ORIGINAL RESEARCH

lncRNA-based study of epigenetic regulations in diabetic peripheral neuropathy

Muhamad Fachrul[1] · Didik H. Utomo[2,3] · Arli A. Parikesit[1]

Abstract

Diabetes remains one of the most prevalent non-communicable diseases in the world, affecting over 400 million of people worldwide, causing serious complications leading to amputations and even death. Over the years, researchers have found that, in addition to genomic mutations, epigenetic mechanisms also play a role in the development of diabetes-specifically type-2 diabetes. Long noncoding RNAs (lncRNAs) have been linked to mediate epigenetic mechanisms, including those in late-stage diabetes. This study attempts to assess the unexplored topic of how lncRNAs could be used to assess the epigenetic mechanisms present in diabetic peripheral neuropathy (DPN); a serious complication of the disease often leading to amputation. Differential lncRNA expression analysis was done with a dataset containing DPN and healthy patients. Standard and corrected t test, and also LIMMA was applied. Results of this study indicates the usefulness of lncRNAs as an exploratory tool to elucidate the complexity of the epigenetic mechanisms of human DPN.

Keywords Epigenetic · lncRNA · In Silico · Type-2 diabetes

Introduction

For the past couple of decades, diabetes still reigns as one of the most prevalent non-communicable diseases in the world. The number of people affected by the condition has quadrupled since 1980—reaching up to 422 million people worldwide in 2014 (World Health Organization 2016b). The condition has been well-established to be a major cause of blindness, kidney failure, heart attacks, and lower limb amputation (World Health Organization 2016a). The link between diabetes and genetics has been established for at least a couple of decades; with multiple studies showing how even genomic mutations are a risk factor in type 2 diabetes mellitus (T2DM), despite it is believed being mostly caused by environmental factors (Dean and McEntyre 2004; Hara et al. 2014). However, it is becoming increasingly clear that genomics and proteomics-based points of views are insufficient to comprehend the holistic biomolecular mechanism of life (Parikesit et al. 2014). Thus, more feasible approach should be devised.

In recent years, researchers have found that heritable epigenetic mechanisms also have a role in the development of diabetes, specifically in T2DM (Al-Haddad et al. 2016; Karachanak-Yankova et al. 2015). The most common mechanisms include methylation, which generally results in gene silencing, and histone modifications, which either promotes or represses gene expression (Al-Haddad et al. 2016). Moreover, there are noticeable trends that the proteomics is being supplemented with transcriptomics point of view. In this end, the role of RNA-based regulation is increasingly important. Non-coding RNA that did not translated into protein, is considered playing an important role in the flow of genetic information (Amaral and Mattick 2008; Mattick 2005). Interestingly, some forms of ncRNAs—namely Micro RNAs (miRNA) and long noncoding RNAs (lncRNAs)—were found as regulatory factors in epigenetic mechanisms, including in late-stage diabetes (Pullen and Rutter 2014). Some examples include the lncRNAs HOTAIR and HOTTIP, which recruit the inhibitory polycomb repressive complex (PRC) 2 and the activating Trithotax/MLL

✉ Didik H. Utomo
   didik.huswo@gmail.com

1   Bioinformatics Department, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta, Indonesia

2   Biology Department, Faculty of Sciences, Brawijaya University, Malang, Indonesia

3   Research and Education Center for Bioinformatics, Nusantara Institute for Life Science and Technology, Jakarta, Indonesia

chromatin modifiers; PRC2 and MLL, which mark distinct lysine residues within histones via trimethylation; and also ANRIL, which silences the INK4a tumor suppressor allele through trimethylation (Kornfeld and Brüning 2014). Thus, the important role of histone mark as inseparable regulation factor in the development of the diseases could not be overruled, and this phenomenon already was confirmed with extensive computational studies (Prohaska et al. 2010).

One of the long-term complications of the disease is diabetic peripheral neuropathy (DPN), which is a condition where the peripheral nerves are damaged due to high chronic blood sugar, leading to foot ulceration, Charcot neuroarthropathy, and occasionally amputation (Boulton 2005). In this respect, epigenetics as the state-of-the-art biomarkers that did not undergone genetics-based mutation could play a role in the progression of this disease (Feinberg and Fallin 2015; Swami 2009). The biological mechanisms of the condition have been well-documented, although the role of epigenetics has only recently been suggested (Reddy et al. 2015). As recent development, epigenetics computation is still considered a very new field with limited number of experts for annotating the working database (Steiner et al. 2012). In this end, This study aims to elucidate the epigenetic mechanisms in DPN based on the lncRNA expression.

Methods

Data acquisition

The data series used for the analysis (accession: GSE95849) was acquired from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) from an unpublished diabetic study (Edgar et al. 2002). The expression profile was done on Phalanx Human lncRNA OneArray v1_mRNA platform, which includes 31,741 probesets relating to human mRNA and lncRNA; and used samples from 18 different participants, equally distributed between DPN, and healthy groups (Luo and Xu 2017).

Statistical analysis

Expression set creation and the whole analysis of the data series were done using R version 3.4.0 on RStudio (RStudio Team 2015). Bioconductor, an open-source analysis suite based on R language, was utilized to acquire the necessary analysis packages (Huber et al. 2015). GEOQuery package was installed to process the data series and create the expression set (Davis and Meltzer 2007).

After the expression set was created, normal t test was done between samples in healthy and DPN groups. Then, FDR and Bonferroni were applied to the t test. p value was adjusted to filter in only the most significantly differentially expressed genes. Limma (linear models for microarray data) package was also installed and used to do the differential expression quantification (Ritchie et al. 2015). The output was exported into a csv file, which was further analyzed using Microsoft Excel 2016. Information regarding the genes were taken from the GeneCards database (Belinky et al. 2013).

Visualization

A boxplot was generated from the normal t-test result. Volcano plots were generated for the t test results of both type of corrections. Heatmap was also generated using the results from limma.

Results and discussions

According to normal t test, 14,599 genes in total were found to be differentially expressed between healthy and D patients with (p value < 0.05). Adjusting the p value to under 0.0000001, 38 genes were still found to be differentially expressed (Fig. 1).

Using FDR adjustment, 17 genes were found to be differentially expressed (p value < 0.00005); two of which are partly non-coding, namely MTHFSD and LMAN2L. Using Bonferroni adjustment, only two genes were found to be differentially expressed under the same cut-off, which was expected since the method has been known to be stricter. Those two genes—GDAP2 and TBC1D24—were both considered coding (Fig. 2).
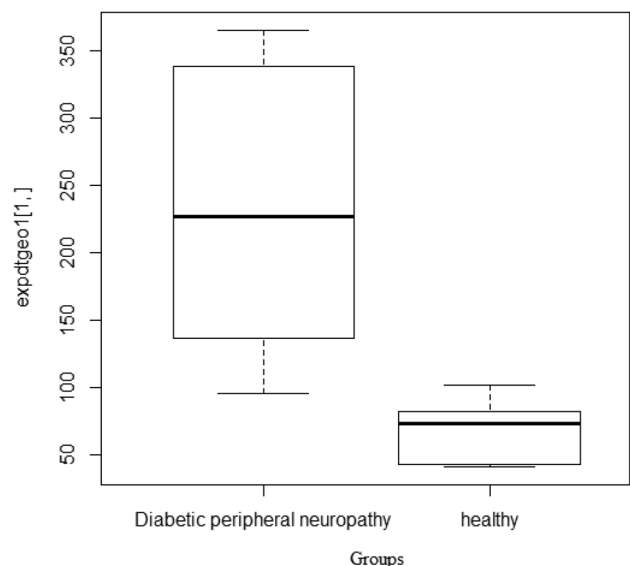
Fig. 1 Boxplot showing the significant difference between healthy and DPN patients
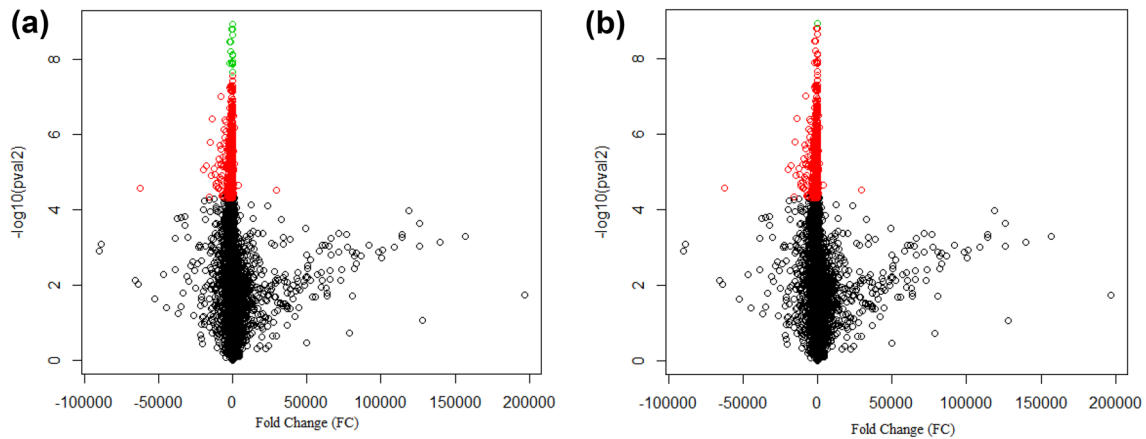
Fig. 2 Volcano plots of the differentially expressed genes from DPN and healthy group resulted from t-test using FDR correction (**a**) and Bonferroni correction (**b**). Red dots represent genes that found to be differentially expressed through normal t test, while green dots represent genes differentially expressed through the respective correction methods (p value < 0.00005)

In comparison, 22 genes were found to be differentially expressed using lncRNA (*p* value < 0.00005). As shown in Table 1 below, only three of them were partly non-coding, and all of them were found to be significantly upregulated in the healthy samples.

As is evident on the generated heatmap, all the upregulated and downregulated genes were uniformly clustered across the samples, showing clear differentiation of the expression in the two conditions (Fig. 3).

Across all three methods, both GDAP2 and TBC1D24 were found to be globally differentially expressed. However, both are considered coding, hence not of interest to the analysis. No purely non-coding gene was actually found to be differentially expressed in any of the method with the

Table 1 The 22 most significantly differentially expressed genes based on the limma result (p value < 0.00005), ordered by level of significance

	logFC	AveExpr	t	P value	Adj.P Val	B	Transcript type	Gene_symbol
PH_hs_0002557	198.1722652	160.1043	23.49167	1.81E−10	1.16E−06	−3.926417315	Coding	GDAP2
PH_hs_0016338	1087.268854	825.4318	25.6805	7.17E−11	1.16E−06	−3.924010444	Coding	ORAI3
PH_hs_004871 5	576.0160043	473.676	25.00625	9.47E−11	1.16E−06	−3.924686516	Coding	C14orf101
PH_hs_0024228	230.9747311	180.3551	23.47885	1.82E−10	1.16E−06	−3.926433373	Coding	ATRN
PH_hs_0033322	249.6737199	184.7529	24.33634	1.26E−10	1.16E−06	−3.925413261	Coding	CFDP1
PH_hs_001 081 2	562.9579024	456.5773	21.73393	4.06E−10	1.34E−06	−3.928882813	Coding	ICMT
PH_hs_0003322	*762.3157775*	*483.6633*	*22.02929*	*3.53E−10*	*1.34E−06*	*−3.928428182*	*Coding, non-coding*	*MKKS*
PH_hs_0014787	638.0567641	385.9836	22.2321	3.21E−10	1.34E−06	−3.928126121	Coding	SLC3583
PH_hs_0016485	182.854354	151.8733	21.45752	4.64E−10	1.34E−06	−3.929324834	Coding	SLC7A605
PH_hs_0031 51 I	*434.7131981*	*285.526*	*21.62874*	*4.27E−10*	*1.34E−06*	*−3.929049097*	*Coding, non-coding*	*NDUFAF1*
PH_hs_0033430	411.9880082	300.9828	21.85007	3.84E−10	1.34E−06	−3.928701908	Coding	PWWP2A
PH_hs_0023957	151.6927762	115.3525	20.56248	7.21E−10	1.40E−06	−3.930876306	Coding	TBC1D24
PH_hs_0012263	98.79692443	98.79301	20.47929	7.52E−10	1.40E−06	−3.931030558	Coding	ELOVL6
PH_hs_0016643	*238.4405371*	*207.1681*	*21.008*	*5.78E−10*	*1.40E−06*	*−3.930080081*	*Coding, non-coding*	*MTHFSD*
PH_hs_0001712	169.5071385	102.6325	20.79776	6.41E−10	1.40E−06	−3.930449656	Coding	METTL13
PH_hs_000521 7	2103.482659	1142.744	20.47895	7.52E−10	1.40E−06	−3.931031196	Coding	GIMAP8
PH_hs_0014722	516.1806891	349.9946	20.53174	7.32E−10	1.40E−06	−3.930933087	Coding	AGGF1
PH_hs_0049581	2059.348668	1427.644	20.322	8.14E−10	1.44E−06	−3.931327241	Coding	MTIF3
PH_hs_0048441	118.0026118	96.16244	19.96768	9.76E−10	1.61E−06	−3.932020468	Coding	UVRAG
PH_hs_0009328	533.9497413	368.2993	19.79654	1.07E−09	1.61E−06	−3.932368222	Coding	TFCP2
PH_hs_0022762	174.7790849	204.6318	19.8491	1.04E−09	1.61E−06	−3.932260497	Coding	SMYD4
PH_hs_0048732	15,046.57014	8854.292	19.47687	1.26E−09	1.82E−06	−3.933041547	Coding	MPEG1
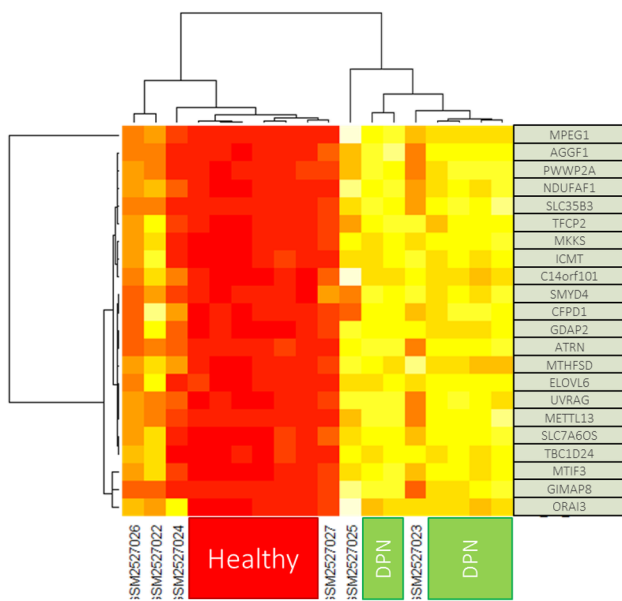
Fig. 3 Heatmap generated from the limma result for comparison DPN and healthy group. iSM25270XX indicates the label for diabetic group that exclude for analysis

cut-off at p value < 0.00005. The only partly non-coding gene found to be differentially expressed in at least two methods was MTHFSD, which is a mostly protein-coding gene that has a role in nucleic acid binding and nucleotide binding (Ota et al. 2004). No information is currently available on the non-coding fragment of the gene, though one particular lncRNA, FENDRR, has been documented to be a prominent enhancer of MTHFSD (Safran et al. 2010; Stelzer et al. 2016). FENDRR itself has been found to be linked to both the PRC2 and TrxG/MLL complexes, suggesting that it acts as modulator of chromatin signatures and promotes methylation of the target genes (Grote et al. 2013).

LMAN2L was another partly non-coding gene selected based on the significance result (p value < 0.00005) in two statistical methods. The gene is also mostly protein coding, and no annotation is available on the function of the non-coding fragments. Similar to MTHFSD, there was also an lncRNA that has been established as a significant enhancer to LMAN2L, which was LOC100506036 (Safran et al. 2010; Stelzer et al. 2016). LOC100506036 has been found to regulate immune functions in rheumatoid arthritis, possibly taking part In the hypomethylation and histone hyperacetylation mechanisms (Kolarz and Majdan 2017; Lu et al. 2016).

This study explored the 31,741 probe sets related to human mRNA and lncRNA from 18 participants. All data were normally distributed that were selected for the most significant only. Based solely on the limma result, there were three lncRNA genes found to be significantly differentially expressed. The most significantly differentially expressed of them all, LINC00324, has been found to be a target of both

DNA methylation and histone modifications (Rouillard et al. 2016). TUBA4B, which was found to be overexpressed in the DPN samples, has been well-studied to be a CpG methylation signature, specifically in diffuse gliomas (Bhat et al. 2016; Louis et al. 2014). DHRS4-AS1 mediates repressive histone modifications and methylation of the dehydrogenase/reductase genes DHRS4, DHRS4L1 and DHRS4L2 (Li et al. 2012). These genes could be considered for further study by providing many samples and comprehensive analysis of gene expression between healthy and DPN group. However, there are only two genes found that make the findings weak. Therefore, a study with more participant will reveal many genes that significant different between healthy group and DPN group.

Conclusion

As limited as they are, the utilization of lncRNA-based analysis for assessment of epigenetic regulations may serve as a promising starting point for exploratory studies. The findings from this analysis show how dynamic the epigenetic mechanisms are in DPN, ranging from hypomethylation and hypermethylation, to histone modifications. Further studies with a larger dataset and validation through MeDIP and ChIP should give researchers a more holistic view of the epigenetic mechanisms in play in the disease.

References

Al-Haddad R, Karnib N, Assaad RA, Bilen Y, Emmanuel N, Ghanem A et al (2016) Epigenetic changes in diabetes. Neurosci Lett 625:64–69. https://doi.org/10.1016/j.neulet.2016.04.046

Amaral PP, Mattick JS (2008) Noncoding RNA in development. Mamm Genome. https://doi.org/10.1007/s00335-008-9136-7

Belinky F, Bahir I, Stelzer G, Zimmerman S, Rosen N, Nativ N et al (2013) Non-redundant compendium of human ncRNA genes in GeneCards. Bioinform 29(2):255–261. https://doi.org/10.1093/bioinformatics/bts676

Bhat SA, Ahmad SM, Mumtaz PT, Malik AA, Dar MA, Urwat U et al (2016) Long non-coding RNAs: mechanism of action and functional utility. NonCoding RNA Res 1(1):43–50. https://doi.org/10.1016/j.ncrna.2016.11.002

Boulton AJM (2005) Management of diabetic peripheral neuropathy. Clin Diabetes 23(1):9–15

Davis S, Meltzer PS (2007) GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinform 23(14):1846–1847. https://doi.org/10.1093/bioinformatics/btm254

Dean L, McEntyre J (2004) The genetic landscape of diabetes. National Center for Biotechnology Information (US), Bethesda (MD). https://www.ncbi.nlm.nih.gov/books/NBK1667/

Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30(1):207–210

Feinberg AP, Fallin MD (2015) Epigenetics at the crossroads of genes and the environment. JAMA 314(11):1129–1130. https://doi.org/10.1001/jama.2015.10414

Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A et al (2013) The tissue-specific lncRNA fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell 24(2):206–214. https://doi.org/10.1016/j.devcel.2012.12.012

Hara K, Shojima N, Hosoe J, Kadowaki T (2014) Genetic architecture of type 2 diabetes. Biochem Biophys Res Commun 452:213–220. https://doi.org/10.1016/j.bbrc.2014.08.012

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS et al (2015) Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12(2):115–121. https://doi.org/10.1038/nmeth.3252

Karachanak-Yankova S, Dimova R, Nikolova D, Nesheva D, Koprinarova M, Maslyankov S et al (2015) Epigenetic alterations in patients with type 2 diabetes mellitus. Balkan J Med Genet 18(2):15–24. https://doi.org/10.1515/bjmg-2015-0081

Kolarz B, Majdan M (2017) Epigenetic aspects of rheumatoid arthritis: contribution of non-coding RNAs. Semin Arthritis Rheum. https://doi.org/10.1016/j.semarthrit.2017.01.003

Kornfeld JW, Brüning JC (2014) Regulation of metabolism by long, non-coding RNAs. Front Genet. https://doi.org/10.3389/fgene.2014.00057

Li Q, Su Z, Xu X, Liu G, Song X, Wang R et al (2012) AS1DHRS4, a head-to-head natural antisense transcript, silences the DHRS4 gene cluster in cis and trans. Proc Natl Acad Sci 109(35):14110–14115. https://doi.org/10.1073/pnas.1116597109

Louis DN, Perry A, Burger P, Ellison DW, Reifenberger G, von Deimling A (2014) International society of neuropathology-haarlem consensus guidelines for nervous system tumor classification and grading. Brain Pathol 24(5):429–435. https://doi.org/10.1111/bpa.12171

Lu MC, Yu HC, Yu CL, Huang HB, Koo M, Tung CH, Lai NS (2016) Increased expression of long noncoding RNAs LOC100652951 and LOC100506036 in T cells from patients with rheumatoid arthritis facilitates the inflammatory responses. Immunol Res 64(2):576–583. https://doi.org/10.1007/s12026-015-8756-8

Luo L, Xu J (2017) Transcriptional profiling of diabetic peripheral neuropathy patients, diabetic patients, and healthy participants. J Diabetes Res. https://doi.org/10.1155/2017/8103904

Mattick JS (2005) The functional genomics of noncoding RNA. Science 309(5740):1527–1528. https://doi.org/10.1126/science.1117806

Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 36(1):40–45. https://doi.org/10.1038/ng1285

Parikesit AA, Steiner L, Stadler PF, Prohaska SJ (2014) Pitfalls of ascertainment biases in genome annotations—computing comparable protein domain distributions in Eukarya. Malays J Fundam Appl Sci 10(2):65–75. https://doi.org/10.11113/mjfas.v10n2.57

Prohaska SJ, Stadler PF, Krakauer DC (2010) Innovation in gene regulation: the case of chromatin computation. J Theor Biol 265(1):27–44. https://doi.org/10.1016/j.jtbi.2010.03.011

Pullen TJ, Rutter GA (2014) Roles of lncRNAs in pancreatic beta cell identity and diabetes susceptibility. Front Genet. https://doi.org/10.3389/fgene.2014.00193

Reddy MA, Zhang E, Natarajan R (2015) Epigenetic mechanisms in diabetic complications and metabolic memory. Diabetologia 58(3):443–455. https://doi.org/10.1007/s00125-014-3462-y

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43(7):e47–e47. https://doi.org/10.1093/nar/gkv007

Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database. https://doi.org/10.1093/database/baw100

RStudio Team (2015) Integrated development for R. RStudio, Inc. R. RStudio, Inc., Boston

Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M et al (2010) GeneCards Version 3: the human gene integrator. Database J Biol Databases Curation. https://doi.org/10.1093/database/baq020

Steiner L, Hopp L, Wirth H, Galle J, Binder H, Prohaska SJ, Rohlf T (2012) A global genome segmentation method for exploration of epigenetic patterns. PLoS ONE 7(10):e46811. https://doi.org/10.1371/journal.pone.0046811

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S et al (2016) The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Prot Bioinf. https://doi.org/10.1002/cpbi.5

Swami M (2009) Epigenetics: misreading the code. Nat Publ Group 9:461. https://doi.org/10.1038/nrc2679

World Health Organization (2016a) Global report on diabetes. Isbn, 978, 88. ISBN: 978 92 4 156525 7

World Health Organization (2016b) WHO | diabetes. WHO, Geneva