 doi:10.20944/preprints202103.0676.v1

Article

MET exon 14 skipping: a case study for the detection of genetic variants in cancer driver genes by deep learning.

Vladimir Nosi [1], Alessandrì Luca [1], Melissa Milan [2], Maddalena Arigoni [1], Silvia Benvenuti [2], Davide Cacchiarelli [3], Marcella Cesana [3], Sara Riccardo [3], Lucio Di Filippo [3], Francesca Cordero [4], Marco Beccuti [4], Paolo M Comoglio [5,†] and Raffaele A Calogero [1,*,†]

[1] Dept. of Molecular Biotechnology and Health Sciences, University of Torino, Italy 1; vladimir.nosi@unito.it, alessandri.luca1991@gmail.com, maddalena.arigoni@unito.it , raffaele.calogero@unito.it
[2] Candiolo Cancer Institute-FPO, IRCCS, Candiolo, Italy; melissa.milan@ircc.it, silvia.benvenuti@ircc.it
[3] TIGEM Telethon Institute of Genetics and Medicine, Italy; d.cacchiarelli@tigem.it, m.cesana@tigem.it, sara.riccardo@ngdx.eu, lucio.difilippo@ngdx.eu
[4] Dept. of Computer Sciences, University of Torino, Italy; francesca.cordero@unito.it, marco.beccuti@unito.it
[5] IFOM-FIRC Institute of Molecular Oncology, Italy; pcomoglio@gmail.com
* Correspondence: raffaele.calogero@unito.it
† Both equally supervised the present work.

Abstract: Background: Disruption of alternative splicing (AS) is frequently observed in cancer and it might represent an important signature for tumor progression and therapy. Exon skipping (ES) represents one of the most frequent AS events and in non-small cell lung cancer (NSCLC) MET exon 14 skipping was shown to be targetable. Methods: We constructed a neural network (NN) specifically designed to detect MET exon 14 skipping events using RNAseq data. Furthermore, for discovery purpose we also developed a sparsely connected autoencoder to identify uncharacterized MET isoforms. Results: The NN had 100% Met exon 14 skipping detection rate, when tested on a manually curated set of 690 TCGA bronchus and lung samples. When globally applied to 2605 TCGA samples, we observed that the majority of false positives was characterized by a blurry coverage of exon 14, but interesting they share a common coverage peak in the second intron and we speculate that this event could be the transcription signature of a LINE1-MET fusion. Conclusions: Taken together our results indicate that neural networks can be an effective tool to provide a quick classification of pathological transcription events and sparsely connected autoencoders could represent the basis for the development of an effective discovery tool.

Keywords: Neural network; MET; exon skipping

Citation: Nosi, V.; Alessandrì Luca; Milan, M.; Arigoni, M.; Benvenuti, S.; Cacchiarelli, D.; Cesana, M.; Riccardo, S.; Filippo, L.D.; Cordero, F.; et al. MET exon 14 skipping: a case study for the detection of genetic variants in cancer driver genes by deep learning. *Int. J. Mol. Sci.* 2021, *22*, x. https://doi.org/10.3390/xxxxx

1. Introduction

It is known that in eukaryotes, alternative splicing plays an important role in defining the protein diversity and enhancing the complexity of gene expression regulation [1]. In humans, the majority of multi-exon genes is affected by alternative splicing, which generates proteins with different functions in distinct cellular processes [2]. Disruption of alternative splicing (AS) is associated with human diseases [3] and exon skipping (ES) is one of the most observed events [4]. The analyses and studies of alternative splicing advance our understanding of mRNA complexity and its regulation, providing valuable insights to grasp disease etiology, and assisting the development of therapeutic interventions for splicing-related diseases [5]. It has been recently developed ExonSkipDB (https://ccsm.uth.edu/ExonSkipDB/) [6], which is a database collecting ES events affecting disease associated genes. Within the 8266 ExonSkipDB genes, annotated as genes loosing functional features due to in-frame ES events in TCGA (https://portal.gdc.cancer.gov/), 449 are part of the 710 COSMIC census genes [7]. 25 of them (ALK, APC, BAP1, BRCA1, BRCA2, BRIP1, BTK, CDH1, CHEK2, ERBB2, ETV6, EXT1, EZH2, GLI1, JAK2, MDM4, MET, MLH1, MUTYH, NF2, NOTCH2, PIK3R1, PTCH1, SUFU, TP53) have at least an

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2021 doi:10.20944/preprints202103.0676.v1

exon skipping event associated with cancer phenotype reported in at least a published paper (Supplementary Table 1S). Notably, MET exon 14 skipping is the only ES event encompassing a massive number of citations (119 from 2015 to 2021 reported in the PUBMED repository). Champagnac and coworkers [8] observed that genomic alterations affecting MET exon 14 are present in 2.6% of non-small cell lung cancer (NSCLC) patients. MET exon 14 skipping can lead to acquisition of transforming ability and has been identified as potential therapeutic target for NSCLC [9]. Many different mutations at DNA level can cause the aberrant splicing of exon 14, and the only search at genomic level for MET exon 14 skipping does not guarantee that the mutated MET transcript is actively expressed. Furthermore, given the relatively small deletion, it remains a question whether antibodies can be developed with enough specificity against this splice variant [10]. RNA sequencing is today a straightforward approach thanks to the possibility to perform targeted RNAseq in paraffine embedded samples [11]. However, to efficiently detect MET exon 14 skipping an effective computing detection algorithm for this specific ES event is also required. In this manuscript, we describe the development and validation of a neural network (NN) specifically devoted to the effective and rapid detection of MET exon 14 skipping events using RNAseq data.

2. Results

2.1. Neural network for the detection of MET exon 14 skipping (METΔ14).

To detect MET exon 14 skipping events, a NN made of six layers was built, Figure 1A.

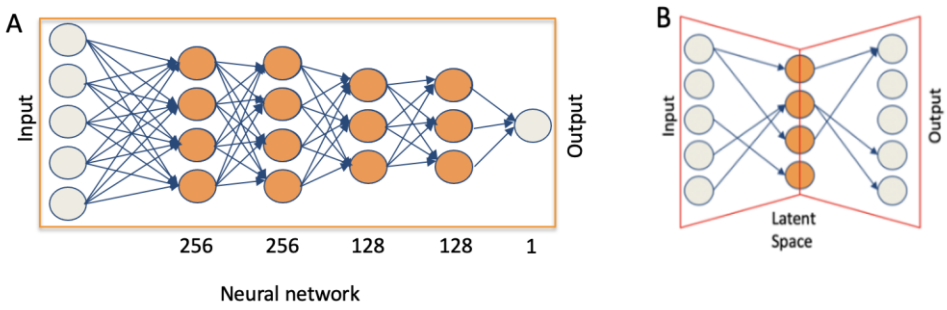

Figure 1. Neural networks used for the analysis of MET variants. A) NN for the detection of exon 14 skipping events. B) Sparsely-connected autoencoders for the detection of MET transcription variants.

As training set for the NN we used data from amplified WT MET and exon 14 skipping MET (Table 1).

Table 1. MET cell line RNAseq data.

Cell line	Status	RNAseq (Million reads)	MET (Thousand reads)
EBC-1	Amplified MET	113	1447
Hs746T	Amplified METΔ14	95	846
A549	MET	115	109
NCI-H596	METΔ14	118	114

Specifically, we split the MET reads in random non-overlapping subgroups of 1000 reads. Although, at 1000 reads coverage the detection of METΔ14 becomes a bit blurry, Figure 2 (d), this threshold allows a generation of a large number of MET (1447) and METΔ14 (846) not overlapping subsamples, and high numerosity of training data is an important element for an efficient learning of the NN.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2021 doi:10.20944/preprints202103.0676.v1

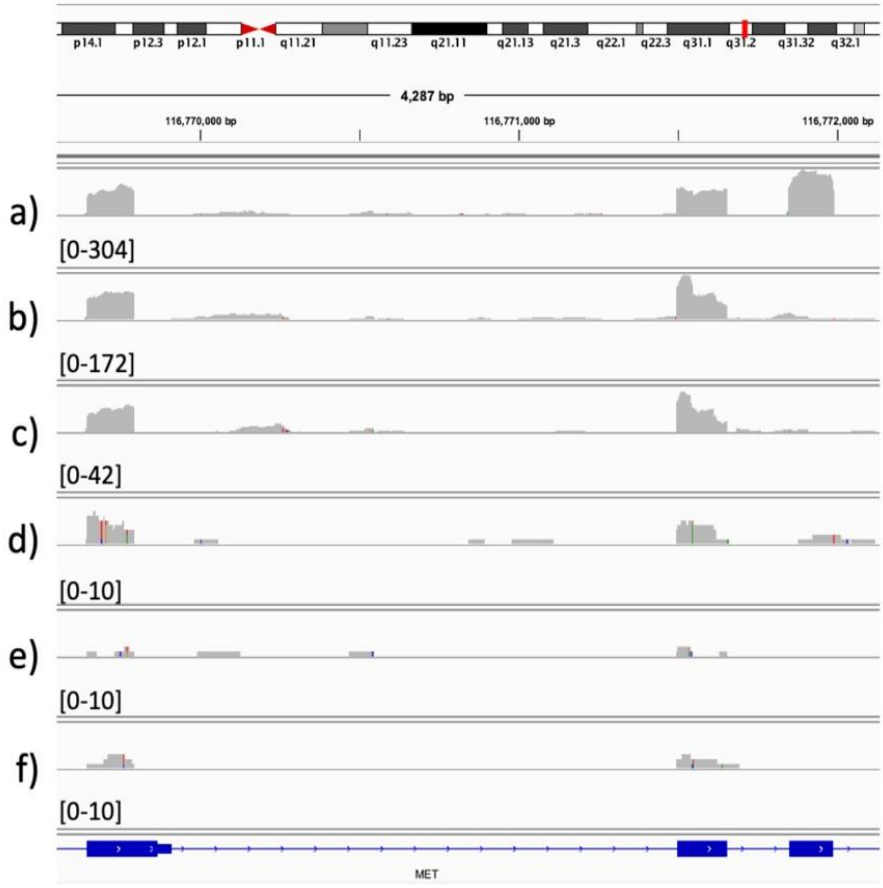

Figure 2: Expected coverage for exons 13, 14 and 15. a) WT MET from A549 RNAseq sample (33 million reads), 27152 reads mapping on MET locus, b) METΔ14 from NCI-H596 RNAseq sample (27 million reads), 24850 reads mapping on MET locus, c) 5000 reads randomly selected from (b), d) 1000 reads randomly selected from (b), e) 500 reads randomly selected from (b), f) 250 reads randomly selected from (b).

Each of the above-mentioned subgroups was converted in 31 and 16 k-mers. MET expression was represented by the amount of each k-mers spanning over MET exons and these data were used to train the NN. We observed that the learning curve at 16 k-mers was slightly better than the one at 31 k-mers (not shown), thus we run the following analyses using the 16 k-mers representation of MET. As training sets, we also used k-mer count frequency [17] for full MET locus, k-mer count frequency for MET exons 13÷15 and coverage frequency for MET exons 13÷15.

As test sets, we used subsets of WT and exon 14 skipping MET from cell lines characterized by a physiological MET expression. NN performance was investigated using as test set: i) subsets made of random not overlapping subgroups of 500, 1000 and 5000 reads, converted in 16 k-mer counts, ii) k-mer count frequency on full MET locus, iii) k-mer count frequency on MET exons 13÷15 and iv) coverage frequency for MET exons 13÷15.

The detection efficiency of METΔ14 using 16 k-mers frequency counts showed best performances at 500 and 1000 reads coverage, Figure 3A,B, as instead at 5000 reads coverage all the different test sets performed in the same way, Figure 3C.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2021 doi:10.20944/preprints202103.0676.v1

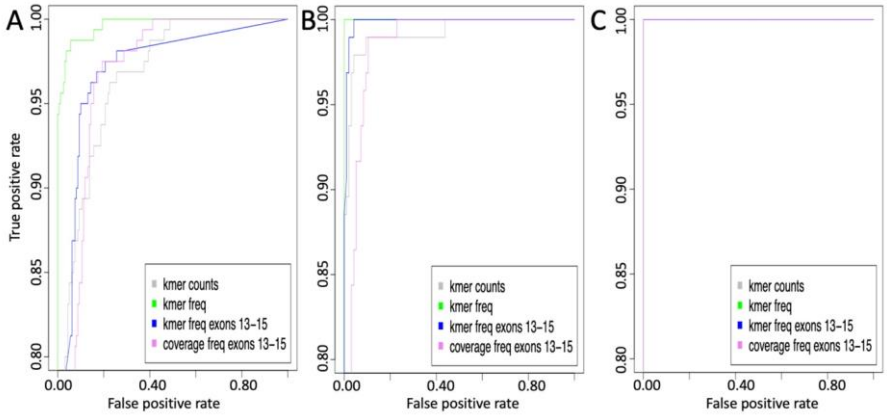

Figure 3. ROC curve for NN prediction. A) Training based on 1000 reads coverage for MET locus and test with a coverage of 500 reads. B) Training based on 1000 reads coverage for MET locus and test with a coverage of 1000 reads. C) Training based on 1000 reads coverage for MET locus and test with a coverage of 5000 reads. Grey line training and test using k-mer counts, green line training and test using k-mer counts frequency, blue line training and test using k-mer counts frequency only for exons 13, 14 and 15, violet line training and test using coverage counts frequency only for exons 13, 14 and 15.

2.2. Neural network validation and discovery on TCGA samples.

To validate the METΔ14 discovery potential of the above-described NN, we used a set of 690 RNAseq samples from the TCGA bronchus and lung dataset. The 690 samples were manually inspected using the Broad's integrative genomics viewer [18] and we detected 17 exon 14 skipping events (2.4%), which is in line with the frequency of the exon 14 skipping events observed by Champagnac [8]. We tested on this tumor set the NN trained with k-mer counts frequency, which predicted 4 samples out of 17 as METΔ14, but only one was a real exon skipping events (sensitivity 5.88%, specificity 99.5%, supplementary table 2S). The NN trained with exons 13÷15 MET k-mer counts frequency improved the detection of METΔ14 events, 9 out of 17 (sensitivity 52.9%), but this prediction included a massive increase of false positives, 129 samples, (specificity 81.3%, supplementary table 2S). The best results were obtained using the NN trained using only the coverage frequency for MET exons 13÷15, which predicted 18 skipping events, including all 17 true skipping events (sensitivity 100%) and one false positive (specificity 99.8%, supplementary Table 2S).

Using the NN trained with the coverage frequency for MET exons 13÷15, we extended the METΔ14 discovery to 2605 TCGA tissues, Table 2.

Table 2. TCGA samples inspected for the presence of METΔ14.

TCGA tissue	# inspected tissue	# detected METΔ14	# detected false METΔ14
Adrenal gland	10	0	0
Bladder	280	1	0
Brain	28	0	0
Breast	162	0	0
Bronchus and lung	690	17	1
Cervix (uterus)	236	0	6
Corpus uteri	109	0	4
Esophagus	165	0	0
Hearth/mediastinum/pleura	78	0	1
Kidney	435	0	3
Pancreas	89	0	0
Skin	288	0	1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2021 doi:10.20944/preprints202103.0676.v1

Soft tissues	35	0	0

We could detect only one METΔ14 in 280 bladder samples. Then, we detected few false METΔ14 in cervix, corpus uteri, heart/mediastinum/pleura, kidney and skin samples, Table 2. The six transcripts detected in cervix, Table 2, were erroneously detected as METΔ14, because they have a blurry coverage on exons 13÷15, Figure 4A. However, when the full MET locus is observed, Figure 4B, it is clear that these METΔ14 false positives are a completely different type of transcripts. A shared characteristic of these transcripts is the high accumulation of reads in the second intron (approx. chr7:116,715,690-116,717,329), in the 6th exon and in the last non-coding MET exon, Figure 4B.

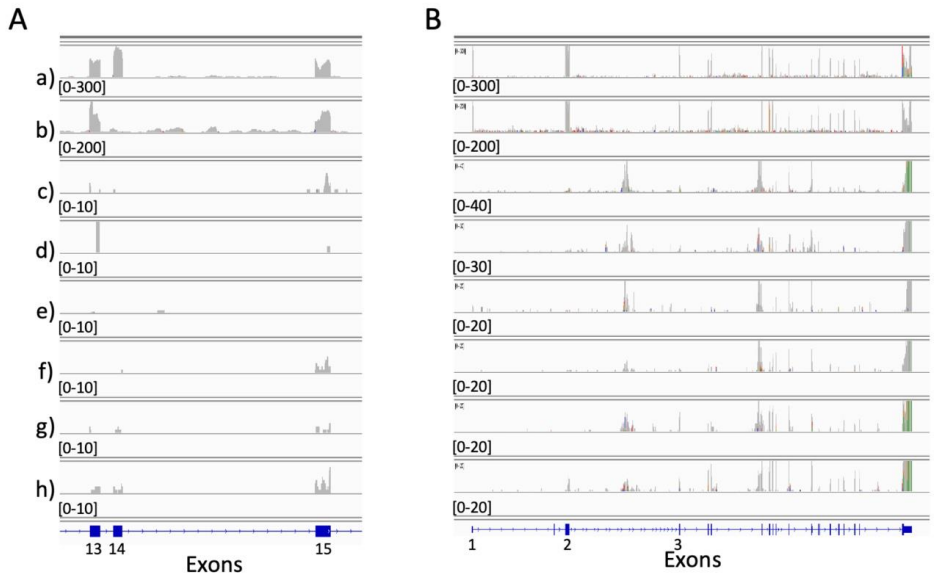

Figure 4. METΔ14 false positive detected in cervix. a) WT MET from A549 RNAseq sample (33 million reads), 27152 reads mapping on MET locus, b) METΔ14 from NCI-H596 RNAseq sample (27 million reads), 24850 reads mapping on MET locus, c-h) False METΔ14. A) Zoom in the 13-15 exons region. B) Full Met locus.

The above observation also applies to the other false METΔ14 detected in corpus uteri, heart/mediastinum/pleura, kidney and skin samples, supplementary Figure 1S.

A possible explanation could be that we are observing the transcriptional effect of a LINE1-MET fusion, which was firstly described few years ago in triple negative breast cancers [19]. We further investigated this point searching for LINE1 alignment, in the subset of MET reads, where only one of the two pair-end reads maps on MET. Indeed, in 10 out of 15 samples, detected as characterized by a transcription peak in MET second intron, we detected LINE1 mapping reads, Supplementary Table 4S. From the samples shown in Supplementary Table 4S, we extracted the paired reads associated with MET reads, i.e. only one read of the pair is mapping in MET locus. We blasted [20] these reads on a LINE1 sequence (chr1:62194249-62212928, hg38) and indeed, some of these reads map to LINE1 sequence, supplementary Table 5S. On the basis of the MET read position we could identify the putative fusion point with MET, which is mainly located in MET intronic regions and in the last non-coding exon. Unfortunately, we cannot pair the TCGA RNAseq samples to genomics data to further validate the presence of a LINE1 insertion on the basis of genome sequencing data.

2.3. Sparsely connected autoencoders (SCA) to detect MET non-canonical isoforms

Our group has recently published a paper on the use of SCA for the identification of hidden functional regulatory elements in single cell RNAseq data [21]. We tested this type of autoencoder to see if we could grasp non-canonical isoforms from the analysis of the TCGA samples used in the previous paragraph. The SCA was designed to take as input

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2021 doi:10.20944/preprints202103.0676.v1

k-mer count frequency or coverage frequency of MET exons. The SCA hidden layer, i.e. latent space, is representing MET exons. Input nodes are only connected to the exon nodes they are associated, Figure 1B. We trained the SCA with the 2605 TCGA samples and we clustered the latent space data using gridFLOW [22]. To estimate the stability of clusters, generated using the SCA latent space, we compared thousands of pairs of clusters generated by SCA latent space clustering, as previously described by us [21]. The rational of this approach is that, if a clusters organization is conserved, it should be depicted by the multiple comparisons of randomly paired latent space cluster representations [21]. The best results were obtained using normalized [23] MET coverage frequency data, Figure 5A. Unfortunately, the stability of the clusters was very poor, Figure 5A. However, an inspection of a random subsets of samples associated with cluster 2, Figure 5B, suggests that at least cluster 2 seems to be made mainly of transcripts recalling the organization of LINE1-MET fusion, which we have described in the previous paragraph.

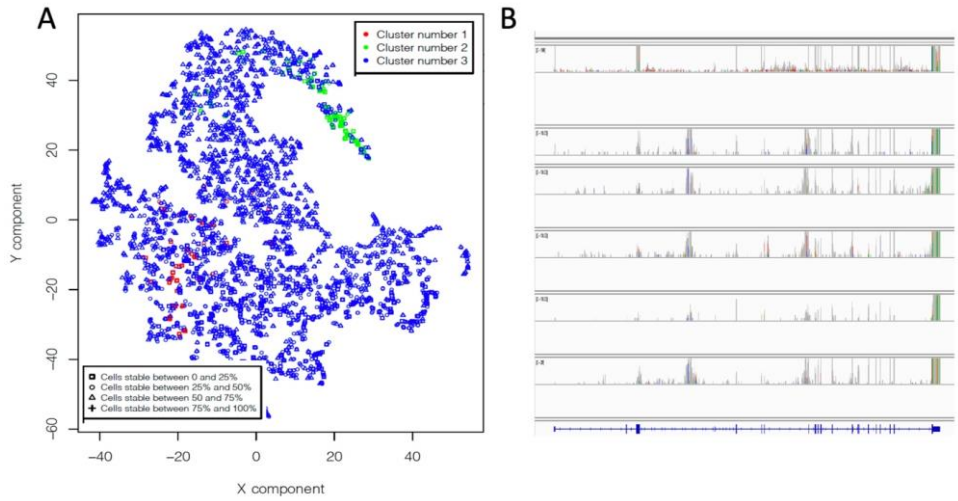

Figure 5. Autoencoder on saver normalized data. A) Clustering results of the latent space trained with 2605 TCGA samples. B) A limited number of samples (green group in A) is characterized the presence of a transcription patter, i.e. the coverage peak in intron 2, which resembles the presence of a LINE1-MET fusion. In B, it is shown a set of samples randomly picked from cluster 2.

3. Discussion

We used MET exon 14 skipping as a case study for the detection of genetic variants in cancer driver genes by deep learning. In recent years, a lot of evidences indicates that MET inhibitors have a good anti-tumor effect in patients with MET exon 14 skipping mutation, suggesting that MET exon 14 skipping may be a new target for NSCLC patients [24]. Thus, the availability of effective tools for the detection of MET exon 14 skipping are needed for a fast identification of patients suitable for MET targeted therapy. Here, we show that we can identify with high sensitivity MET exon 14 skipping event, using a neural network trained with RNAseq coverage data encompassing the region between MET exon 13 and exon 15. Our analysis of a subset of TCGA samples highlights that MET exon 14 skipping is a peculiar event of lung specimens. Then, mainly in uterine cancers, we detected a set of MET exon 14 skipping false positives, sharing a common feature: an unexpected peak of coverage in the MET intron 2. This observation brought us to speculate that we were observing a transcriptional signature for a LINE1-MET fusion event [19]. This hypothesis, has been supported by the identification of MET paired-end reads, having one read mapping on MET and the other on LINE1 sequence. Notably, transcription of the LINE1-MET fusion was observed in advanced stages of cancer [19, 25], but very little is still known about the effect of the LINE1-MET chimera in cancer. Having identified more than one artifactual event in MET, we investigated the possibility to discover those anomalous events by the integration of a particular type of deep learning tool, sparsely connected autoencoders [21], with clustering techniques used in multicolor cytometry.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 March 2021 doi:10.20944/preprints202103.0676.v1

Although we have to further improve its precision and sensitivity, we were able to detect from TCGA specimens a set of tumors sharing the putative LINE1-MET fusion.

4. Materials and Methods

4.1. Generating the data for the neural network training and test set

We have generated RNAseq data from EBC-1 [12], a non-small cell lung cancer (NSCLC) cell line, harboring MET amplification and from Hs746T, a gastric cancer cell line, harboring amplified MET exon 14 skipped isoform (METΔ14) [13]. Furthermore, we have performed RNAseq on human lung adenocarcinoma cell line A549, expressing c-Met [14] and on NCI-H596, derived from an NSCLC, expressing exon 14 skipped MET [15], Table 1. Both cell lines express physiological levels of MET. Total RNA was extracted from cell lines using Trizol reagent (Invitrogen), following the manufacturer indication. Total RNA quality was evaluated using Bioanalyser (Agilent). Total RNA was quantified using Qbit (Thermo Fisher Scientific). RNAseq was done using TruSeq library kit version 2 (Illumina), following the manufacturer indications for the generation of polyA enriched transcripts. Libraries were sequenced on Novaseq (Illumina) as 150 nts paired end protocol.

4.1.1.16/31. k-mer training set

The training set for the neural network (NN) was generated using the cell lines with MET amplification: EBC-1 and Hs746T. EBC-1 and Hs746T reads were organized in subgroups of 1000 reads, randomly selected and not overlapping. This approach generated a large set of samples for the NN training, i.e. 1447 subsets for EBC-1 and 846 for Hs746T. Subsampled reads were associated with MET exons (supplementary Table 3S) and converted in 16/31 k-mers using BFcounter [16].

4.1.2.16/31. k-mer test set

The test set for the neural network (NN) was generated using the cell lines with physiological MET expression: A549 and NCI-H596. A549 and NCI-H596 reads were organized in subgroups of 500, 1000 and 5000 reads, randomly selected and not overlapping. Subsampled reads were associated to MET exons (supplementary Table 3S) and converted in 16/31 k-mers using BFcounter [16].

4.1.3. Coverage training and test set

The training and test set for the neural network (NN) were generated using the RNAseq data used for the 16/31 k-mers training and test sets. For the training, reads were organized in subgroups of 1000 reads, randomly selected and not overlapping, for the test set, reads were organized in subgroups of 500, 1000 and 5000 reads. Subsampled reads were used to calculate coverage associated with MET exons 13, 14 and 15.

4.2. TCGA RNAseq datasets

We registered to TCGA a project for the study of MET exon 14 skipping events, to obtain access to TCGA raw sequencing data, i.e. RNAseq BAM files. Since the size of the TCGA transcription data exceeds 200 TB, we progressively downloaded the BAM files on the basis of the cancer tissue locus. Then, from each BAM file, we extracted the reads encompassing MET locus (chr7:116672196-116798377, hg38 human genome assembly). We kept only samples where the MET locus was covered by at least 5000 reads. To define 5000 reads as the minimal coverage for MET, we checked the expected coverage for exons 13, 14 and 15 in A549 (WT MET cell line), in NCI-H596 (METΔ14 cell line) and in random subsets of 5000, 1000, 500 and 250 reads from NCI-H596, Fig. 2. We observed that the detection of exon 14 skipping become blurry below 5000 reads coverage. Together with the MET linked reads we also extracted the MET paired reads, where only one of the two reads maps on MET locus.

 doi:10.20944/preprints202103.0676.v1

4.3. Model coding and hyperparameter selection for NN

We constructed a NN made of 6 layers. The input layer has variable size depending on the type of input (k-mers or coverage). 1st and 2nd hidden layers are made of 256 nodes, 3rd and 4th are made of 128 nodes, all using RELU (rectified linear unit) as activation function and 0.1 as dropout rate. The output layer is made by 1 node, associated with a sigmoid activation function. We implemented the models in python (version 3.7) using TensorFlow package (version 2.0.0), Keras (version 2.3.1), pandas (version 0.25.3), numpy (version 1.17.4), matplotlib (version 3.1.2), sklearn (version 0.22), scipy (version 1.3.3). Optimization was done using Adam (Adaptive moment estimation), with the following parameters lr=0.01, beta_1=0.9, beta_2=0.999, epsilon=1e−08,decay=0.0, loss='mean_squared_error'.

The trained NN is implemented in a docker container together with all tools needed to extract MET reads from fastq data. The NN can be used for the discovery of METΔ14 using conventional RNAseq or MET targeted RNAseq. The tool is provided free of charge to Accademia and non-profit organizations for research use only.

4.4. Model coding and hyperparameter selection for Sparsely Connected autoencoders (SCA)

Autoencoders learning is based on an encoder function that projects input data onto a lower dimensional space. Then, autodecoder function recovers the input data from the low-dimensional projections minimizing the reconstruction. We implemented the models in python (version 3.7) using TensorFlow package (version 2.0.0), Keras (version 2.3.1), pandas (version 0.25.3), numpy (version 1.17.4), matplotlib (version 3.1.2), sklearn (version 0.22), scipy (version 1.3.3). Optimization was done using Adam (Adaptive moment estimation) with the following parameters lr=0.01, beta_1=0.9, beta_2=0.999, epsilon=1e−08, decay=0.0, loss='mean_squared_error'. RELU (rectified linear unit) was used as activation function for the dense layer.

5. Conclusions

Taken together our results indicate that neural networks can be an effective tool to provide a quick classification of pathological transcription events and sparsely connected autoencoders could represent the basis for the development of an effective discovery tool in this field.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: METΔ14 false positive detected in corpus uteri. a) WT MET from A549 RNAseq sample (33 million reads), 27152 reads mapping on MET locus, b) METΔ14 from NCI-H596 RNAseq sample (27 million reads), 24850 reads mapping on MET locus, c-f) False METΔ14 in corpus uteri samples, g) False METΔ14 in heart/mediastinum/pleura samples, h-l) False METΔ14 in kidney samples, m) False METΔ14 in skin samples. Table 1S: Set of COSMIC census genes present in the ExonSkipDB. The first column is the set of genes associated to articles describing the presence of a skipping event in that gene linked to cancer. Table 2S: samples in TCGA bronchus and lung predicted as METΔ14 using different training configurations. METΔ14 score ≤ 0.1 predicts a skipped event. NN quality score ranges between 1 and 0, where 1 indicates an optimal NN performances. Table 3S: MET exons. Table 4S: Samples characterized by unexpected coverage peak in intron 2. Table 5S: Samples characterized by the presence of a MET read(s) paired with a read mapping in LINE1 sequence.

Author Contributions: VN and LA equally worked at the developed of the neural network and of the autoencoder. DC, MC, SR, LD provided the RNA sequencing data used for training and testing the neural network. FC and MB optimized the autoencoder parameters and tested it. MA, MM and BS retrieved and processed TCGA data. PMC and RAC supervised the work and wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) (AIRC-IG, Ref: 23820 to PMC).

Data Availability Statement: the NN is available as tool at docker.io/repbioinfo/metobservatory.2021.01. Examples and instructions for its usage are available at https://github.com/kendomaniac/metObservatory. RNAseq used for training and test data are available at https://github.com/kendomaniac/metObservatory.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Graveley, B.R., Alternative splicing: increasing diversity in the proteomic world. Trends Genet, 2001. 17(2): p. 100-7.
2. Pan, Q., et al., Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet, 2008. 40(12): p. 1413-5.
3. Tazi, J., N. Bakkour, and S. Stamm, Alternative splicing and disease. Biochim Biophys Acta, 2009. 1792(1): p. 14-26.
4. Florea, L., L. Song, and S.L. Salzberg, Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Res, 2013. 2: p. 188.
5. Jiang, W. and L. Chen, Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. Comput Struct Biotechnol J, 2021. 19: p. 183-195.
6. Kim, P., et al., ExonSkipDB: functional annotation of exon skipping event in human. Nucleic Acids Res, 2020. 48(D1): p. D896-D907.
7. Sondka, Z., et al., The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer, 2018. 18(11): p. 696-705.
8. Champagnac, A., et al., Frequency of MET exon 14 skipping mutations in non-small cell lung cancer according to technical approach in routine diagnosis: results from a real-life cohort of 2,369 patients. J Thorac Dis, 2020. 12(5): p. 2172-2178.
9. Awad, M.M., et al., MET Exon 14 Mutations in Non-Small-Cell Lung Cancer Are Associated With Advanced Age and Stage-Dependent MET Genomic Amplification and c-Met Overexpression. J Clin Oncol, 2016. 34(7): p. 721-30.
10. Van Der Steen, N., et al., cMET Exon 14 Skipping: From the Structure to the Clinic. J Thorac Oncol, 2016. 11(9): p. 1423-32.
11. Marczyk, M., et al., The impact of RNA extraction method on accurate RNA sequencing from formalin-fixed paraffin-embedded tissues. BMC Cancer, 2019. 19(1): p. 1189.
12. Lutterbach, B., et al., Lung cancer cell lines harboring MET gene amplification are dependent on Met for growth and survival. Cancer Res, 2007. 67(5): p. 2081-8.
13. Asaoka, Y., et al., Gastric cancer cell line Hs746T harbors a splice site mutation of c-Met causing juxtamembrane domain deletion. Biochem Biophys Res Commun, 2010. 394(4): p. 1042-6.
14. Li, B., et al., Higher levels of c-Met expression and phosphorylation identify cell lines with increased sensitivity to AMG-458, a novel selective c-Met inhibitor with radiosensitizing effects. Int J Radiat Oncol Biol Phys, 2012. 84(4): p. e525-31.
15. Kong-Beltran, M., et al., Somatic mutations lead to an oncogenic deletion of met in lung cancer. Cancer Res, 2006. 66(1): p. 283-9.
16. Melsted, P. and J.K. Pritchard, Efficient counting of k-mers in DNA sequences using a bloom filter. BMC Bioinformatics, 2011. 12: p. 333.
17. Wen, J., et al., A classification model for lncRNA and mRNA based on k-mers and a convolutional neural network. BMC Bioinformatics, 2019. 20(1): p. 469.
18. Robinson, J.T., et al., Integrative genomics viewer. Nat Biotechnol, 2011. 29(1): p. 24-6.
19. Miglio, U., et al., The expression of LINE1-MET chimeric transcript identifies a subgroup of aggressive breast cancers. Int J Cancer, 2018. 143(11): p. 2838-2848.
20. Altschul, S.F., et al., Basic local alignment search tool. J Mol Biol, 1990. 215(3): p. 403-10.
21. Alessandri, L., et al., Sparsely-connected autoencoder (SCA) for single cell RNAseq data mining. NPJ Syst Biol Appl, 2021. 7(1): p. 1.
22. Ye, X. and J.W.K. Ho, Ultrafast clustering of single-cell flow cytometry data using FlowGrid. BMC Syst Biol, 2019. 13(Suppl 2): p. 35.
23. Huang, M., et al., SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods, 2018. 15(7): p. 539-542.
24. Huang, C., et al., Management of Non-small Cell Lung Cancer Patients with MET Exon 14 Skipping Mutations. Curr Treat Options Oncol, 2020. 21(4): p. 33.
25. Hur, K., et al., Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut, 2014. 63(4): p. 635-46.