University ID No. (OTL will assign)

**WORCESTER POLYTECHNIC INSTITUTE**

**INVENTION DISCLOSURE FORM**

*Forward this completed and signed form to the Office of Technology Licensing*

**Title:** An ultra-sensitive variant detection model for low depth targeted next-generation sequencing data

**Department: Biomedical Engineering Dean:**

**1. Big Picture: What is the ultimate “one sentence” possible product?**

RVD2 is an ultra-sensitive variant detection model for low depth targeted next-generation sequencing data.

**2. Inventor(s) - Name, position, phone, email. (Identify all individuals who have made significant intellectual contributions to this invention's advance over prior technology, but do not include anyone merely because s/he has carried out some of the experimental work.)**

Patrick Flaherty, Assistant Professor, (508) 831-6583, [pjflaherty@wpi.edu](mailto:pjflaherty@wpi.edu)

Yuting He, Graduate student, (774) 578-5183, [yhe2@wpi.edu](mailto:yhe2@wpi.edu)

Fan Zhang, Graduate student, (508) 736-2008, [fzhang@wpi.edu](mailto:fzhang@wpi.edu)

**3. Specify any other inventor(s) who is/are an employee or an organization other than WPI and the institutional affiliation.**

None

**4. Background (To successfully determine the patentability of this invention, it will be necessary to compare it to existing technology, referred to as "prior art." Provide any references to assist in this evaluation.)**

**A. If possible, identify any references to the prior art by patent number or journal article identification.**

**B. Specify any deficiency in the prior art improved upon by this invention or any limitation which it extends.**

Next-generation sequencing (NGS) technology has enabled the systematic interrogation of the genome for a fraction of the cost of traditional assays (Koboldt et al., 2013). Protocol and platform engineering improvements have reduced the cost of the assay to the point where it is now possible to generate 1×109 bases of sequence data in 27 hours for approximately $1000 (Quail et al., 2012). As a result NGS is increasingly being used as a general platform for research assays for copy number variation (Alkan et al., 2009), promoter occupancy (Ouyang et al., 2009) and others (Rivera and Ren, 2013). Recently, NGS has been used as a clinical diagnostic tool for testing for mutations in non-invasive fetal DNA (Kitzman et al., 2012), infectious disease samples (Capobianchi et al., 2012), cancer (Navin et al., 2010), and human microbial communities (Consortium, 2013).

Clinical samples are often genetically heterogeneous. For example, non-invasive fetal DNA testing uses a data from maternal blood sample and filters for the minor fraction of reads that originate from cell-free DNA from the fetus (Fan et al., 2008). Infectious diseases such as HIV and influenza may contain many sub-populations (Flaherty et al., 2011 ;). DNA sequencing of individual regions of a solid tumor has revealed genetic heterogeneous within an individual sample (Navin et al., 2010).

However, the primary statistical tools for calling variants from NGS data are optimized for homogeneous samples. Most analysis pipelines make use of preprocessing or postprocessing or both to eliminate error prone reads and false positives. GATK uses a naive Bayesian decision rule to call a variant as one of two possible alleles (DePristo et al.2011). VarScan2-mpileup uses information from multiple replicates to call homogeneous and heterogeneous genotypes (Koboldt et al., 2012).

Recently, researchers have developed algorithms to call low-frequency or rare variants in heterogeneous samples. MuTect is able to handle low-purity samples by computing a log-odds score (Cibulskis et al., 2013). If the likelihood of the data assuming the locus is mutated exceeds the likelihood of the data assuming the site is not mutated by a threshold then a variant is called. There is an additional log-odds check on the normal sample to filter false positives. Strelka models the joint distribution for a tumor and normal sample as a mixture model using continuous values allele frequencies (Saunders et al., 2012). VarScan2 has a subcommand to call somatic mutations. Recently VarScan2 was shown to outperform many other algorithms for calling low-frequency (5%) variants.

**5. Briefly describe the invention (use additional sheets, as necessary). Indicate specifically what is considered to be the invention, as distinct from the prior art.**

**The description may be by reference to another document, which should be attached to this disclosure (e.g., copy of a report, preprint, excerpt from a proposal, etc.).**

**Also attach any sketches, flow charts, structural formulas, circuit diagrams, etc. that are appropriate to and necessary for full disclosure. Please identify any such attachment(s) positively by having each page signed, dated, and witnessed.**

See paper. Python codes are available upon request.

**6. What level of proof do you have for the invention? Working prototype, proof of concept experiments, etc?**

See paper.

**7. Has this invention been disclosed to others, either verbally or in written form (date, place, to whom, method of disclosure)?**

Siemens Diagnotics via phone call. Hanlee Ji & Mark Holodniy (Stanford University) phone call.

**8. Indicate any pending disclosures (date, place, to whom, method of disclosure).**

Paper (attached) by end of 2013.

**9. What does the market look like for this invention? Indicate the potential commercial use of this invention (e.g., fields of use, advantages, estimate of value)**

RVD2 can be used as a clinical diagnostic tool for testing for mutations in non-invasive fetal DNA (Kitzman et al., 2012), infectious disease samples (Capobianchi et al., 2012), cancer (Navin et al., 2010), and human microbial communities (Consortium, 2013).

**10. Indicate any potential commercial licensees that may be interested in this invention.**

Siemens Diagnostics. Qiagen next gen diagnostics. Contact Stanford OTL.

**11. Identify any grants, sponsors or projects (provide fund number) under which either conception or first reduction to practice occurred, including partial funding and Federal "formula" funding. Also list any related projects and/or inventions and any other potential claimants to rights in this invention. NOTE: This is very important to have the correct grant number in the proper format as WPI needs to report any inventions developed under federal grant money.**

Grant or contract funding: none

Sponsor Grant # Principal Investigator

Federal formula funds (Hatch or McIntyre-Stennis). Specify:

Other Sources of Funds (Describe, ie. EPSCOR, Industry). Specify:

**(OSP Initials \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) Verified that the grant information shown above is complete as listed with the University’s offices of Sponsored Programs.**

**12. Were any University funds or other resources used in making this invention (if yes, please explain).**

Yes. WPI seed funding.

**13. If funded by an external sponsor, has the sponsor been notified of this invention, either directly, in a progress or other report, or in an application for additional funds (date, sponsor, method of disclosure)?**

**n/a**

**This disclosure will become the first official University record of this invention.**

**Before signing, please ensure, to the best of your knowledge, that all information provided herein is complete and accurate.**

**Signed and submitted by:**

**Inventor's Signature Date Citizenship**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_**

**Print name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Home Address, Including City, State and Zip, non WPI email?**

**Inventor's Signature Date Citizenship**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_**

**Print name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Home Address, Including City, State and Zip, non WPI email?**

**Inventor's Signature Date Citizenship**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_**

**Print name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Home Address, Including City, State and Zip, non WPI email?**

**Inventor's Signature Date Citizenship**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_**

**Print name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Home Address, Including City, State and Zip, non WPI email?**

NOTE: WPI will assume that any eventual revenue from this invention will be split equally, unless there is a signed note designating a different split from all inventors.

**Departmental Endorsement:** To the best of my knowledge, the above information is correct.

Department Head Signature Date

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_**

**Advisor Endorsement for Inventions by WPI Students:** To the best of my knowledge,

the above information is correct.

Student Advisor Signature Date

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_**

For Technology Licensing Office Use

Date Received Acknowledged by

Sponsorship Rights Verified: Yes No

Copies Attached: Yes No

**Reference**

Alkan, C., Kidd, J. M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., ... & Eichler, E. E. (2009). Personalized copy number and segmental duplication maps using next-generation sequencing. Nature genetics, 41(10), 1061-1067.

Capobianchi, M. R., Giombini, E., & Rozera, G. (2013). Next‐generation sequencing technology in clinical virology. Clinical Microbiology and Infection,19(1), 15-22.

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., ... & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature biotechnology,31(3), 213-219.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics, 43(5), 491-498.

Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L., & Quake, S. R. (2008). Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proceedings of the National Academy of Sciences, 105(42), 16266-16271.

Flaherty, P., Natsoulis, G., Muralidharan, O., Winters, M., Buenrostro, J., Bell, J., ... & Ji, H. P. (2012). Ultrasensitive detection of rare mutations using next-generation targeted resequencing. Nucleic acids research, 40(1), e2-e2.

Kitzman, J. O., Snyder, M. W., Ventura, M., Lewis, A. P., Qiu, R., Simmons, L. E., ... & Shendure, J. (2012). Non-invasive whole genome sequencing of a human fetus. Science translational medicine, 4(137), 137ra76-137ra76.

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). The Next-Generation Sequencing Revolution and Its Impact on Genomics. Cell, 155(1), 27-38.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., ... & Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research, 22(3), 568-576.

Methé, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., ... & Buck, G. A. (2012). A framework for human microbiome research. Nature, 486(7402), 215-221.

Navin, N., Krasnitz, A., Rodgers, L., Cook, K., Meth, J., Kendall, J., ... & Wigler, M. (2010). Inferring tumor progression from genomic heterogeneity.Genome research, 20(1), 68-80.

Ouyang, Z., Zhou, Q., & Wong, W. H. (2009). ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells.Proceedings of the National Academy of Sciences, 106(51), 21521-21526.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., ... & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.BMC genomics, 13(1), 341.

Rivera, C. M., & Ren, B. (2013). Mapping Human Epigenomes. Cell, 155(1), 39-55.

Saunders, C. T., Wong, W. S., Swamy, S., Becq, J., Murray, L. J., & Cheetham, R. K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics, 28(14), 1811-1817.