FFPENET: SOMATIC VARIANT CALLING IN FFPE TUMOR SAMPLES USING DEEP TRANSFER LEARNING

Kiran Krishnamachari, Anders Jacobsen Skanderup
Laboratory of Computational Cancer Genomics, Genome Institute of Singapore (GIS)
Agency for Science, Technology and Research (A*STAR)
60 Biopolis Street, Genome, Singapore 138672, Republic of Singapore
{kiran_krishnamachari,skanderupamj}@gis.a-star.edu.sg

ABSTRACT

Detecting somatic mutations in tumors is fundamental to improving our understanding of cancer and to recommend customized treatment strategies. In clinical settings, next-generation sequencing is performed predominantly on tumor biopsies that have undergone formalin fixation and paraffin embedding (FFPE), which allows laboratories to store and transport samples at room temperature without significantly affecting their morphology. However, the FFPE process is known to cause chemical changes in DNA and significantly affects the accuracy of variant calling algorithms, which are primarily developed for research-grade fresh frozen (FF) samples. Hence, there is a mismatch between software tools used in research settings and what is needed in clinical settings. Here we develop FFPENet, a convolutional neural network (CNN) trained to identify and remove FFPE artifacts in the output of somatic variant callers on FFPE tumor samples. FFPENet encodes raw reads overlapping each mutation in a multi-channel tensor input and fine-tunes a CNN pre-trained on a large number of FF tumor whole genome samples using a multi-sector cohort of patient-matched FF and FFPE tumor samples. We demonstrate the improved effectiveness of this approach over existing methods by benchmarking on a multi-sector whole-exome sequencing (WES) cohort of lung cancer. Models will be released at https://github.com/skandlab/VarNet.

1 INTRODUCTION

Identification of mutations in cancer genomes is an important step in the study of cancer genomics to improve our understanding as well as treatment of cancer. Formalin-fixed paraffin-embedded (FFPE) tumor tissue biopsies are commonly used for translational research and clinical diagnostics as they enable economical storage and transportation of biological samples. Next-generation sequencing (NGS) of DNA derived from FFPE tumor samples is frequently performed for both research as well as therapeutic purposes (Figure 1). However, the FFPE preservation process introduces significant DNA damage that results in misread bases during NGS. This makes the accurate determination of somatic mutations from FFPE tumor samples significantly more challenging compared to research grade fresh-frozen (FF) tumor samples (Oh et al., 2015; Srinivasan et al., 2002). FFPE artifacts tend to occur at low-allele-frequencies and the most prominent of these are C→T/G→A changes due to the hydrolytic deamination of cytosines (Steiert et al., 2023). FFPE artifacts especially confound the detection of true low-allele-frequency somatic mutations that could be of clinical relevance.

Simply filtering low-allele-frequency mutation calls may mislead downstream clinical analysis. Existing somatic mutation callers such Strelka2 (Kim et al., 2018), Mutect2 (Cibulskis et al., 2013), Varscan (Koboldt et al., 2012), Vardict (Lai et al., 2015), Freebayes-somatic (Garrison & Marth, 2012) and VarNet (Krishnamachari et al., 2022) are usually developed and benchmarked on research-grade FF tumor samples, which do not contain significant DNA damage. Hence, their performance significantly degrades on FFPE tumor samples mainly due to the presence of a large proportion of false-positive mutations or artifacts in their outputs (Steiert et al., 2023; Yost et al., 2012). Mutation callers can be evaluated on FFPE tumor samples using matched FF samples derived from the same patient and tumor. Mutation calls identified in FFPE samples but absent in the matched FF tumor

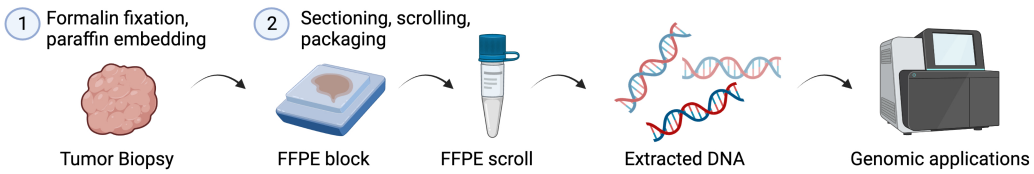

Figure 1: Next-generation sequencing of formalin-fixated paraffin-embedded (FFPE) tumor biopsies is frequently performed for translational research and clinical diagnostics.

samples are considered putative FFPE artifacts (de Schaetzen van Brienen et al., 2020), although intra-tumor heterogeneity (ITH) may introduce complexity.

Here we develop FFPENet, a deep transfer learned model that filters out artifacts from the outputs of callers such as VarNet, Strelka2 and Mutect2 on FFPE tumor samples. As FFPE tumor datasets with established ground truth mutations are lacking, we hypothesize that a transfer learning approach using a model pre-trained on a large number of FF tumor samples would be effective in removing FFPE artifacts. Hence, we use VarNet (Krishnamachari et al., 2022), a convolutional neural network, as the pre-trained deep learning model for training FFPENet as it was trained on a large number (> 300) of cancer whole genomes comprising multiple cancer types. We also hypothesize that encoding aligned sequencing reads in a multi-channel image-like representation would enable such an approach without having to manually encode FFPE related signatures. We demonstrate using a cohort of matched FF and FFPE lung cancer samples that FFPENet can significantly improve accuracy compared to existing methods.

2 Related work

As most popular somatic variant callers are optimized for FF tumor samples, their performance significantly degrades on FFPE samples due to a large number of false-positive mutation artifacts. Recent work has attempted to improve the quality of variant call-sets by performing post-hoc removal of artifacts from the outputs of popular variant callers. For example, IdeaFix (Tellaetxe-Abete et al., 2021) filters likely mutation artifacts from the output of Mutect2 using a decision tree-based approach exploiting multiple features such as read pair orientation bias, genomic context and variant allele frequency. IdeaFix annotates C$\rightarrow$T/G$\rightarrow$A calls made by Mutect2 as either true variants or artifacts. SOBDetector (Diossy et al., 2021) proposed a method to filter artifacts from the output of any mutation caller using the *strand orientation bias* feature. FFPolish (Dodani et al., 2022) proposed a logistic regression model of multiple features including variant allele frequency and variant read-quality metrics to filter artifacts from the outputs of mutation callers. Other work has also proposed using mutation calls made by *any two callers* on FFPE tumor samples as a simple baseline to reduce artifacts (de Schaetzen van Brienen et al., 2020). This strategy however would not exclude artifacts that are misclassified as mutations by more than one caller. cisCall is a tool for variant calling from Illumina sequencing data from FFPE samples (Kato et al., 2018). In our work, we adapt VarNet (Krishnamachari et al., 2022) as it is an accurate end-to-end deep learning based somatic variant caller that can be fine-tuned on new datasets[1].

3 FFPENet

FFPE artifacts are associated with many features of the sequencing data such as their position in reads, genomic context, orientation bias etc. For example, FFPE deaminations have been observed to be enriched at the ends of molecules, due to an increased sensitivity to deaminate at overhanging ends (Briggs et al., 2007; Lindahl & Nyberg, 1972). As FFPE tumor samples with established ground truth mutations are scarce, we use a transfer learning approach using a deep learning model, VarNet (Krishnamachari et al., 2022). VarNet is an accurate model pre-trained on a large number of FF tumor whole genome samples, hence, it is expected to have learned useful low-level features that are relevant for variant calling. VarNet's image-based encoding of raw sequencing data surrounding

[1]Pre-trained models were retrieved from https://github.com/skandlab/VarNet

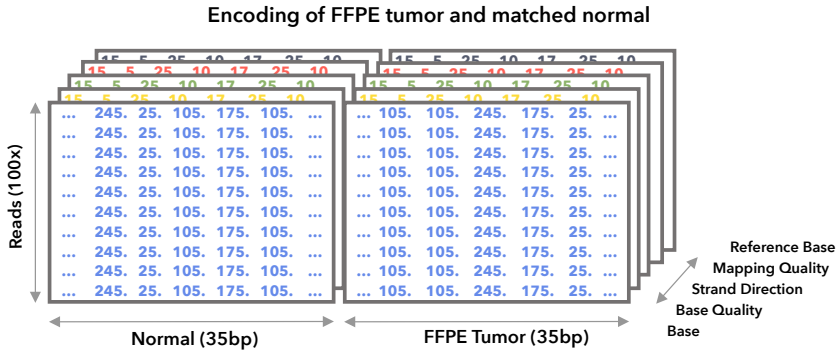

Figure 2: Multi-channel representation of aligned tumor and normal reads for each candidate variant.

mutation candidates enables the model to learn to multiple features to filter artifacts. Features of each aligned read i.e. base, base quality, mapping quality, strand bias, are numerically encoded in a tensor input to the convolutional neural network (Figure 2). The FFPENet model is initialized with pre-trained weights from the VarNet model and is further fine-tuned to remove artifacts from VarNet's output on FFPE samples. FFPENet is run as a post-processing step after variant calls are first generated by a somatic variant caller. Hence, it is computationally inexpensive to run.

We trained FFPENet using 9-fold cross validation using a cohort of nine patients with matched FF and FFPE tumor samples as well as matched germline samples. In each cross-validation fold, we left out one patient for testing. We report aggregated results across all test samples below. We used cross-validation to choose the following training hyper-parameters, 1) number of fully-connected layers to fine-tune 2) learning-rate 3) label-smoothing parameter 4) weight-decay parameter and 5) training batch-size. The hyper-parameters search ranges are detailed in the Appendix (Table 3).

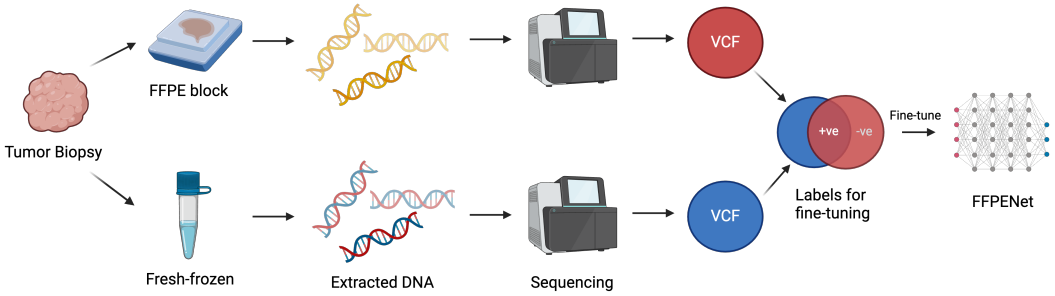

Figure 3: Matched FF and FFPE tumor samples were sequenced to create a dataset to train FFPENet to differentiate between *mutations* and *artifacts*.

3.1 Training Data

We obtained whole-exome-sequencing (WES) data for a cohort of nine lung cancer patients with matched FF and FFPE tumor biopsies as well as matched normal samples. The FF and FFPE samples were derived from tumor biopsies taken at the same time. We generated mutation calls for both FF and FFPE tumor samples using VarNet, Strelka2 and Mutect2. For each patient and variant caller, mutation calls made in the FFPE tumor sample but not in any matched FF sample are considered *putative FFPE artifacts*. Mutation calls made in both FF and FFPE tumor samples are considered somatic *mutations* as shared mutations are unlikely to be due to FFPE damage. We used this strategy to create a labeled class-balanced dataset of FFPE tumor samples to train a model to differentiate between somatic *mutations* and *FFPE artifacts* (Figure 3).

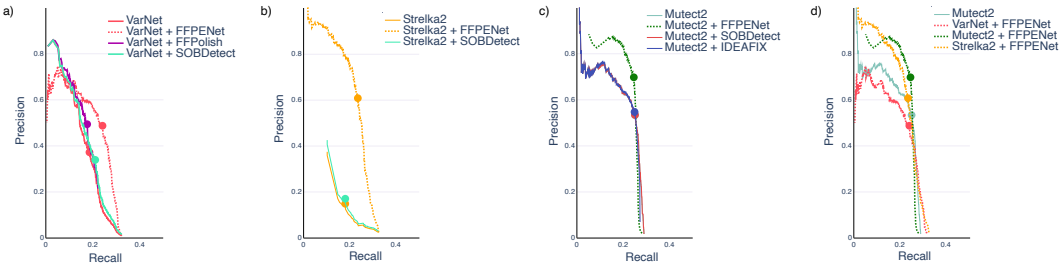

Figure 4: Precision/recall curves for SNV calling using a) VarNet b) Strelka2 c) Mutect2 and d) top-4 callers. Solid circles in the curves indicate the highest F1 accuracy.

3.2 Multi-sector data

Intra-tumor heterogeneity (ITH) is an important factor in the treatment and study of tumors. ITH refers to distinct cell populations within a tumor leading to differences in drug sensitivity, tumor growth rate and prognosis. Hence, it is important to identify mutations in different cell populations within a tumor. Sub-clonal mutations in tumors can occur at low enough variant allele frequencies (VAF) that they can be mistaken for FFPE-related artifacts if not found in the matched FF samples derived from other distinct cell populations within the tumor. This can introduce incorrectly labeled data in our training set. To mitigate this issue, we used multi-sector (multi-region) whole exome sequencing of tumors in our cohort to help differentiate between FFPE artifacts and sub-clonal variants. Hence, we defined FFPE artifacts as variants identified in the FFPE tumor sample but not in any matched FF sample. We obtained 3-4 FF sectors per matched tumor sample to provide a broad view of the heterogeneity within the tumor. In the absence of multi-sector data, it would be difficult to differentiate artifacts from real sub-clonal variants in the FFPE tumor sample. To the best of our knowledge, this is a novel approach in the study of FFPE-artifacts and has not been reported before.

3.3 Evaluation

We defined a ground-truth mutation set to evaluate variant calling accuracy on the FFPE tumor samples using a combination of variant calls identified in the matched FF samples. We treated variants called by both SMuRF (Huang et al., 2019) and VarNet on the matched FF samples as ground-truth (union of variant calls in multiple frozen sectors). To ensure fair representation of all the baseline variant callers included in this study (Strelka2, Mutect2, Varscan, Vardict, Freebayes, and VarNet), we established the ground-truth mutation set using the intersection of calls made by SMuRF, an ensemble of five methods, and VarNet on the matched FF samples. For each benchmarked method, we classified variants it detects in an FFPE tumor sample but not found in the ground-truth mutation set as false positives (FPs). The formulas used to compute the accuracy metrics of callers are:

$$recall = \frac{Caller \cap (VarNet \cap SMuRF)}{VarNet \cap SMuRF}$$

$$precision = \frac{Caller \cap (VarNet \cap SMuRF)}{Caller}$$

$$F1 = Harmonic\ mean(recall,\ precision)$$

4 Results

Using the matched FF samples and the definitions above, we evaluated FFPENet and other existing methods i.e. SOBDetect, FFPolish and IdeaFix, based on their ability to retain true somatic mutations while filtering out artifacts in FFPE tumor samples. We applied each of these methods to the mutation calls made by each baseline variant caller on FFPE tumor samples[2]. At its default prediction threshold, VarNet (Krishnamachari et al., 2022) produced a large number of artifacts (20,222)

[2]IdeaFix (Tellaetxe-Abete et al., 2021) only supports Mutect2.

Table 1: Performance of callers on SNV calling. True Positives (TPs) and False Positive (FPs) calls are defined using the definition above. F1 scores are calculated using both the default PASS threshold (Default F1) and the best threshold (Max F1) of each caller.

Caller	**Truth**	**TPs**	**FPs**	**Default F1**	**Max F1**
VarNet	1,226	379	20,222	0.035	0.247
VarNet + FFPENet	1,226	298	367	0.315	0.322
VarNet + SOBDetect	1,226	379	13,374	0.050	0.259
VarNet + FFPolish	1,226	275	1,001	0.220	0.259
Strelka2	1,226	399	16,645	0.044	0.163
Strelka2 + FFPENet	1,226	273	141	0.333	0.339
Strelka2 + SOBDetect	1,226	398	13,211	0.054	0.175
Mutect2	1,226	356	34,166	0.020	0.341
Mutect2 + FFPENet	1,226	280	93	**0.350**	**0.363**
Mutect2 + SOBDetect	1,226	355	27,330	0.025	0.341
Mutect2 + IdeaFix	1,226	334	4,523	0.110	0.342
Vardict	1,226	387	20,537	0.035	0.035
Vardict + SOBDetect	1,226	386	17,646	0.040	0.040
Varscan	1,226	389	19,727	0.036	0.096
Varscan + SOBDetect	1,226	389	18,230	0.039	0.097
Freebayes	1,226	360	32,324	0.021	0.032
Freebayes + SOBDetect	1,226	360	26,094	0.026	0.032
SMuRF	1,226	484	314,054	0.003	0.081
SMuRF + SOBDetect	1,226	482	283,246	0.003	0.082

on all nine FFPE tumor samples combined (Table 1), leading to a low F1-score (0.035). Running FFPENet on the output of VarNet significantly reduced the number of artifacts (367 vs 20,222) leading to a higher F1-score for VarNet + FFPENet compared to VarNet (0.315 vs 0.035) (Table 1). At the default prediction threshold, FFPENet provided less recall than VarNet alone (0.243 vs 0.309). However, FFPENet is able to provide higher recall with better precision when its prediction threshold is modified (Figure 4). We note that none of the callers exceeded 40% recall (sensitivity). This is due to the fact that the ground-truth SNVs are defined using the *union* of multiple sectors in the matched frozen tumor. The FFPE tumor sample is not expected to carry all mutations found in multiple sectors due to intra-tumor heterogeneity. These results suggest that callers are likely identifying the clonal mutations that are found in both the FFPE sample as well as the matched frozen samples.

FFPENet significantly enhances the performance of VarNet, Strelka2, and Mutect2 compared to other post-processing methods such as SOBDetect, FFPolish, and IdeaFix (Figure 4 and Table 1). At their default score thresholds, most variant callers tend to generate an excessive number of false-positive mutation calls. By applying FFPENet, we observed substantial improvements in the precision and overall performance of VarNet, Strelka2, and Mutect2.

Table 2: Percentage improvement in F1 scores by applying FFPE post-processing methods to baseline variant callers (computed at default and optimal score thresholds). FFPENet achieves the highest improvements in performance per caller.

Baseline Caller	**FFPENet (ours)**		**SOBDetect**		**FFPolish**		**IdeaFix**	
	Default F1	**Best F1**	**Default F1**	**Best F1**	**Default F1**	**Best F1**	**Default F1**	**Best F1**
VarNet	**800.00%**	**30.36%**	42.86%	4.86%	528.57%	4.86%	-	-
Strelka2	**656.82%**	**107.98%**	22.73%	7.36%	-	-	-	-
Mutect2	**1650.00%**	**6.45%**	25.00%	0.00%	-	-	450.00%	0.29%

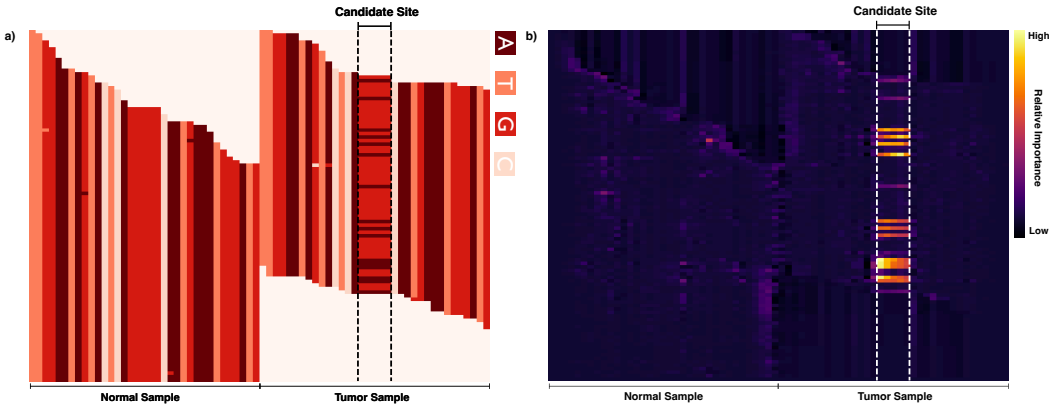

Figure 5: a) Input representation (only base channel is shown) and b) importance heatmap of an FFPE artifact presented to FFPENet. The model is able to identify variant alleles in the candidate mutation site.

To provide a comprehensive evaluation, we report both the default performance of each method and the maximum F1 score achieved by varying the score threshold. This is important because downstream studies typically rely on the default quality thresholds set by each caller. Moreover, tuning these thresholds on new test samples is challenging in the absence of ground-truth mutations. Our results underscore the need for caution when using default quality thresholds for FFPE tumor samples. We recommend either optimizing the quality threshold for each caller or using a post-processing tool such as FFPENet to refine variant calls.

Among all tested approaches, the combination of Mutect2 + FFPENet yielded the highest F1 accuracy on our test samples, both at default and optimized quality thresholds. Notably, three of the top four performing methods incorporated FFPENet, highlighting its effectiveness (Figure 4d). A detailed summary of FFPENet's performance improvements when applied to baseline variant callers is presented in Table 2.

4.1 Interpreting FFPENet

After training FFPENet, we analyzed the features learned by the model by visualizing the "importance" it assigns to different input "pixels" using Guided Backpropagation (Springenberg et al., 2015). This method highlights the input regions that most strongly influence the model's predictions, allowing us to interpret its decision-making process.

Our analysis revealed that FFPENet has automatically learned to recognize variant alleles at the candidate mutation site while also attending to contextual mutational patterns at other genomic positions (Figure 5). This suggests that the model is not only detecting individual variants but is also leveraging broader mutational signatures that are indicative of FFPE-related artifacts. By capturing these spatial dependencies, FFPENet can distinguish FFPE artifacts from genuine variants more effectively. This approach contrasts with other methods, which rely on explicitly encoding only a predefined set of known FFPE-related properties. Instead of being constrained by hand-engineered features, FFPENet autonomously discovers FFPE-specific signatures directly from sequencing data. This flexibility enables it to potentially uncover novel artifact patterns that may not have been previously characterized.

5 Conclusion

As FFPE tumor datasets with established ground truth are scarce, we hypothesized that a transfer learning approach using a model pre-trained on a large number of FF samples would be effective. We also hypothesized that encoding sequencing reads in a multi-channel image-like representation would enable such an approach without needing to manually encode FFPE related signatures. We used a novel approach using multi-sector tumor sequencing data to train and evaluate somatic vari-

ant calling in FFPE tumor samples. We demonstrated that FFPENet learned useful features for identifying FFPE artifacts from raw data and outperformed existing methods when combined with state-of-the-art somatic variant callers such as Strelka2, VarNet and Mutect2. Our work is a step towards improving the utility of sequencing FFPE-preserved tumor biopsies for translational and basic research. Future work involves training our model on larger cohorts for robust performance across cancer types.

REFERENCES

Adrian W. Briggs, Udo Stenzel, Philip L. F. Johnson, Richard E. Green, Janet Kelso, Kay Prüfer, Matthias Meyer, Johannes Krause, Michael T. Ronan, Michael Lachmann, and Svante Pääbo. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, 104(37):14616–14621, 2007. doi: 10.1073/pnas.0704665104. URL https://www.pnas.org/doi/abs/10.1073/pnas.0704665104. _eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.0704665104.

Kristian Cibulskis, Michael S Lawrence, Scott L Carter, Andrey Sivachenko, David Jaffe, Carrie Sougnez, Stacey Gabriel, Matthew Meyerson, Eric S Lander, and Gad Getz. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3):213–219, 2013. ISSN 1546-1696. doi: 10.1038/nbt.2514. URL https://doi.org/10.1038/nbt.2514.

Louise de Schaetzen van Brienen, Maarten Larmuseau, Kim Van der Eecken, Frederic De Ryck, Pauline Robbe, Anna Schuh, Jan Fostier, Piet Ost, and Kathleen Marchal. Comparative analysis of somatic variant calling on matched FF and FFPE WGS samples. *BMC Medical Genomics*, 13 (1):94, July 2020. ISSN 1755-8794. doi: 10.1186/s12920-020-00746-5. URL https://doi.org/10.1186/s12920-020-00746-5.

Miklos Diossy, Zsofia Sztupinszki, Marcin Krzystanek, Judit Borcsok, Aron C Eklund, István Csabai, Anders Gorm Pedersen, and Zoltan Szallasi. Strand Orientation Bias Detector to determine the probability of FFPE sequencing artifacts. *Briefings in Bioinformatics*, 22(6), May 2021. ISSN 1477-4054. doi: 10.1093/bib/bbab186. URL https://doi.org/10.1093/bib/bbab186. _eprint: https://academic.oup.com/bib/article-pdf/22/6/bbab186/41087878/bbab186.pdf.

Dollina D. Dodani, Matthew H. Nguyen, Ryan D. Morin, Marco A. Marra, and Richard D. Corbett. Combinatorial and Machine Learning Approaches for Improved Somatic Variant Calling From Formalin-Fixed Paraffin-Embedded Genome Sequence Data. *Frontiers in Genetics*, 13, 2022. ISSN 1664-8021. doi: 10.3389/fgene.2022.834764. URL https://www.frontiersin.org/articles/10.3389/fgene.2022.834764.

Erik Garrison and Gabor Marth. *Haplotype-based variant detection from short-read sequencing*. 2012. _eprint: 1207.3907.

Weitai Huang, Yu Amanda Guo, Karthik Muthukumar, Probhonjon Baruah, Mei Mei Chang, and Anders Jacobsen Skanderup. SMuRF: Portable and accurate ensemble prediction of somatic mutations. *Bioinformatics (Oxford, England)*, January 2019. ISSN 1367-4811. doi: 10.1093/bioinformatics/btz018. URL http://www.ncbi.nlm.nih.gov/pubmed/30649191.

Mamoru Kato, Hiromi Nakamura, Momoko Nagai, Takashi Kubo, Asmaa Elzawahry, Yasushi Totoki, Yuko Tanabe, Eisaku Furukawa, Joe Miyamoto, Hiromi Sakamoto, Shingo Matsumoto, Kuniko Sunami, Yasuhito Arai, Yutaka Suzuki, Teruhiko Yoshida, Katsuya Tsuchihara, Kenji Tamura, Noboru Yamamoto, Hitoshi Ichikawa, Takashi Kohno, and Tatsuhiro Shibata. A computational tool to detect DNA alterations tailored to formalin-fixed paraffin-embedded samples in cancer clinical sequencing. *Genome Medicine*, 10(1):44, June 2018. ISSN 1756-994X. doi: 10.1186/s13073-018-0547-0. URL https://doi.org/10.1186/s13073-018-0547-0.

Sangtae Kim, Konrad Scheffler, Aaron L. Halpern, Mitchell A. Bekritsky, Eunho Noh, Morten Källberg, Xiaoyu Chen, Yeonbin Kim, Doruk Beyter, Peter Krusche, and Christopher T. Saunders. Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods*, 15(8):591–594, August 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0051-x. URL

https://www.nature.com/articles/s41592-018-0051-x. Number: 8 Publisher: Nature Publishing Group.

Daniel C Koboldt, Qunyuan Zhang, David E Larson, Dong Shen, Michael D McLellan, Ling Lin, Christopher A Miller, Elaine R Mardis, Li Ding, and Richard K Wilson. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*, 22(3):568–576, 2012. Publisher: Cold Spring Harbor Lab.

Kiran Krishnamachari, Dylan Lu, Alexander Swift-Scott, Anuar Yeraliyev, Kayla Lee, Weitai Huang, Sim Ngak Leng, and Anders Jacobsen Skanderup. Accurate somatic variant detection using weakly supervised deep learning. *Nature Communications*, 13(1):4248, July 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-31765-8. URL https://doi.org/10.1038/s41467-022-31765-8.

Zhongwu Lai, Aleksandra Markovets, Miika Ahdesmaki, and Justin Johnson. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Cancer Research*, 75(15 Supplement):4864 LP – 4864, August 2015. doi: 10.1158/1538-7445.AM2015-4864. URL http://cancerres.aacrjournals.org/content/75/15_Supplement/4864.abstract.

Tomas Lindahl and Barbro Nyberg. Rate of depurination of native deoxyribonucleic acid. *Biochemistry*, 11(19):3610–3618, September 1972. ISSN 0006-2960. doi: 10.1021/bi00769a018. URL https://doi.org/10.1021/bi00769a018. Publisher: American Chemical Society.

Ensel Oh, Yoon-La Choi, Mi Jeong Kwon, Ryong Nam Kim, Yu Jin Kim, Ji-Young Song, Kyung Soo Jung, and Young Kee Shin. Comparison of Accuracy of Whole-Exome Sequencing with Formalin-Fixed Paraffin-Embedded and Fresh Frozen Tissue Samples. *PLOS ONE*, 10 (12):1–13, December 2015. doi: 10.1371/journal.pone.0144162. URL https://doi.org/10.1371/journal.pone.0144162. Publisher: Public Library of Science.

Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin A. Riedmiller. Striving for Simplicity: The All Convolutional Net. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6806.

Mythily Srinivasan, Daniel Sedmak, and Scott Jewell. Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids. *The American Journal of Pathology*, 161(6): 1961–1971, December 2002. ISSN 0002-9440. doi: 10.1016/S0002-9440(10)64472-0. URL https://doi.org/10.1016/S0002-9440(10)64472-0. Publisher: Elsevier.

Tim A Steiert, Genís Parra, Marta Gut, Norbert Arnold, Jean-Rémi Trotta, Raúl Tonda, Alice Moussy, Zuzana Gerber, Peter M Abuja, Kurt Zatloukal, Christoph Röcken, Trine Folseraas, Marit M Grimsrud, Arndt Vogel, Benjamin Goeppert, Stephanie Roessler, Sebastian Hinz, Clemens Schafmayer, Philip Rosenstiel, Jean-François Deleuze, Ivo G Gut, Andre Franke, and Michael Forster. A critical spotlight on the paradigms of FFPE-DNA sequencing. *Nucleic Acids Research*, pp. gkad519, June 2023. ISSN 0305-1048. doi: 10.1093/nar/gkad519. URL https://doi.org/10.1093/nar/gkad519.

Maitena Tellaetxe-Abete, Borja Calvo, and Charles Lawrie. Ideafix: a decision tree-based method for the refinement of variants in FFPE DNA sequencing data. *NAR genomics and bioinformatics*, 3(4):lqab092, December 2021. ISSN 2631-9268. doi: 10.1093/nargab/lqab092. Place: England.

Shawn E. Yost, Erin N. Smith, Richard B. Schwab, Lei Bao, HyunChul Jung, Xiaoyun Wang, Emile Voest, John P. Pierce, Karen Messer, Barbara A. Parker, Olivier Harismendy, and Kelly A. Frazer. Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. *Nucleic Acids Research*, 40(14):e107–e107, April 2012. ISSN 0305-1048. doi: 10.1093/nar/gks299. URL https://doi.org/10.1093/nar/gks299._eprint: https://academic.oup.com/nar/article-pdf/40/14/e107/25343085/gks299.pdf.

A Appendix

A.1 Training hyperparameters

Table 3: Hyperparameters search space for training FFPENet model.

Parameter	**Random distribution**
Learning rate	$10^{\text{Uniform}(-5.5,-2)}$
Fine-tune dense layers only	RandomChoice($[false, true]$)
Number of dense layers to finetune	RandomChoice($[1, 2, 3, 4]$)
Reinitialize weights of fine-tuned layers	RandomChoice($[false, true]$)
Optimizer	RandomChoice($[Adam, SGD]$)
Weight decay	$10^{\text{Uniform}(-6,-3)}$
Batch size	$2^{\text{Uniform}(4,7)}$
Label smoothing parameter	Uniform$(0, 0.1)$