

OPEN

Identification of somatic alterations in lipoma using whole exome sequencing

Deepika Kanojia¹, Pushkar Dakle¹, Anand Mayakonda^{1,2}, Rajeev Parameswaran³, Mark E. Puhaindran⁴, Victor Lee Kwan Min⁵, Vikas Madan¹ & Phillip Koeffler^{1,6,7}

Lipomas are benign fatty tumors with a high prevalence rate, mostly found in adults but have a good prognosis. Until now, reason for lipoma occurrence not been identified. We performed whole exome sequencing to define the mutational spectrum in ten lipoma patients along with their matching control samples. We presented genomic insight into the development of lipomas, the most common benign tumor of soft tissue. Our analysis identified 412 somatic variants including missense mutations, splice site variants, frameshift indels, and stop gain/lost. Copy number variation analysis highlighted minor aberrations in patients. Kinase genes and transcriptions factors were among the validated mutated genes critical for cell proliferation and survival. Pathway analysis revealed enrichment of calcium, Wnt and phospholipase D signaling in patients. In conclusion, whole exome sequencing in lipomas identified mutations in genes with a possible role in development and progression of lipomas.

Lipomas, soft tissue fat tumors, are the most common benign mesenchymal tumors¹. Lipomas are slow growing tumors and often occur under the skin on the neck, shoulders, arms, back, abdomen and thighs. However, occasionally lipomas may be present in deep location or originate within muscle². Lipoma affects only 1% of the population, although it is probably underreported³. These benign tumors are commonly found in adults and their prevalence is higher in men than in women. Lipomas are generally small round masses of less than 5 cm and typically are harmless unless they compress an organ. Patients may have either a solitary lipoma or multiple lipomas. Lipomas commonly are diagnosed by clinicians and histologically examined after surgical resection. Interestingly, prognosis for lipomas is very good with rare chance of recurrence after removal of the tumor⁴.

A more aggressive form of fat tumor is liposarcoma, which is also rare in occurrence but frequently develop dangerous metastatic tumors which is symptomatically different from lipomas⁴. Genomic pattern of liposarcomas showed copy number alterations (amplification of Chr 12q) as a major driver of initiation of malignancy along with several putative driver somatic mutations of PI3KCA, TP53, NF1, and EGFR⁵⁻⁷. Exact cause of lipomas is still not known, but lipomas have been identified with genetic rearrangements including structural changes at 12q13-15, 13q and 6p21-23 regions⁸. Approximately, 55–75% of solitary lipomas have cytogenetic abnormalities involving *HMGA2* gene rearrangements (truncation or fusion)⁹. Studies revealed enhanced adipogenesis and a normal rate of adipocyte apoptosis in lipoma tissue compared to normal fat tissue¹⁰. Lipoma appears to be close to well-differentiated liposarcomas in terms of their *HMGA2* structural alterations and high adipogenicity¹¹. However, no convincing literature existed to show liposarcomas arises from lipomas.

Genomic characterization of tumors improved our understanding of their molecular genesis for better management of the disease. Therefore, we screened superficial subcutaneous lipomas using whole exome sequencing to systematically profile their mutational signatures, copy number variations and identify affected pathways for better understanding of their molecular pathogenesis.

¹Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore. ²Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany. ³Division of Surgical Oncology, National University Cancer Institute, Singapore, Singapore. ⁴Department of Hand and Reconstructive Microsurgery, National University Hospital, Singapore, Singapore. ⁵Department of Pathology, National University Hospital, Singapore, Singapore. ⁶Division of Hematology/Oncology, Cedars-Sinai Medical Center, University of California, School of Medicine, Los Angeles, California, USA. ⁷National University Cancer Institute, National University Hospital, Singapore, Singapore. Correspondence and requests for materials should be addressed to D.K. (email: csidk@nus.edu.sg)

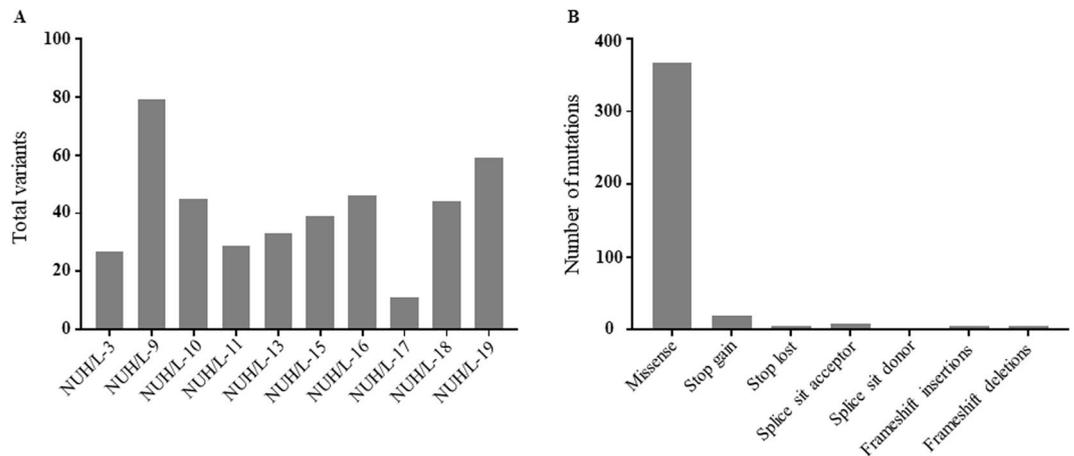


Figure 1. Variants identified in ten lipomas using whole exome sequencing. **(A)** Total number of variants found in each lipoma sample. **(B)** Number of each type of variant identified in ten lipoma samples.

Results and Discussion

We performed a genomic analysis to profile the somatic mutational landscape of lipomas by whole exome sequencing. To best of our knowledge, this is first study to perform exome sequencing on lipomas.

Clinical characteristics of lipoma patients. Primary subcutaneous lipomas [ordinary lipomas] were obtained from patients with their consent undergoing surgical resection at the National University Hospital. Detailed clinical features associated with these benign tumors presented in Supplementary Table 1. We collected lipomas from ten patients (all males) with a median age of 52 years. We performed whole exome sequencing on ten lipomas and normal controls from the same patient (either blood sample or adjacent normal tissue). Detail of normal controls used for each patient is given in Supplementary Table 1. Lipomas generally include various histologic subtypes but the samples used in the study are the common superficial primary subcutaneous solitary lipomas found on either back or arm.

Sequencing data. Whole exome sequencing of lipoma samples achieved an average of 135 million reads per sample (range 111–188 million). The control samples from the same patients found to have an average of 128 million reads per sample (range 112–164 million). Sequencing obtained a mean depth of 93 and 85 across the sequencing regions in lipomas and matching controls, respectively.

Allele frequency (AF) implicate how early a mutation occurred during progression of the tumor and our data suggest that the mutations occurred both early (variants with high AF) and later (variants with low AF) during development of their lipoma. In total, 5 frameshift deletions and 5 frameshift insertions occurred with 5–20% AF.

Mutational landscape in lipomas. Using our data pipeline analysis, we identified 412 somatic variants with a median of 42 per patient sample (range 11–79) [Fig. 1A and Supplementary Tables 2–11]. The type of mutations include 367 missense mutations, 19 stop gained, 5 stop lost, 9 splice acceptor, 2 splice site donor, 5 frameshift insertions and 5 frameshift deletions [Fig. 1B]. Details of types of mutation for each sample listed in Supplementary Table 12. Similar to other benign tumors and sarcomas, overall mutation burden in lipomas observed to be relatively low.

We validated the variants identified by data analysis of whole exome sequencing using Sanger sequencing as secondary technology to avoid the possibility of reporting out a false-positive result. However, Sanger sequencing has a limit of detecting somatic variants with AF about 10%, and higher. Variants with more than 10% of AF subjected to Sanger sequencing. We validated 42 variants out of 129 in total confirming these as true positive mutations in the patients. These 42 variants included 37 missense mutations, 1 frameshift deletion, 2 in-frame deletions, 1 splice donor and 1 stop gained mutation. Detailed list of variants, their mutation type and known function of genes provided in Supplementary Table 13. Validated mutations included kinase genes; ACVRIC (AF 28%), PTPRT (AF 14%), ERBB2 (AF 14%) and MAPK7 (AF 20%) and transcription factors ATMIN (AF 7%), ZNF317 (AF 26%) and FOSL1 (AF 19%); each are crucial for cell survival and proliferation. No single mutated gene was common in all ten tumors. Several of the genes are known to play functional role in oncogenesis such as APC (AF 26%), FOSL1 (AF 19%), ERBB2 (AF 14%), NFATC3 (AF 10%), PDE1A (AF 11%), ATMIN (AF 7%), and MAPK7 (AF 20%). In addition, experimental evidence have also indicated modulation of FOSL1¹², ERBB2¹³, and PDE1A¹⁴ regulates adipogenesis.

We also explored the differences or similarities of lipomas with cancerous liposarcomas and found few similarities with malignant liposarcomas. We compared the mutations identified in lipomas with mutation data of several liposarcoma studies^{5,15,16} and found three genes APC, RYR2 & MAPK7 mutated in both lipomas and liposarcomas. We also screened gene variants mutated in lipomas in sarcomas TCGA database and observed majority of these genes showed copy number alterations with very few mutations. However, variants identified in lipomas are different from sarcomas and not driver variants causing the disease. Therefore, we could not inferred

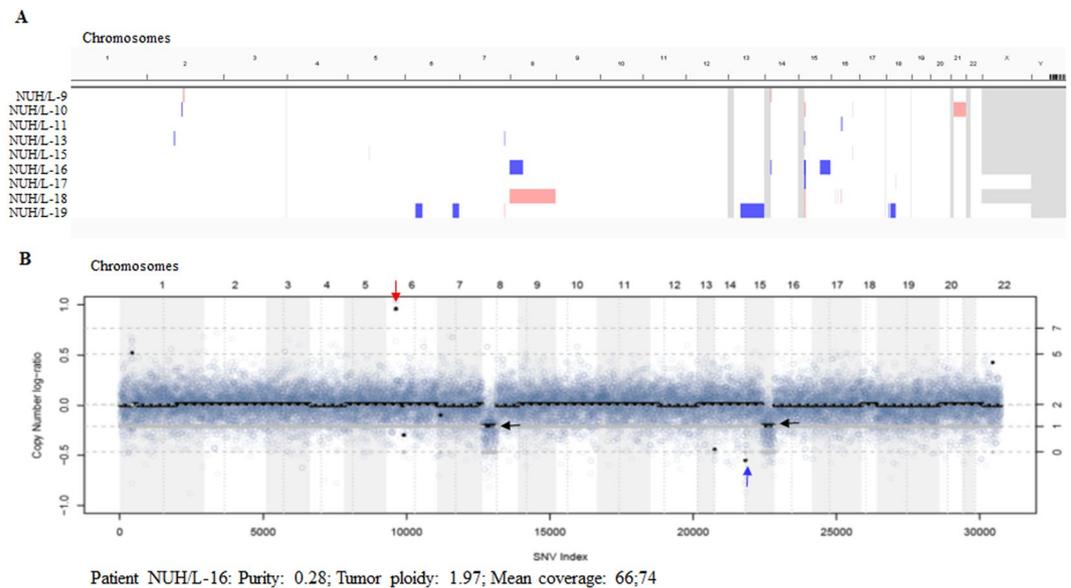


Figure 2. Copy number alterations in lipoma samples. **(A)** Genome-wide copy number variations in lipoma samples showing minor alterations. Chromosomes are represented from left to right in the order. Light blue indicates LOH (or shallow deletion), light red indicates three copies gain, and dark red indicates copy number equivalent to more than three copies. **(B)** Copy number log-ratios of sample NUH/L-16 analysed by PureCN. A dot represents a germline SNP and background colors visualize chromosomes and vertical dotted lines centromere positions. Grey line represents the expected allelic fractions in the segment, which are calculated using the estimated purity and segment copy numbers. Red arrow marks amplification, blue arrow marks deletion and black arrow marks LOH.

any relation between lipomas and liposarcomas as it is difficult to interpret based on few cases of lipomas with no recurrent alterations in lipomas.

Copy number variation analysis in lipoma patients. PureCN Rpackage was used to calculate the tumor purity and ploidy to call copy number in lipoma samples¹⁷. The samples used for whole exome sequencing were overall not highly pure, and two samples were flagged by PureCN. The tumor purity and ploidy calculated for each patient samples given in Supplementary Table 14. Due to very low tumor purity of sample NUH/L-3, we excluded the sample from this analysis. Low tumor purity affected the allelic frequency of the variants identified in tumors in our study and decreased sensitivity of detection. The tumor samples used in our study found to be contaminated with normal probably due to small size of these tumors, little control over purity at clinic level and few obtained from biopsy samples instead of surgical resections.

Copy number analysis indicated lipomas are mostly genomically silent as only few minor copy number aberrations were observed in these patients [Fig. 2A]. These results are in concordant with low mutation burden in lipoma samples. Analysis identified few amplifications, deletions and loss of heterozygosity (LOH) in lipoma patients, which are tabulated in Supplementary Tables 15 and 16. As a representative, copy number of sample NUH/L-16 presented as log-ratio demonstrating amplification (red arrow), deletion (blue arrow) and LOH (black arrow) [Fig. 2B]. Plots of copy number log-ratio of all other samples shown in Supplementary Fig. 1. Three recurrent copy number changes found in lipoma samples; deletion of NFSAP1, LINC02203 and OR4M2 genes in two patients. Few LOH regions were observed in two or more than two lipoma patients such as chr15:20169937-21236751, chr6:31948633-32014371, and chr8:7272427-7754188. List of LOH regions identified in patient samples provided in Supplementary Table 16.

Mutational patterns in lipoma patient samples. We performed mutational spectrum analysis of lipomas to categorize their mutational signature and to identify functional mutagenic processes in lipomas. Patient wise signature analysis is shown in Supplementary Fig. 2. No mutational signature was detected for patient NUH/L-17 due to lower mutational burden and low allele frequency of variants. Analysis revealed mutational signature 15 as the major signature affecting lipoma patients [Fig. 3A,B]. Signature 15 is one of the mutational signature associated with defective DNA mismatch repair that contributes to large numbers of substitutions and small indels at nucleotide repeats. In addition, this signature found in several samples of lung and stomach cancer. Surprisingly, this is the most common signature in a benign tumor. We had an interesting observation that mutational signature 1 affects malignant liposarcomas in previous study⁵ whereas none of the lipoma patients showed this signature. Therefore, these results suggests different biological mechanisms involved in mutagenesis of lipoma.

Altered pathways in lipomas. Pathway analysis was conducted to identify pathways and biological processes affected by mutated genes in lipomas. This analysis included gene variants validated and confirmed with

Methods

Tissue collection and sample preparation. Lipomas, normal adjacent control tissues and patient's blood samples were obtained from National University Hospital, Tissue repository, Singapore in accordance with ethical guidelines of National University of Singapore, Institutional Review Board (NUS-IRB). According to their requirement, samples were collected after obtaining informed written consent from patients. Tissue samples or blood received after surgery was snap frozen and stored in liquid nitrogen until processing. Genomic DNAs were isolated from tumors, control tissues and blood samples using QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. Genomic DNAs were quantified by Qubit dsDNA BR assay kit (Life technologies) and their quality verified by agarose gel electrophoresis. All the experimental protocols were approved by the National University of Singapore, Office of safety, health and environment (NUS, OSHE).

Whole exome sequencing. Exome libraries were prepared using the Agilent SureSelect All Human Exon v5 capture kit⁵. Briefly, the qualified genomic DNA samples were randomly fragmented into base pair peak of 150 to 200 bp, and adapters ligated to both ends of the resulting fragments. Adapter-ligated templates were purified by the Agencourt AMPure SPRI beads, and fragments with insert size of about 200 bp were excised. Extracted DNA was amplified by ligation-mediated polymerase chain reaction (LM-PCR), purified, and hybridized to the SureSelect Biotinylated RNA Library (BAITS) for enrichment. Hybridized fragments were bound to the streptavidin beads whereas non-hybridized fragments were washed away after 24 h. Captured LM-PCR products were subjected to Agilent 2100 Bioanalyzer to estimate the magnitude of enrichment. Each captured library was independently loaded on Hiseq 4000 platform for high-throughput sequencing to ensure that each sample meets the desired average fold-coverage. Raw image files were processed by Illumina base calling Software 1.7 with default parameters, and the sequences were generated as 90 bp paired-end reads. All the experimental protocols were approved by the NUS, OSHE.

Analysis of whole exome sequencing. Paired end sequencing reads were aligned to the human reference genome (hg19) using bwa mem. Samblaster was subsequently used to mark PCR duplicates²². Systematic errors in base quality scores were then corrected using GATK4 Base Recalibrator²³. Somatic single nucleotide variants and short indels were detected against a matched normal using both GATK4 Mutect2 and Strelka2²⁴. Variants calls from Mutect2 further filtered with GATK4 FilterMutectCalls. vcf2maf was used to convert the .vcf files to MAF format and also annotated using Ensembl VEP (Version 92_GRCh37)²⁵. Depth of coverage for all samples calculated using GATK3 DepthOfCoverage.

We excluded any variants which were not tagged PASS by FilterMutectCalls or Strelka2. Variants were further filtered to remove those with global minor allele frequency (GMAF) lower than 1% or variants tagged as common variants by vcf2maf based on allele count in Exac populations, while still retaining those classified as "pathogenic" in ClinVar. Variants with either LOW/MODIFIER impact as per VEP or with variant AF less than 0.01 were also removed.

Mutation signature analysis. MuSiCa, online tool was used for somatic characterization of mutation signatures present in the patient's variants list²⁶. This tool based on MutationalPatterns and built using the Shiny framework in R language. MutationalPatterns extracts de novo signatures using the original NMF algorithm and quantifies COSMIC-reported signatures. A series of mutational profiles extracted using the information of somatic single nucleotide variants and 96 possibilities assessed to detect processes accountable for the same substitutions but in different contexts.

Pathway analysis. Refseq-annotated mutations (genes) were used for pathway analysis using PathScan algorithm integrated into MuSic pipeline^{27,28}. P values calculated in KEGG pathway analysis, where, first it identifies the subset of genes, and proteins, that contribute in a particular KEGG pathway for the input. Then, it evaluates the fraction of those biological features that overlaps with the set of features of input genes. In the final step, KEGG computes the significance of the overlap using the Fisher exact test, which represents as p-value interpreted as a measurement of the confidence that this overlap is due to chance.

Copy number variation analysis. PureCN¹⁷ was used to estimate tumor purity and call copy number alterations. Coverage values for the targeted regions were calculated and normalized for GC-content and replication bias. Replication timing data for K562 cell line (wgEncodeUwRepliSeqK562WaveSignalRep1.bigWig) was obtained from the UCSC genome browser. A variant caller format (VCF) file containing germline variants present in 3 or more normal samples was used for obtaining read mapping bias estimates. Heterozygous short-nucleotide polymorphisms (SNPs) in targeted regions including 50 bp flanks were used to calculate the allele-specific copy numbers. Genes were called amplified if the segment copy number was ≥ 6 for focal amplifications or ≥ 7 for non-focal amplifications. Genes with a segment copy number of ≤ 0.5 were called as deletions.

Sanger sequencing validations. Gene variants with AF greater than 10% were validated by Sanger sequencing. Polymerase chain reaction was performed using 25 ng template DNA of lipoma tissue and normal control sample from the same individual. PCR products run on 1% agarose gel and purified using Wizard[®] SV Gel and PCR Clean-Up System (Promega). Purified PCR products were sequenced and analysed with the Sequence Scanner Software 2 Version 2 (Applied Biosystems).

Ethics approval and consent to participate. Patient's tissues and blood samples collected in accordance with ethical guidelines of National University of Singapore, Institutional Review Board (NUS-IRB) and after obtaining informed written consent from patients. All the experimental protocols were approved by the National University of Singapore, Office of safety, health and environment (NUS, OSHE).

References

- Kransdorf, M. J., Moser, R. P. Jr., Meis, J. M. & Meyer, C. A. Fat-containing soft-tissue masses of the extremities. *Radiographics* **11**, 81–106, <https://doi.org/10.1148/radiographics.11.1.1996399> (1991).
- McTighe, S. & Chernev, I. Intramuscular lipoma: a review of the literature. *Orthop Rev (Pavia)* **6**, 5618, <https://doi.org/10.4081/or.2014.5618> (2014).
- Luba, M. C., Bangs, S. A., Mohler, A. M. & Stulberg, D. L. Common benign skin tumors. *Am Fam Physician* **67**, 729–738 (2003).
- Dalal, K. M., Antonescu, C. R. & Singer, S. Diagnosis and management of lipomatous tumors. *J Surg Oncol* **97**, 298–313, <https://doi.org/10.1002/jso.20975> (2008).
- Kanojia, D. *et al.* Genomic landscape of liposarcoma. *Oncotarget* **6**, 42429–42444, <https://doi.org/10.18632/oncotarget.6464> (2015).
- Beird, H. C. *et al.* Genomic profiling of dedifferentiated liposarcoma compared to matched well-differentiated liposarcoma reveals higher genomic complexity and a common origin. *Cold Spring Harb Mol Case Stud* **4**, <https://doi.org/10.1101/mcs.a002386> (2018).
- Cancer Genome Atlas Research Network, Electronic address, e. d. s. c. & Cancer Genome Atlas Research, N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* **171**, 950–965 e928, <https://doi.org/10.1016/j.cell.2017.10.014> (2017).
- Nishio, J. Contributions of cytogenetics and molecular cytogenetics to the diagnosis of adipocytic tumors. *J Biomed Biotechnol* **2011**, 524067, <https://doi.org/10.1155/2011/524067> (2011).
- Hatano, H. *et al.* Clinicopathological features of lipomas with gene fusions involving HMGA2. *Anticancer Res* **28**, 535–538 (2008).
- Suga, H. *et al.* Cellular and molecular features of lipoma tissue: comparison with normal adipose tissue. *Br J Dermatol* **161**, 819–825, <https://doi.org/10.1111/j.1365-2133.2009.09272.x> (2009).
- Hatano, H. *et al.* Well-differentiated liposarcoma associated with benign lipoma. *Anticancer Res* **24**, 1039–1044 (2004).
- Luther, J. *et al.* Elevated Fra-1 expression causes severe lipodystrophy. *J Cell Sci* **124**, 1465–1476, <https://doi.org/10.1242/jcs.079855> (2011).
- Harrington, M., Pond-Tor, S. & Boney, C. M. Role of epidermal growth factor and ErbB2 receptors in 3T3-L1 adipogenesis. *Obesity (Silver Spring)* **15**, 563–571, <https://doi.org/10.1038/oby.2007.562> (2007).
- Snyder, P. B., Esselstyn, J. M., Loughney, K., Wolda, S. L. & Florio, V. A. The role of cyclic nucleotide phosphodiesterases in the regulation of adipocyte lipolysis. *J Lipid Res* **46**, 494–503, <https://doi.org/10.1194/jlr.M400362-JLR200> (2005).
- Neeta Somaiah, H. C. B. *et al.* and David Hong. Targeted next generation sequencing of well-differentiated/dedifferentiated liposarcoma reveals novel gene amplifications and mutations. *Oncotarget* **9**, 19891–19899 (2018).
- Ali Amin-Mansour, S. G. *et al.* Wagner. Genomic Evolutionary Patterns of Leiomyosarcoma and Liposarcoma. *Clinical Cancer Research* **25**, 5135–5142 (2019).
- Markus, R. *et al.* PureCN: copy number calling and SNV classification using targeted short read sequencing. *Source Code for Biology and Medicine* **11** (2016).
- Pramme-Steinwachs, I., Jastroch, M. & Ussar, S. Extracellular calcium modulates brown adipocyte differentiation and identity. *Sci Rep* **7**, 8888, <https://doi.org/10.1038/s41598-017-09025-3> (2017).
- Bae, Y. K. *et al.* Intracellular Calcium Determines the Adipogenic Differentiation Potential of Human Umbilical Cord Blood-Derived Mesenchymal Stem Cells via the Wnt5a/beta-Catenin Signaling Pathway. *Stem Cells Int* **2018**, 6545071, <https://doi.org/10.1155/2018/6545071> (2018).
- Christodoulides, C., Lagathu, C., Sethi, J. K., Vidal-Puig, A. & Adipogenesis, W. N. T. signalling. *Trends Endocrinol Metab* **20**, 16–24, <https://doi.org/10.1016/j.tem.2008.09.002> (2009).
- Song, H. I. & Yoon, M. S. PLD1 regulates adipogenic differentiation through mTOR - IRS-1 phosphorylation at serine 636/639. *Sci Rep* **6**, 36968, <https://doi.org/10.1038/srep36968> (2016).
- Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505, <https://doi.org/10.1093/bioinformatics/btu314> (2014).
- McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303, <https://doi.org/10.1101/gr.107524.110> (2010).
- Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**, 591–594, <https://doi.org/10.1038/s41592-018-0051-x> (2018).
- McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, <https://doi.org/10.1186/s13059-016-0974-4> (2016).
- Diaz-Gay, M. *et al.* Mutational Signatures in Cancer (MuSiCa): a web application to implement mutational signatures analysis in cancer samples. *BMC Bioinformatics* **19**, 224, <https://doi.org/10.1186/s12859-018-2234-y> (2018).
- Dees, N. D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Res* **22**, 1589–1598, <https://doi.org/10.1101/gr.134635.111> (2012).
- Wendl, M. C. *et al.* PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* **27**, 1595–1602, <https://doi.org/10.1093/bioinformatics/btr193> (2011).

Acknowledgements

This study is supported by the National Institute of Health (R01-CA200992-03), National Research Foundation Singapore under its Singapore Translational Research (STaR) Investigator Award (NMRC/STaR/0021/2014) and administered by the Singapore Ministry of Health's National Medical Research Council (NMRC), the NMRC Centre Grant awarded to National University Cancer Institute of Singapore, the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiatives and is additionally supported by the Wendy Walk Foundation.

Author Contributions

D.K. and P.K. designed the experiments; D.K. performed the experiments; D.K., P.D., A.M. and V.M. analysed the results; R.P., M.E.P. and V.L.K. provided the patient samples; P.K. supervised the project and D.K., P.D. and P.K. wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-50805-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019