thameem F, Wilson G, Wong TY, Njølstad PR, Levy JC, Mangino M, Bonnycastle LL, Schwarzmayr T, Fadista J, Surdulescu GL, Herder C, Groves CJ, Wieland T, Bork-Jensen J, Brandslund I, Christensen C, Koistinen HA, Doney ASF, Kinnunen L, Esko T, Farmer AJ, Hakaste L, Hodgkiss D, Kravic J, Lyssenko V, Hollensted M, Jørgensen ME, Jørgensen T, Ladenvall C, Justesen JM, Käräjämäki A, Kriebel J, Rathmann W, Lannfelt L, Lauritzen T, Narisu N, Linneberg A, Melander O, Milani L, Neville M, Orho-Melander M, Qi L, Qi Q, Roden M, Rolandsson O, Swift A, Rosengren AH, Stirrups K, Wood AR, Mihailov E, Blancher C, Carneiro MO, Maguire J, Poplin R, Shakir K, Fennell T, DePristo M, Hrabéde Angelis M, Deloukas P, Gjesing AP, Jun G, Nilsson P, Murphy J, Onofrio R, Thorand B, Hansen T, Meisinger C, Hu FB, Isomaa B, Karpe F, Liang L, Peters A, Huth C, O'Rahilly SP, Palmer CNA, Pedersen O, Rauramaa R, Tuomilehto J, Salomaa V, Watanabe RM, Syvänen AC, Bergman RN, Bharadwaj D, Bottinger EP, Cho YS, Chandak GR, Chan JCN, Chia KS, Daly MJ, Ebrahim SB, Langenberg C, Elliott P, Jablonski KA, Lehman DM, Jia W, Ma RCW, Pollin TI, Sandhu M, Tandon N, Froguel P, Barroso I, Teo YY, Zeggini E, Loos RJF, Small KS, Ried JS, DeFronzo RA, Grallert H, Glaser B, Metspalu A, Wareham NJ, Walker M, Banks E, Gieger C, Ingelsson E, Im HK, Illig T, Franks PW, Buck G, Trakalo J, Buck D, Prokopenko I, Mägi R, Lind L, Farjoun Y, Owen KR, Gloyn AL, Strauch K, Tuomi T, Kooner JS, Lee JY, Park T, Donnelly P, Morris AD, Hattersley AT, Bowden DW, Collins FS, Atzmon G, Chambers JC, Spector TD, Laakso M, Strom TM, Bell GI, Blangero J, Duggirala R, Tai ES, McVean G, Hanis CL, Wilson JG, Seielstad M, Frayling TM, Meigs JB, Cox NJ, Sladek R, Lander ES, Gabriel S, Burtt NP, Mohlke KL, Meitinger T, Groop L, Abecasis G, Florez JC, Scott LJ, Morris AP, Kang HM, Boehnke M, Altshuler D, and McCarthy MI, 2016. The genetic architecture of type 2 diabetes. *Nature*, 536(7614):41–47.

Jun G, Wing MK, Abecasis GR, and Kang HM, 2015. An efficient and scalable analysis framework for variant extraction and refinement from population-scale dna sequence data. *Genome Res*, 25(6):918–25.

Flickinger M, Jun G, Abecasis GR, Boehnke M, and Kang HM, 2015. Correcting for sample contamination in genotype calling of dna sequence data. *The American Journal of Human Genetics*, 97(2):284–290.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MHY, et al., 2015. An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571):75–81.

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. *Nature*, 526(7571):68–74.

Williams AL, Genovese G, Dyer T, Altemose N, Truax K, Jun G, Patterson N, Myers SR, Curran JE, Duggirala R, Blangero J, Reich D, and Przeworski M, 2015. Non-crossover gene conversions show strong gc bias and unexpected clustering in humans. *eLife*, 4:e04637.

Mahajan A, Sim X, Ng HJ, Manning A, Rivas MA, Highland HM, Locke AE, Grarup N, Im HK, Cingolani P, Flannick J, Fontanillas P, Fuchsberger C, Gaulton KJ, Teslovich TM, Rayner NW, Robertson NR, Beer NL, Rundle JK, Bork-Jensen J, Ladenvall C, Blancher C, Buck D, Buck G, Burtt NP, Gabriel S, Gjesing AP, Groves CJ, Hollensted M, Huyghe J, Jackson AU, Jun G, et al., 2015. Identification and functional characterization of g6pc2 coding variants influencing glycemic traits define an effector transcript at the g6pc2-abcb11 locus. *PLoS genetics*, 11(1):e1004876–e1004876.

Wood AR, Tuke MA, Nalls M, Hernandez D, Gibbs JR, Lin H, Xu CS, Li Q, Shen J, Jun G, Almeida M, Tanaka T, Perry JRB, Gaulton K, Rivas M, Pearson R, Curran JE, Johnson MP, Goring HHH, Duggirala R, Blangero J, Mccarthy MI, Bandinelli S, Murray A, Weedon MN, Singleton A, Melzer D, Ferrucci L, and Frayling TM, 2015. Whole-genome sequencing to understand the genetic architecture of common gene expression and biomarker phenotypes. *Human Molecular Genetics*, 24(5):1504–1512.

Vrieze SI, Malone SM, Vaidyanathan U, Kwong A, Kang HM, Zhan X, Flickinger M, Irons D, Jun G, Locke AE, Pistis G, Porcu E, Levy S, Myers RM, Oetting W, McGue M, Abecasis G, and Iacono WG, 2014. In search of rare variants: Preliminary results from whole genome sequencing of 1, 325 individuals with psychophysiological endophenotypes. *Psychophysiology*, 51(12):1309–1320.

Almasy L, Dyer TD, Peralta JM, Jun G, Wood AR, Fuchsberger C, Almeida MA, Kent JW, Fowler S, Blackwell TW, Puppala S, Kumar S, Curran JE, Lehman D, Abecasis G, Duggirala R, and Blangero J, 2014. Data for genetic analysis workshop 18: human whole genome sequence, blood pressure, and simulated phenotypes in extended pedigrees. *BMC Proceedings*, 8(Suppl 1):S2.

Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, Lange LA, Lu Y, Tang Zz, Zhang H, Hindy G, Masca N, Stirrups K, Kanoni S, Do R, Jun G, Hu Y, Kang HM, Xue C, Goel A, Farrall M, Duga S, Merlini PA, Asselta R, Girelli D, Olivieri O, Martinelli N, Yin W, Reilly D, Speliotes E, Fox CS, Hveem K, Holmen OL, Nikpay M, NHLBI Exome Sequencing Project, Farlow DN, Assimes TL, Franceschini N, Robinson J, North KE, Martin LW, DePristo M, Gupta N, Escher SA, Jansson JH, Van Zuydam N, Palmer CN, Wareham N, Koch W, Meitinger T, Peters A, Lieb W, Erbel R, Konig IR, Kruppa J, Degenhardt F, Gottesman O, Bottinger EP, O'Donnell CJ,

Psaty BM, Ballantyne CM, Abecasis G, Ordovas JM, Melander O, Watkins H, Orho-Melander M, Ardissino D, Loos RJ, McPherson R, Willer CJ, Erdmann J, Hall AS, Samani NJ, Deloukas P, Schunkert H, Wilson JG, Kooperberg C, Rich SS, Tracy RP, Lin DY, Altshuler D, Gabriel S, Nickerson DA, Jarvik GP, Cupples LA, Reiner AP, Boerwinkle E, and Kathiresan S, 2014. Loss-of-function mutations in the apoc3 gene, plasma triglycerides, and risk for coronary heart disease. *New England Journal of Medicine*, 371(1):22–31.

Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, Bizon C, Lange EM, Smith JD, Turner EH, Jun G, Kang HM, Peloso G, Auer P, ping Li K, Flannick J, Zhang J, Fuchsberger C, Gaulton K, Lindgren C, Locke A, Manning A, Sim X, Rivas MA, Holmen OL, Gottesman O, Lu Y, Ruderfer D, Stahl EA, Duan Q, Li Y, Durda P, Jiao S, Isaacs A, Hofman A, Bis JC, Correa A, Griswold ME, Jakobsdottir J, Smith AV, Schreiner PJ, Feitosa MF, Zhang Q, Huffman JE, Crosby J, Wassel CL, Do R, Franceschini N, Martin LW, Robinson JG, Assimes TL, Crosslin DR, Rosenthal EA, Tsai M, Rieder MJ, Farlow DN, Folsom AR, Lumley T, Fox ER, Carlson CS, Peters U, Jackson RD, van Duijn CM, Uitterlinden AG, Levy D, Rotter JI, Taylor HA, Jr. VG, Siscovick DS, Fornage M, Borecki IB, Hayward C, Rudan I, Chen YE, Bottinger EP, Loos RJ, Sætrom P, Hveem K, Boehnke M, Groop L, McCarthy M, Meitinger T, Ballantyne CM, Gabriel SB, O'Donnell CJ, Post WS, North KE, Reiner AP, Boerwinkle E, Psaty BM, Altshuler D, Kathiresan S, Lin DY, Jarvik GP, Cupples LA, Kooperberg C, Wilson JG, Nickerson DA, Abecasis GR, Rich SS, Tracy RP, and Willer CJ, 2014. Whole-exome sequencing identifies rare and low-frequency coding variants associated with ldl cholesterol. *The American Journal of Human Genetics*, 94(2):233 – 245.

Duan Q, Liu EY, Auer PL, Zhang G, Lange EM, Jun G, Bizon C, Jiao S, Buyske S, Franceschini N, et al., 2013. Imputation of coding variants in african americans: better performance using data from the exome sequencing project. *Bioinformatics*, .

Zhan X, Larson DE, Wang C, Koboldt DC, Sergeev YV, Fulton RS, Fulton LL, Fronick CC, Branham KE, Bragg-Gresham J, Jun G, Hu Y, Kang HM, Liu D, Othman M, Brooks M, Ratnapriya R, Boleda A, Grassmann F, von Strachwitz C, Olson LM, Buitendijk GHS, Hofman A, van Duijn CM, Cipriani V, Moore AT, Shahid H, Jiang Y, Conley YP, Morgan DJ, Kim IK, Johnson MP, Cantsilieris S, Richardson AJ, Guymer RH, Luo H, Ouyang H, Licht C, Pluthero FG, Zhang MM, Zhang K, Baird PN, Blangero J, Klein ML, Farrer LA, DeAngelis MM, Weeks DE, Gorin MB, Yates JRW, Klaver CCW, Pericak-Vance MA, Haines JL, Weber BHF, Wilson RK, Heckenlively JR, Chew EY, Stambolian D, Mardis ER, Swaroop A, and Abecasis GR, 2013. Identification of a rare coding variant in complement 3 associated with age-related macular degeneration. *Nature Genetics*, .

Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Gabriel S, Shendure J, Nickerson DA, Bamshad M, and Akey JM, 2013. Analysis of 6,515 exomes reveals a very recent origin of most human protein-coding variants. *Nature*, 493(7431):216–220.

Jun G and Ghosh J, 2013. Semisupervised learning of hyperspectral data with unknown land-cover classes. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(1):273–282.

Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M, and Kang HM, 2012. Detecting and estimating contamination of human dna samples in sequencing and array-based genotype data. *American journal of human genetics*, 91(5):839–848.

The 1000 Genomes Consortium, 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM, GO B, GO S, and on behalf of the NHLBI Exome Sequencing Project, 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, 337(6090):64–69.

Jun G and Ghosh J, 2011. Spatially adaptive classification of land cover with remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7):2662 –2673.

Jun G and Ghosh J, 2011. Spatially adaptive semi-supervised learning with gaussian processes for hyperspectral data analysis. *Statistical Analysis and Data Mining*, 4(4):358–371.

Liu A, Jun G, and Ghosh J, 2009. A self-training approach to cost sensitive uncertainty sampling. *Machine Learning*, 76(2-3):257–270.

Peer-reviewed Conference Proceedings

Symons C, Vatsavai R, Jun G, and Arel I, 2012. Bias selection using task-targeted random subspaces for robust application of graph-based semi-supervised learning. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 1, pages 415–420.

Vatsavai RR, Symons CT, Chandola V, and Jun G, 2011. Gx-means: A model-based divide and merge algorithm for geospatial image clustering. In *Procedia Computer Science 4, Proceedings of the International Conference on Computational Science (ICCS 2011)*, volume 4, pages 186 – 195.

Jun G and Ghosh J, 2010. Nearest-manifold classification with gaussian processes. In *20th International Conference on Pattern Recognition (ICPR '10)*.

Jun G, Ghosh J, Radosavljevic V, and Obradovic Z, 2010. Predicting ground-based aerosol optical depth with satellite images via guassian processes. In *Proceedings of the 2010 International Conference on Knowledge Discovery and Information Retrieval (KDIR2010)*.

Jun G, Vatsavai RR, and Ghosh J, 2009. Spatially adaptive classification and active learning of multispectral data with gaussian processes. In *Proc. ICDM Workshop on Spatial and Spatio-temporal Data Mining (SSTDM/ICDMW09)*.

Jun G and Ghosh J, 2009. Spatially adaptive classification of hyperspectral data with gaussian processes. In *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS 09)*.

Liu A, Jun G, and Ghosh J, 2009. Active learning of hyperspectral data with spatially dependent label acquisition costs. In *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS09)*.

Liu A, Jun G, and Ghosh J, 2009. Spatially cost-sensitive active learning. In *Proc. SIAM International Conference on Data Mining (SDM09)*.

Jun G and Ghosh J, 2009. Hybrid hierarchical classifiers for hyperspectral data analysis. In Benediktsson JA, Kittler J, and Roli F, editors, *Multiple Classifier Systems*. LNCS, Springer.

Jun G and Ghosh J, 2009. Multi-class boosting with class hierarchies. In Benediktsson JA, Kittler J, and Roli F, editors, *Multiple Classifier Systems*. LNCS, Springer.

Jun G and Ghosh J, 2008. An efficient active learning algorithm with knowledge transfer for hyperspectral remote sensing data. In *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS 08)*.

Jun G, Aggarwal J, and Gokmen M, 2008. Tracking and segmentation of highway vehicles in cluttered and crowded scenes. In *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*, pages 1–6.

Jun G, 2005. Home media center and media clients for multi-room audio and video applications. In *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pages 257–260.

Recent Conference Presentations

Talks rogression and Subtypes of Prediabetes with Metabolomics Profiling in Starr County Mexican Americans, *77th Scientific Sessions, American Diabetes Association*, Young Investigator Travel Grant, San Diego, CA, USA, 2017.

Talks Deep whole-genome sequencing in pedigrees to quantify the contribution of private variants to type 2 diabetes and related metabolic traits, *63rd Annual Meeting of American Society of Human Genetics (ASHG)*, Boston, MA, USA, 2013.

Detecting Functional Rare Variants Relating to Type 2 Diabetes Using Deep Whole Genome Sequencing, *72nd Scientific Sessions, American Diabetes Association*, Young Investigator Travel Grant, Philadelphia, PA, USA, 2012.

Joint variant calling and analysis across >4,000 exomes of European and African American ancestry, *International Congress of Human Genetics (ICHG/ASHG)*, Montreal, Canada, 2011.

Posters MultiVAC: Scalable multi-sample variant caller with local de novo assembly, *2016 CSHL Biology of Genomes*, Cold Spring Harbor, NY, USA.

IGAP: Integrated Genetic Analysis Platform for Web-based Interactive Association Analysis and Visualization of Large-scale Genotype/Phenotype Data, *The Allied Genetics Conference 2016*, Orlando, FL, USA.

Robust Multi-sample Calling of Structural Variations from Complete Genomics Sequencing Data, *2014 CSHL Biology of Genomes*, Cold Spring Harbor, NY, USA.

Analyzing Deep Whole Genome Sequence and Genotype Data of >1,000 Individuals from Large Mexican-American Pedigrees in the T2D-GENES Study, *American Society of Human Genetics 2012*, San Francisco, CA, USA.

A profile of DNA variation across 2,500 exomes of European and African American ancestry, *2011 CSHL Biology of Genomes*, Cold Spring Harbor, NY, USA.

Lectures/Invited Talks

June2014 **Lecture**, *Sequence Analysis Workshop (SeqShop)*, University of Michigan, Ann Arbor, Michigan.
- Variant calling and filtering for SNPs

June 2013 **Invited Talk**, *The 4th International Conference for Statistical Methods in the Biomedical Science and Bioinformatics*, Choong-Ang University, Seoul, Korea.
- Methods for modeling and correcting sample contamination in next-generation sequencing data

March 2013 **Lecture**, *International Workshop on Statistical Genetic Methods for Human Complex Traits*, University of Colorado, Boulder, Colorado.
- Sequencing: Mapping, alignment, and variant calling with practical
- Sequencing: Visualization and QC of variant calls

Softwares

verifyBamID — Detecting and estimating sample contaminations in sequencing data
`http://genome.sph.umich.edu/wiki/VerifyBamID`

verifyIDintensity — Detecting and estimating sample contamination in genotype array data
`http://genome.sph.umich.edu/wiki/VerifyIDintensity`

gotCloud — Software pipeline for SNP detection and refinement
`http://genome.sph.umich.edu/wiki/GotCloud`

Patents

Issued — US Patent 9,342,141 and 8,504,637, KR Patent 1005773620000, Audio/video device, apparatus and method for controlling audio/video device

US Patent 8,260,843, KR Patent 1007728610000, Apparatus and method for providing remote user interface

US Patent 8,196,170, KR Patent 1007556950000, Method of connecting to internet via broadcast receiving device and apparatus for the same

KR Patent 1004777050000, Method for displaying main screen of home media center

KR Patent 1005974040000, Method and apparatus for transferring home media center's media contents apart from home media center

KR Patent 1007886520000, Apparatus and method for dialing auto sound

Professional Activities

Program Committees and Reviewer Services

Reviewer — Bioinformatics

PLoS Computational Biology

BMC Bioinformatics

Genome Research

IEEE Transactions on Neural Network and Learning Systems

IEEE Transactions on Image Processing

Patter Recognition

RemoteSensing

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

IEEE Geoscience and Remote Sensing Letters

Program Committee — International Workshop on Spatial and Spatiotemporal Data Mining (SSTDM), held in cooperation with IEEE International Conference on Data Mining (ICDM)

International Workshop on Knowledge Discovery from Sensor Data (SensorKDD), held in conjunction with the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2011

Professional Membership

American Society of Human Genetics (ASHG)

Institute of Electrical and Electronic Engineers (IEEE)