We have extensive experience in large-scale variant calling and interpretation: This experience is through our membership in the 1000 Genomes Consortium, particularly from our participation in the analysis working group and the structural variant (SV) and functional interpretation (FIG) subgroups of the consortium, where the majority of the variant calling tools were developed, deployed and interpreted [1]–[3].

Structural variations (SVs) are important contributors to human polymorphism, have great functional impact and are often implicated in diseases including cancer. We have developed a number of SV calling algorithms, including BreakSeq which compares raw reads with a breakpoint library (junction mapping) [4], CNVnator which measures read depth [5], AGE which refines local alignment [6], and PEMer which uses paired ends [7]. We have also developed array-based approaches [8] and a sequencing-based Bayesian model [9]. Furthermore, we have studied the distinct features of SVs that originate from different mechanisms, and showed how creation processes may have potentially divergent functional impacts [10], [11].

*Read depth based identification of copy number variation by CNVnator.*

We have developed ways of prioritizing high-functional impact variants in coding regions: We have developed a number of software tools for evaluating the impact of non-synonymous SNVs within coding regions of the genome. These tools utilize VAT [12], our software for variant annotation, and take as input X-ray crystallographic models of protein structures downloaded from the Protein Data Bank (PDB).

One tool within our newly-built software suite named STRESS, employs models of conformational change to predict allosteric residues by finding essential surface pockets and information-flow bottlenecks [13]. This tool is computationally fast, thereby enabling its implementation across the entire repertoire of coding elements for which crystallographic data is available. STRESS captures conserved regions, and has also been used to rationalize known disease-associated SNVs. A second tool within our software suite, uses localized structural frustration [14] to identify SNVs for which local perturbations may severely impact protein functionality without strongly disrupting global stability [15]. We have previously leveraged this method to elucidate the differential impacts of somatic SNVs associated with oncogenes and tumor suppressor genes (TSGs). As a third component of this software suite, we are now submitting work which describes our newly-developed ALoFT software, that integrates multiple features in order to predict and quantify the impact of potential LOF variants in protein-coding genes.

We have developed ways of prioritizing high-functional impact variants in non-coding regions: We have completed extensive analysis of patterns of variation in non-coding regions, and their coding targets [16]–[18]. We used metrics such as diversity, and fraction of rare variants, to characterize selection on various classes and subclasses of functional annotations [16], [19]. In addition, we have also defined variants that are disruptive to a TF-binding motif in a regulatory region [20]. Further studies showed relationships between selection and protein network topology (for instance, quantifying selection in hubs relative to proteins on the network periphery) [21], [22]. In recent studies [10], we have integrated and extended these methods to develop a prioritization pipeline called FunSeq [23]. It identifies sensitive and ultra-sensitive regions -- those annotations under strong selective pressure, as determined using genomes from many individuals from diverse populations. It identifies deleterious variants in many non-coding functional elements, including TF binding sites, enhancer elements, and regions of open chromatin corresponding to DNase I hypersensitive sites. It also detects their specific disruptiveness to TF binding sites, annotating both loss-of (LOF) and gain-of function (GOF) events. Integrating large-scale data from various resources (including ENCODE and the 1000 Genomes Project) with cancer genomics data, FunSeq is able to prioritize known TERT promoter driver mutations, and score recurrent somatic mutations higher than non-recurrent mutations. Using FunSeq, we identified ~100 non-coding candidate drivers in ~90 WGS medulloblastoma, breast, and prostate cancer samples [10]. Drawing on this experience, we are currently co-leading the International Cancer Genomics Consortium (ICGC) pan-cancer analysis-working group (PCAWG)-2 (analysis of mutations in regulatory regions) group.

*Workflow for Funseq based variant prioritization.*

We have also used allelic variability to prioritize regions of the genome. That is, we prioritize regions that differ in functional genomic response. This includes, allele-specific expression and binding between the maternal and paternal alleles. Our variant analysis work includes AlleleSeq [24], a computational pipeline to identify allele-specific events, and AlleleDB, our database connecting single nucleotide variants with allele-specific binding and expression [25].

We have developed tools for somatic and germline burden tests: We have worked on statistical methods for analysis of non-coding regulatory regions. LARVA (Large-scale Analysis of Recurrent Variants in noncoding Annotations) identifies significant mutation enrichments in noncoding elements, by comparing observed mutation counts with expected counts under a whole genome background mutation model [26]. LARVA includes corrections for biases in mutation rate owing to DNA replication timing. LARVA can be targeted to coding regions to prioritize genes. We used this tool in a pan-cancer analysis of variants in 760 cancer whole genomes, spanning a number of cancer data portals and published datasets. Our analyses demonstrated that LARVA can recapitulate previously established coding and noncoding cancer drivers, including the TERT and TP53 promoters [26]. Furthermore, we have developed MOAT (Mutations Overburdening Annotations Tool), an alternative empirical mutation burden approach that evaluates mutation enrichments based upon permutations of the input data (submitted). Both annotation-based and variant-based permutation is supported.

*Comparison of β-binomial distribution fit (turquoise) and binomial distribution fit (pink) to observed cancer somatic mutation counts. The β-binomial distribution betters models the empirical distribution’s (black) overdispersion.*

We have identified regions associated with kidney cancer through our involvement in the papillary TCGA team: Related to Yale’s expertise in the clinical management and genetics of kidney cancer, we were invited to participate in TCGA kidney cancer projects. Our role in the TCGA KICH (chromophobe RCC) included coordination of the Cancer Cell manuscript. Our team performed the analysis of the whole genome sequencing for the TCGA KIRP (pRCC), now published in New England Journal of Medicine [27]. In recent work, we leverage our expertise of non-coding region in the first whole genome analysis of pRCC (under revision). Our work finds significant genomic alterations beyond traditional known drivers of pRCC. We hypothesize these alterations may have non-canonical effect on known tumorigenesis pathways (for example, MET in type 1 pRCC). We discovered genomic markers in MET and NEAT1 that predict prognosis. Last, we investigate mutational signatures and mutational landscape in pRCC and pin down several meaningful etiological factors explaining inter-patient genomic variation in pRCC. This provided further experience with available RCC genomic datasets. Finally, our team has participated in two ongoing pan-RCC manuscripts serving a central role assessing evaluating the cluster of cluster assignments (COCA) immunologic profile from gene and miRNA expression datasets. Together with other published results on RCC [28]–[32], we have assembled an extensive list of impactful and statistically significant regions of RCC genomes.

References:

[1] R. E. Mills, K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheetham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. K. Lam, J. Leng, R. Li, Y. Li, C.-Y. Lin, R. Luo, X. J. Mu, J. Nemesh, H. E. Peckham, T. Rausch, A. Scally, X. Shi, M. P. Stromberg, A. M. Stütz, A. E. Urban, J. A. Walker, J. Wu, Y. Zhang, Z. D. Zhang, M. A. Batzer, L. Ding, G. T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, K. Ye, E. E. Eichler, M. B. Gerstein, M. E. Hurles, C. Lee, S. A. McCarroll, J. O. Korbel, and 1000 Genomes Project, “Mapping copy number variation by population-scale genome sequencing.,” *Nature*, vol. 470, no. 7332, pp. 59–65, Feb. 2011.

[2] 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, and G. A. McVean, “A map of human genome variation from population-scale sequencing.,” *Nature*, vol. 467, no. 7319, pp. 1061–73, Oct. 2010.

[3] G. Project Consortium, G. Consortium Participants are arranged by project role, then by institution alphabetically, finally alphabetically within institutions except for Principal Investigators, P. Leaders, as indicated, C. author, S. committee, P. group, B. College of Medicine, B. Institute of MIT, E. Bioinformatics Institute, M. Planck Institute for Molecular Genetics, U. National Institutes of Health, U. of Oxford, W. University in St Louis, W. Trust Sanger Institute, A. group, A. Einstein College of Medicine, B. College, W. Hospital, C. Spring Harbor Laboratory, D. University, E. Molecular Biology Laboratory, C. University, H. University, H. Gene Mutation Database, L. University Medical Center, L. State University, M. General Hospital, P. State University, S. University, T.-A. University, T. Genomics Research Institute, U. of California, S. Diego, S. Francisco, S. Cruz, U. of Chicago, U. College London, U. of Geneva, U. of Maryland School of Medicine, U. of Medicine, D. of New Jersey, U. of Michigan, U. of Montréal, U. of Puerto Rico, U. of Texas Health Sciences Center at Houston, U. of Utah, U. of Washington, Y. University, S. variation group, E. group, iversity of Geneva, F. interpretation group, D. coordination centre group, E. group, S. collection, B. from England, C. in Medellín, H. Chinese South, F. in Finland, I. populations in Spain, P. Ricans in Puerto Rico, S. management, and W. group, “An integrated map of genetic variation from 1,092 human genomes,” *Nature*, vol. 490, 2012.

[4] H. Y. K. Lam, X. J. Mu, A. M. Stütz, A. Tanzer, P. D. Cayting, M. Snyder, P. M. Kim, J. O. Korbel, and M. B. Gerstein, “Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library.,” *Nat. Biotechnol.*, vol. 28, no. 1, pp. 47–55, Jan. 2010.

[5] A. Abyzov, A. E. Urban, M. Snyder, and M. Gerstein, “CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing.,” *Genome Res.*, vol. 21, no. 6, pp. 974–84, Jun. 2011.

[6] A. Abyzov and M. Gerstein, “AGE: defining breakpoints of genomic structural variants at single-nucleotide resolution, through optimal alignments with gap excision.,” *Bioinformatics*, vol. 27, no. 5, pp. 595–603, Mar. 2011.

[7] J. O. Korbel, A. Abyzov, X. J. Mu, N. Carriero, P. Cayting, Z. Zhang, M. Snyder, and M. B. Gerstein, “PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data.,” *Genome Biol.*, vol. 10, no. 2, p. R23, Feb. 2009.

[8] L.-Y. Wang, A. Abyzov, J. O. Korbel, M. Snyder, and M. Gerstein, “MSB: a mean-shift-based approach for the analysis of structural variation in the genome.,” *Genome Res.*, vol. 19, no. 1, pp. 106–17, Jan. 2009.

[9] Z. D. Zhang and M. B. Gerstein, “Detection of copy number variation from array intensity and sequencing read depth using a stepwise Bayesian model.,” *BMC Bioinformatics*, vol. 11, p. 539, Oct. 2010.

[10] E. Khurana, Y. Fu, V. Colonna, X. J. Mu, H. M. Kang, T. Lappalainen, A. Sboner, L. Lochovsky, J. Chen, A. Harmanci, J. Das, A. Abyzov, S. Balasubramanian, K. Beal, D. Chakravarty, D. Challis, Y. Chen, D. Clarke, L. Clarke, F. Cunningham, U. S. Evani, P. Flicek, R. Fragoza, E. Garrison, R. Gibbs, Z. H. Gümüs, J. Herrero, N. Kitabayashi, Y. Kong, K. Lage, V. Liluashvili, S. M. Lipkin, D. G. MacArthur, G. Marth, D. Muzny, T. H. Pers, G. R. S. Ritchie, J. A. Rosenfeld, C. Sisu, X. Wei, M. Wilson, Y. Xue, F. Yu, 1000 Genomes Project Consortium, E. T. Dermitzakis, H. Yu, M. A. Rubin, C. Tyler-Smith, and M. Gerstein, “Integrative annotation of variants from 1092 humans: application to cancer genomics.,” *Science*, vol. 342, no. 6154, p. 1235587, Oct. 2013.

[11] A. Abyzov, S. Li, D. R. Kim, M. Mohiyuddin, A. M. Stütz, N. F. Parrish, X. J. Mu, W. Clark, K. Chen, M. Hurles, J. O. Korbel, H. Y. K. Lam, C. Lee, and M. B. Gerstein, “Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms.,” *Nat. Commun.*, vol. 6, p. 7256, Jun. 2015.

[12] L. Habegger, S. Balasubramanian, D. Z. Chen, E. Khurana, A. Sboner, A. Harmanci, J. Rozowsky, D. Clarke, M. Snyder, and M. Gerstein, “VAT: a computational framework to functionally annotate variants in personal genomes within a cloud-computing environment,” *Bioinformatics*, vol. 28, no. 17, pp. 2267–2269, Sep. 2012.

[13] D. Clarke, A. Sethi, S. Li, S. Kumar, R. W. F. Chang, J. Chen, and M. Gerstein, “Identifying Allosteric Hotspots with Dynamics: Application to Inter- and Intra-species Conservation.,” *Structure*, vol. 24, no. 5, pp. 826–37, May 2016.

[14] D. U. Ferreiro, E. A. Komives, and P. G. Wolynes, “Frustration in biomolecules.,” *Q. Rev. Biophys.*, vol. 47, no. 4, pp. 285–363, Nov. 2014.

[15] S. Kumar, D. Clarke, and M. Gerstein, “Localized structural frustration for evaluating the impact of sequence variants,” *Nucleic Acids Res.*, p. gkw927, Oct. 2016.

[16] X. J. Mu, Z. J. Lu, Y. Kong, H. Y. K. Lam, and M. B. Gerstein, “Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project.,” *Nucleic Acids Res.*, vol. 39, no. 16, pp. 7058–76, Sep. 2011.

[17] K. Y. Yip, C. Cheng, N. Bhardwaj, J. B. Brown, J. Leng, A. Kundaje, J. Rozowsky, E. Birney, P. Bickel, M. Snyder, and M. Gerstein, “Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors.,” *Genome Biol.*, vol. 13, no. 9, p. R48, Sep. 2012.

[18] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O’Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder, “Architecture of the human regulatory network derived from ENCODE data.,” *Nature*, vol. 489, no. 7414, pp. 91–100, Sep. 2012.

[19] E. Khurana, Y. Fu, D. Chakravarty, F. Demichelis, M. A. Rubin, and M. Gerstein, “Role of non-coding sequence variants in cancer.,” *Nat. Rev. Genet.*, vol. 17, no. 2, pp. 93–108, Feb. 2016.

[20] ENCODE Project Consortium, “An integrated encyclopedia of DNA elements in the human genome.,” *Nature*, vol. 489, no. 7414, pp. 57–74, Sep. 2012.

[21] P. M. Kim, J. O. Korbel, and M. B. Gerstein, “Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 51, pp. 20274–9, Dec. 2007.

[22] E. Khurana, Y. Fu, J. Chen, and M. Gerstein, “Interpretation of genomic variants using a unified biological network approach.,” *PLoS Comput. Biol.*, vol. 9, no. 3, p. e1002886, 2013.

[23] Y. Fu, Z. Liu, S. Lou, J. Bedford, X. J. Mu, K. Y. Yip, E. Khurana, and M. Gerstein, “FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer.,” *Genome Biol.*, vol. 15, no. 10, p. 480, 2014.

[24] J. Rozowsky, A. Abyzov, J. Wang, P. Alves, D. Raha, A. Harmanci, J. Leng, R. Bjornson, Y. Kong, N. Kitabayashi, N. Bhardwaj, M. Rubin, M. Snyder, and M. Gerstein, “AlleleSeq: analysis of allele-specific expression and binding in a network framework.,” *Mol. Syst. Biol.*, vol. 7, p. 522, Aug. 2011.

[25] J. Chen, J. Rozowsky, T. R. Galeev, A. Harmanci, R. Kitchen, J. Bedford, A. Abyzov, Y. Kong, L. Regan, and M. Gerstein, “A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals.,” *Nat. Commun.*, vol. 7, p. 11101, Apr. 2016.

[26] L. Lochovsky, J. Zhang, Y. Fu, E. Khurana, and M. Gerstein, “LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations.,” *Nucleic Acids Res.*, vol. 43, no. 17, pp. 8123–34, Sep. 2015.

[27] Cancer Genome Atlas Research Network, W. M. Linehan, P. T. Spellman, C. J. Ricketts, C. J. Creighton, S. S. Fei, C. Davis, D. A. Wheeler, B. A. Murray, L. Schmidt, C. D. Vocke, M. Peto, A. A. M. Al Mamun, E. Shinbrot, A. Sethi, S. Brooks, W. K. Rathmell, A. N. Brooks, K. A. Hoadley, A. G. Robertson, D. Brooks, R. Bowlby, S. Sadeghi, H. Shen, D. J. Weisenberger, M. Bootwalla, S. B. Baylin, P. W. Laird, A. D. Cherniack, G. Saksena, S. Haake, J. Li, H. Liang, Y. Lu, G. B. Mills, R. Akbani, M. D. M. Leiserson, B. J. Raphael, P. Anur, D. Bottaro, L. Albiges, N. Barnabas, T. K. Choueiri, B. Czerniak, A. K. Godwin, A. A. Hakimi, T. H. Ho, J. Hsieh, M. Ittmann, W. Y. Kim, B. Krishnan, M. J. Merino, K. R. Mills Shaw, V. E. Reuter, E. Reznik, C. S. Shelley, B. Shuch, S. Signoretti, R. Srinivasan, P. Tamboli, G. Thomas, S. Tickoo, K. Burnett, D. Crain, J. Gardner, K. Lau, D. Mallery, S. Morris, J. D. Paulauskis, R. J. Penny, C. Shelton, W. T. Shelton, M. Sherman, E. Thompson, P. Yena, M. T. Avedon, J. Bowen, J. M. Gastier-Foster, M. Gerken, K. M. Leraas, T. M. Lichtenberg, N. C. Ramirez, T. Santos, L. Wise, E. Zmuda, J. A. Demchok, I. Felau, C. M. Hutter, M. Sheth, H. J. Sofia, R. Tarnuzzer, Z. Wang, L. Yang, J. C. Zenklusen, J. Zhang, B. Ayala, J. Baboud, S. Chudamani, J. Liu, L. Lolla, R. Naresh, T. Pihl, Q. Sun, Y. Wan, Y. Wu, A. Ally, M. Balasundaram, S. Balu, R. Beroukhim, T. Bodenheimer, C. Buhay, Y. S. N. Butterfield, R. Carlsen, S. L. Carter, H. Chao, E. Chuah, A. Clarke, K. R. Covington, M. Dahdouli, N. Dewal, N. Dhalla, H. V Doddapaneni, J. A. Drummond, S. B. Gabriel, R. A. Gibbs, R. Guin, W. Hale, A. Hawes, D. N. Hayes, R. A. Holt, A. P. Hoyle, S. R. Jefferys, S. J. M. Jones, C. D. Jones, D. Kalra, C. Kovar, L. Lewis, J. Li, Y. Ma, M. A. Marra, M. Mayo, S. Meng, M. Meyerson, P. A. Mieczkowski, R. A. Moore, D. Morton, L. E. Mose, A. J. Mungall, D. Muzny, J. S. Parker, C. M. Perou, J. Roach, J. E. Schein, S. E. Schumacher, Y. Shi, J. V Simons, P. Sipahimalani, T. Skelly, M. G. Soloway, C. Sougnez, A. Tam, D. Tan, N. Thiessen, U. Veluvolu, M. Wang, M. D. Wilkerson, T. Wong, J. Wu, L. Xi, J. Zhou, J. Bedford, F. Chen, Y. Fu, M. Gerstein, D. Haussler, K. Kasaian, P. Lai, S. Ling, A. Radenbaugh, D. Van Den Berg, J. N. Weinstein, J. Zhu, M. Albert, I. Alexopoulou, J. J. Andersen, J. T. Auman, J. Bartlett, S. Bastacky, J. Bergsten, M. L. Blute, L. Boice, R. J. Bollag, J. Boyd, E. Castle, Y.-B. Chen, J. C. Cheville, E. Curley, B. Davies, A. DeVolk, R. Dhir, L. Dike, J. Eckman, J. Engel, J. Harr, R. Hrebinko, M. Huang, L. Huelsenbeck-Dill, M. Iacocca, B. Jacobs, M. Lobis, J. K. Maranchie, S. McMeekin, J. Myers, J. Nelson, J. Parfitt, A. Parwani, N. Petrelli, B. Rabeno, S. Roy, A. L. Salner, J. Slaton, M. Stanton, R. H. Thompson, L. Thorne, K. Tucker, P. M. Weinberger, C. Winemiller, L. A. Zach, and R. Zuna, “Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma.,” *N. Engl. J. Med.*, vol. 374, no. 2, pp. 135–45, Jan. 2016.

[28] M. Kovac, C. Navas, S. Horswell, M. Salm, C. Bardella, A. Rowan, M. Stares, F. Castro-Giner, R. Fisher, E. C. de Bruin, M. Kovacova, M. Gorman, S. Makino, J. Williams, E. Jaeger, A. Jones, K. Howarth, J. Larkin, L. Pickering, M. Gore, D. L. Nicol, S. Hazell, G. Stamp, T. O’Brien, B. Challacombe, N. Matthews, B. Phillimore, S. Begum, A. Rabinowitz, I. Varela, A. Chandra, C. Horsfield, A. Polson, M. Tran, R. Bhatt, L. Terracciano, S. Eppenberger-Castori, A. Protheroe, E. Maher, M. El Bahrawy, S. Fleming, P. Ratcliffe, K. Heinimann, C. Swanton, and I. Tomlinson, “Recurrent chromosomal gains and heterogeneous driver mutations characterise papillary renal cancer evolution.,” *Nat. Commun.*, vol. 6, p. 6336, Mar. 2015.

[29] S. Durinck, E. W. Stawiski, A. Pavía-Jiménez, Z. Modrusan, P. Kapur, B. S. Jaiswal, N. Zhang, V. Toffessi-Tcheuyap, T. T. Nguyen, K. B. Pahuja, Y.-J. Chen, S. Saleem, S. Chaudhuri, S. Heldens, M. Jackson, S. Peña-Llopis, J. Guillory, K. Toy, C. Ha, C. J. Harris, E. Holloman, H. M. Hill, J. Stinson, C. S. Rivers, V. Janakiraman, W. Wang, L. N. Kinch, N. V Grishin, P. M. Haverty, B. Chow, J. S. Gehring, J. Reeder, G. Pau, T. D. Wu, V. Margulis, Y. Lotan, A. Sagalowsky, I. Pedrosa, F. J. de Sauvage, J. Brugarolas, and S. Seshagiri, “Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes.,” *Nat. Genet.*, vol. 47, no. 1, pp. 13–21, Jan. 2015.

[30] Cancer Genome Atlas Research Network, “Comprehensive molecular characterization of clear cell renal cell carcinoma.,” *Nature*, vol. 499, no. 7456, pp. 43–9, Jul. 2013.

[31] Y. Sato, T. Yoshizato, Y. Shiraishi, S. Maekawa, Y. Okuno, T. Kamura, T. Shimamura, A. Sato-Otsubo, G. Nagae, H. Suzuki, Y. Nagata, K. Yoshida, A. Kon, Y. Suzuki, K. Chiba, H. Tanaka, A. Niida, A. Fujimoto, T. Tsunoda, T. Morikawa, D. Maeda, H. Kume, S. Sugano, M. Fukayama, H. Aburatani, M. Sanada, S. Miyano, Y. Homma, and S. Ogawa, “Integrated molecular analysis of clear-cell renal cell carcinoma.,” *Nat. Genet.*, vol. 45, no. 8, pp. 860–7, Aug. 2013.

[32] Y. Christinat and W. Krek, “Integrated genomic analysis identifies subclasses and prognosis signatures of kidney cancer.,” *Oncotarget*, vol. 6, no. 12, pp. 10521–31, Apr. 2015.