Immune system and zinc are associated with recurrent aphthous stomatitis. An assessment using a network-based approach.

César Rivera.[1,2,3]

Affiliations: [1]Department of Biomedical Sciences, Faculty of Health Sciences, University of Talca, Talca, Chile. [2]Department of Oral Diagnosis, Piracicaba Dental School, University of Campinas, Piracicaba, Brazil. [3]Brazilian Biosciences National Laboratory (LNBio), National Center for Research in Energy and Materials (CNPEM), Campinas, São Paulo, Brazil.

Corresponding author: César Rivera. Jaime Rodríguez Carvajal Building, University of Talca, Lircay Av. s/n Talca, Chile. Phone: (56-71) 2418855. E-mail: cerivera@utalca.cl

Abstract: Objective: The aim of this research was to identify genes, proteins and processes from the biomedical information published on recurrent aphthous stomatitis (RAS) using network-based foci. Methods: The clinical context was defined using *MeSH* terms for RAS and biomarkers, combined with words associated with risk. A set of protein coding genes was prioritized using the Génie web server and classified with PANTHER. For defining biologically relevant proteins, protein-protein interaction networks were constructed using Reactome database and Cytoscape. Top 20 proteins were then subjected to functional enrichment using STRING. Results: From 1,075,576 gene-abstract links, 1,491 genes were prioritized. Proteins were related to signaling molecule proteins (n=221), receptor proteins (n=221) and nucleic acid binding proteins (n=169). The network constructed with these proteins included 3,963 nodes and functional analysis showed that main processes involved immune system and zinc ion binding function. Conclusions: For the first time, bioinformatics tools were used for integrating pathways and networks associated with RAS. Molecules and processes associated with immune system recur robustly in all analyzed information. The molecular zinc ion binding function could be an area for exploring more specific and effective therapeutic interventions.

Keywords: *Recurrent aphthous stomatitis; pathogenesis; protein interaction maps; immune system; zinc.*

INTRODUCTION.

Recurrent aphthous stomatitis (RAS) is the most common disease of the oral mucosa,[1] characterized by multiple recurrent painful ulcers in the oral cavity.[2] Usually it occurs in young healthy individuals, with a high frequency in the labial and buccal mucosa and tongue.[3] RAS occurs in three clinical presentations: minor, major, and herpetiform.[4] Several medical disorders are associated with ulcers.[5] This might suggest that RAS has pathways in common with other diseases. Despite its high prevalence, the etiology and pathogenesis of RAS still remains unclear. An immune dysregulation linked to several triggers may facilitate its development. This alteration could be related to T lymphocyte (CD4 and CD8) subpopulations that play an important role in altered immune responses.[6] To date, the treatment for RAS is non-specific, and is primarily focused on relieving pain.[7] Despite the large number of articles published and the volume of data generated, many challenges still remain. More studies will be necessary to reveal the molecular mechanisms involved in RAS, in order to allow preventive and

therapeutic interventions in the near future.

Bioinformatics tools currently allow research to focus on the integration of large-scale data.[8] Network-based approaches to human disease can have multiple biomedical and clinical applications. A better understanding of the cell interconnections present in a disease can lead to identification of the genes and pathways involved, which in turn could facilitate development of more efficient therapies.[9] Likewise, this might be useful for RAS.

Considering several molecules and processes have been suggested to be involved in etiopathogenesis of RAS, the aim of this research is to present a broad view, from a large volume of published information of genes, proteins, networks and processes that form a part of RAS using bioinformatics tools.

MATERIALS AND METHODS.

The independent variables were molecules and molecular processes; the dependent variable was etiopathogenesis of RAS. The pipeline is presented in Figure 1 and consists of 4 steps as follows:

i) Identification and prioritization of genes associated with RAS.

ii) Characterization of prioritized molecules.

iii) Construction of a contextual network and identification of main activity centers.

iv) Biological interpretation using a protein interaction network and functional enrichment.

Gene prioritization with statistical text mining.

MEDLINE/PubMed database was used to obtain the information associated with RAS. For this, the Génie web tool was used.[10] The Génie algorithm was developed for prioritizing all genes according to biomedical topics using all abstracts available in MEDLINE/PubMed. The text-mining done by Génie contributed to the initial prioritization of various genes associated with RAS through an analysis on a large scale. The topic of interest was defined from the PubMed query (performed on April 4, 2017): stomatitis, aphthous *(MeSH)* and biomarkers *(MeSH)* and risk ratio (Title/Abstract) or relative risk (Title/Abstract) or odds ratio (Title/Abstract) or odds ratio (Title/Abstract) or risk (Title/Abstract) and humans *(MeSH* Terms) and English (lang). In this initial

step, six abstracts were used as training set (PMIDs 10353862, 18512793, 9456687, 15492921, 23772946 and 24848287). In order to ensure that only significant genes were reported the cut-offs were taken as *p*-value <0.01 for abstracts, and false discovery rate <0.01 for gene selections. One-sided Fisher's exact test was computed to define the significance of gene-to-topic relationship. With these definitions, Génie processed 20,922 genes with 1,075,576 gene-abstract links (500,532 unique PMIDs).

Classification of RAS prioritized genes.

Prioritized genes by Génie were analyzed using Protein Analysis Through Evolutionary Relationships (PANTHER) classification system.[11] PANTHER is a comprehensive system which combines gene function, onthology, pathways and statistical analysis tools to allow analysis of a large volume of data.[12] For listing of genes, protein class, GO-slim biological process and GO-slim cellular component were determined. Top 5 of each category were plotted and two main subcategories were explored.

Identification of contextually relevant hubs in RAS biological networks.

Identical numerical values were assigned to all proteins identified in the panel. A network was built with Cytoscape version 3.5 and its CHAT app.[13] First neighbor interactors were sourced from Reactome peer reviewed pathway database.[14] Top 20 contextual hubs (important centers of activity) were identified.

Predicted protein-protein interactions.

Search Tool for the Retrieval of Interacting Genes meta-database (STRING)[15] was used to evaluate and integrate existing protein-protein interactions in top 20 contextual hub ranking. STRING is highly useful for functional studies since it generates networks including all types of interactions.[15] Direct (physical) and indirect (functional) associations were included. A stringent STRING score of 0.9 was chosen to depict a "high confidence" network and protein nodes without any edges were not displayed. A functional enrichment was obtained for biological processes and molecular function pathways, identifying the participating molecules and the false discovery rate. As statistical background the whole human genome was used.

RESULTS.

Over 1,000 genes are associated with RAS. Génie web server was used for identifying protein coding genes associated with RAS. In order to impart a clinical focus, the search was directed to identification of biomarkers with terms associated with risk (possibility that one condition gives this in one group *versus* another).

Génie reported 1,491 statistically relevant genes, which can be referred to in the supplementary dataset (Avalaible at: http://doi.org/10.5281/zenodo.546267). The list is headed by the following genes (UniProtKB accession numbers): BTNL2 (Q9UIR0), FCRL3 (Q96P31), BANK1 (Q8NDB2), SLC22A4 (Q9H015), LOXL1 (Q08397), VDR P11473, TAGAP (Q8N103), HLA-DQB1 (P01920), XRCC2 (O43543), MS4A2 (Q01362).

Prioritized genes for RAS represent components and processes associated with signal transduction. The set of 1,491 genes significantly associated with RAS was analyzed using PANTHER. Figure 2 shows classification and functional analysis of these genes. Top three protein categories were signaling molecule proteins (n=221), receptor proteins (n=221) and nucleic acid binding proteins (n=169) (Figure 2A). In signaling molecule proteins and receptor proteins, the main families were cytokine (52%) and G-protein coupled receptor proteins (46%).

Three main enriched biological processes were cellular processes (n=747), metabolic process (n=564) and response to stimulus (n=424) (Figure 2B). The first two were headed by the cell communication processes (67%) and primary metabolic process (44%). As for the cell localizations represented by the set, plasma membrane stood out above the rest of the categories (Figure 2C). From a broad perspective, these components and processes form a part of transmembrane signaling.

Table 1. Contextual hub analysis of recurrent aphthous stomatitis prioritized gene-set using CHAT Cytoscape app and Reactome database: Top 20 list.

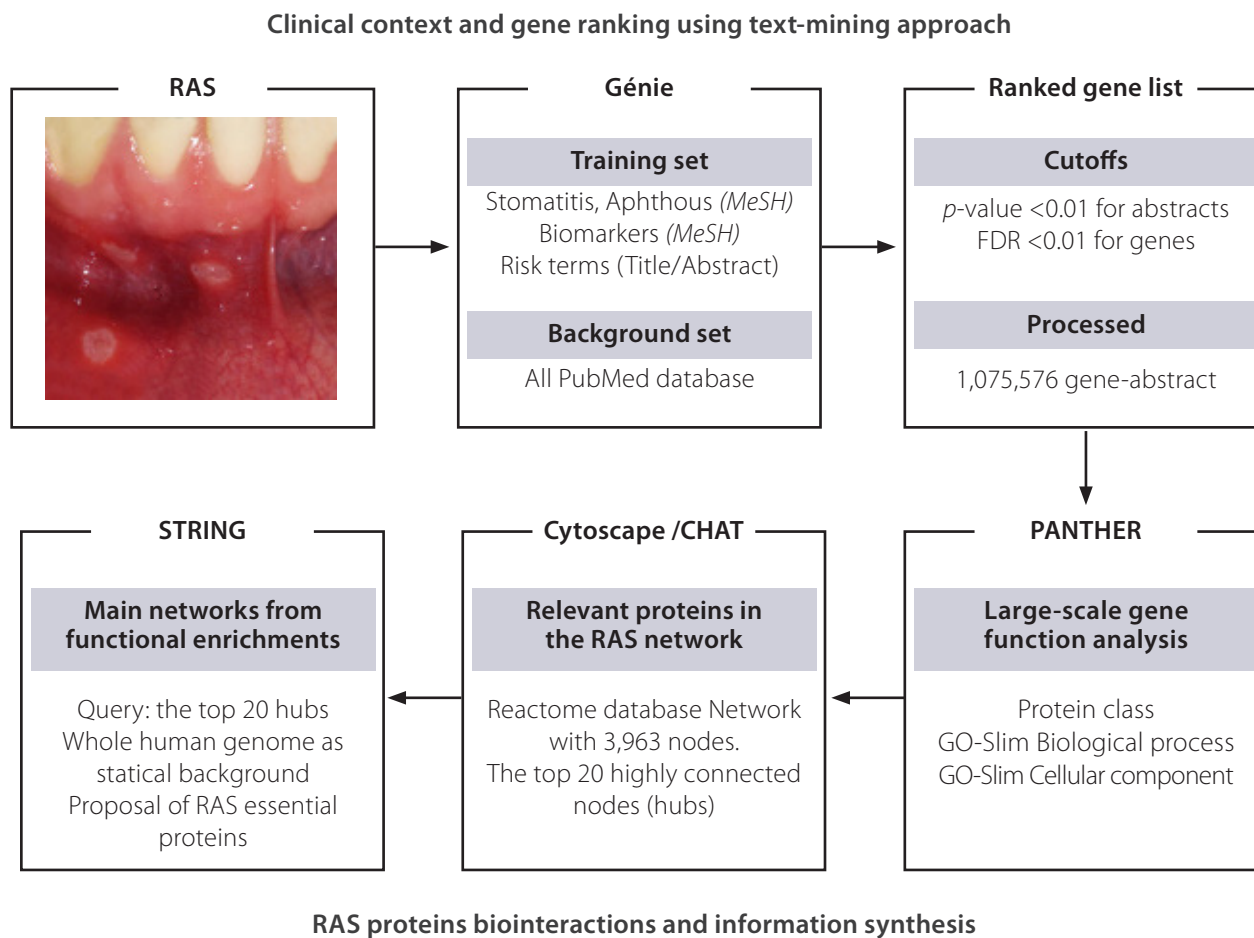| Rank	Gene symbol	UniProtKB	Protein name	*p*-value*
1	CTSS	P25774	Cathepsin S	5.98E-61
2	SAR1B	Q9Y6B6	GTP-binding protein SAR1b	1.08E-59
3	SEC24D	O94855	Protein transport protein Sec24D	1.43E-59
4	SEC31A	O94979	Protein transport protein Sec31A	1.43E-59
5	SEC24A	O95486	Protein transport protein Sec24A	1.43E-59
6	SEC24B	O95487	Protein transport protein Sec24B	1.43E-59
7	SEC24C	P53992	Protein transport protein Sec24C	1.43E-59
8	CANX	P27824	Calnexin	5.25E-58
9	SEC23A	Q15436	Protein transport protein Sec23A	2.74E-57
10	STAT1	P42224-1	Signal transducer and activator of transcription 1-alpha/beta	5.69E-56
11	BCAP31	P51572	B-cell receptor-associated protein 31	3.73E-52
12	-	P04436	T-cell receptor alpha chain V region HPB-MLT	1.11E-51
13	TCRA	P04437	T-cell receptor alpha chain V region PY14	1.11E-51
14	-	P01737	T-cell receptor alpha chain V region PY14	1.11E-51
15	TRAC	P01848	T-cell receptor alpha chain C region	1.11E-51
16	TCRB	P04435	T-cell receptor alpha chain C region	3.53E-51
17	TRBV12-3	P01733	T-cell receptor beta chain V region YT35	3.53E-51
18	TRBC1	P01850	T-cell receptor beta-1 chain C region	3.53E-51
19	ERAP1	Q9NZ08	Endoplasmic reticulum aminopeptidase 1	4.20E-51
20	ERAP2	Q6P179	Endoplasmic reticulum aminopeptidase 2	4.20E-51

*Adjusted. The complete list is available in the supplementary dataset (DOI: 10.5281/zenodo.546267).

Table 2. Main networks* from functional enrichments using STRING database.

| Pathway ID		Pathway description	Matching genes (protein coding)	FDR
Biological process	GO.0002376	Immune system process: any process involved in the development or functioning of the immune system, an organismal system for calibrated responses to potential internal or invasive threats	BCAP31, CANX, CTSS, ERAP1, ERAP2, SAR1B, SEC23A, SEC24A, SEC24B, SEC24C, SEC24D, SEC31A, STAT1	9.16E-10
Molecular function	GO.0008270	Zinc ion binding: Interacting selectively and non-covalently with zinc (Zn) ions	ERAP1, ERAP2, SEC23A, SEC24A, SEC24B, SEC24C, SEC24D	0.0271

*The first place of each functional analysis has been shown. The complete list is available in the supplementary dataset (DOI: 10.5281/zenodo.546267). FDR, false discovery rate.

Figure 1. Flow chart for analysis of evidence associated with recurrent aphthous stomatitis.

Clinical context and gene ranking using text-mining approach

**RAS proteins biointeractions and information synthesis**

The initial step consisted of prioritization of genes in the tool Génie, using as training set a MEDLINE/PubMed query with clinical emphasis. For classification, genes were introduced into the PANTHER classification system. Subsequently a contextual analysis (giving equal importance to each protein) was conducted using Cytoscape, their CHAT application and Reactome database. Major centers of activity were then analyzed in STRING database, in order to obtain a limited list of identifications and the main functional networks.

Figure 2. Large-scale gene ontology analysis using PANTHER classification system.
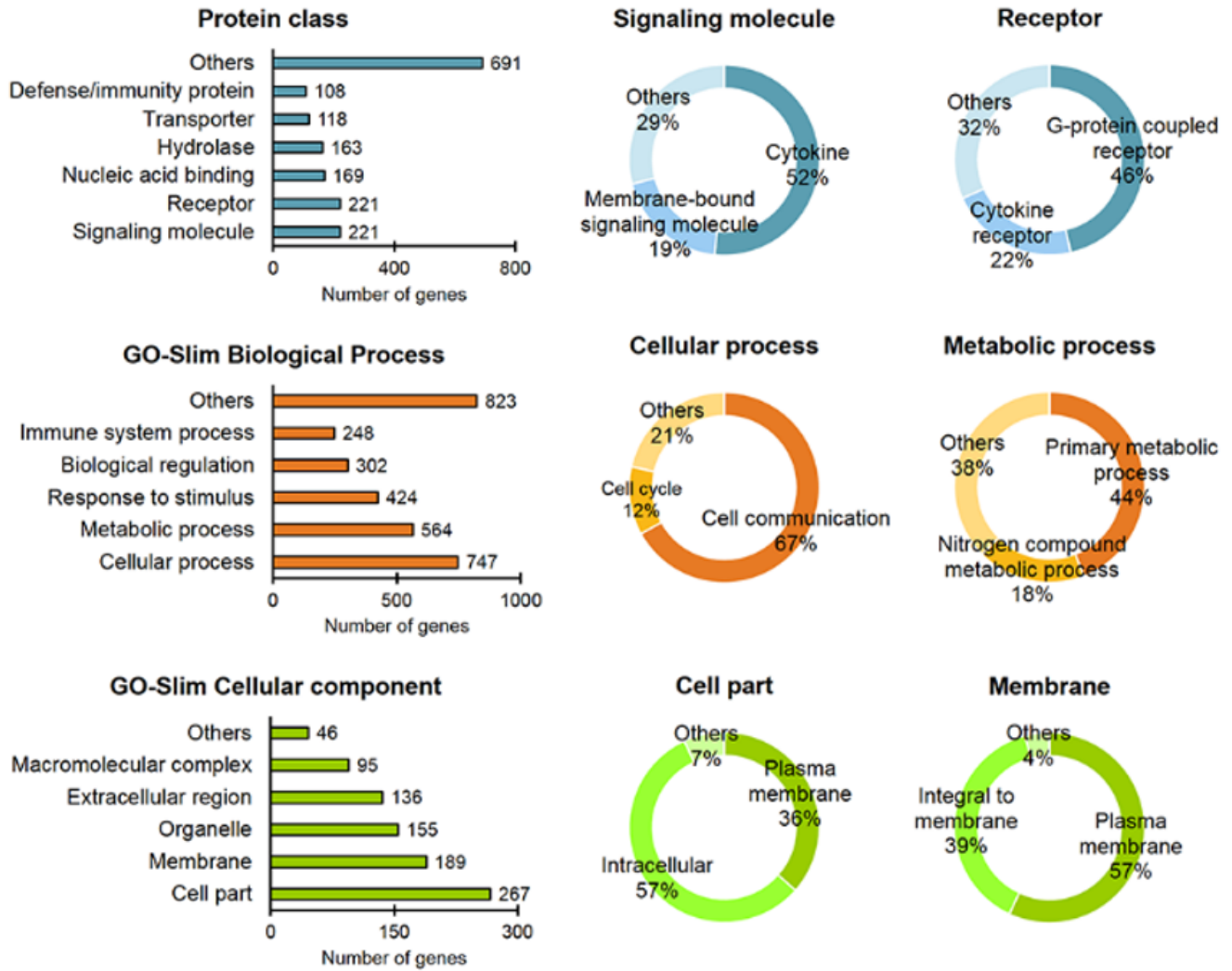
All genes significantly associated with recurrent aphthous stomatitis were then analyzed on PANTHER.
The bar graphs show the analysis according to the protein class, biological process and cell component.
Pie charts show details for the more relevant subcategories of each group. For details see the supplementary dataset (DOI: 10.5281/zenodo.546267).

Figure 3. Protein-protein interaction networks of recurrent aphthous stomatitis.
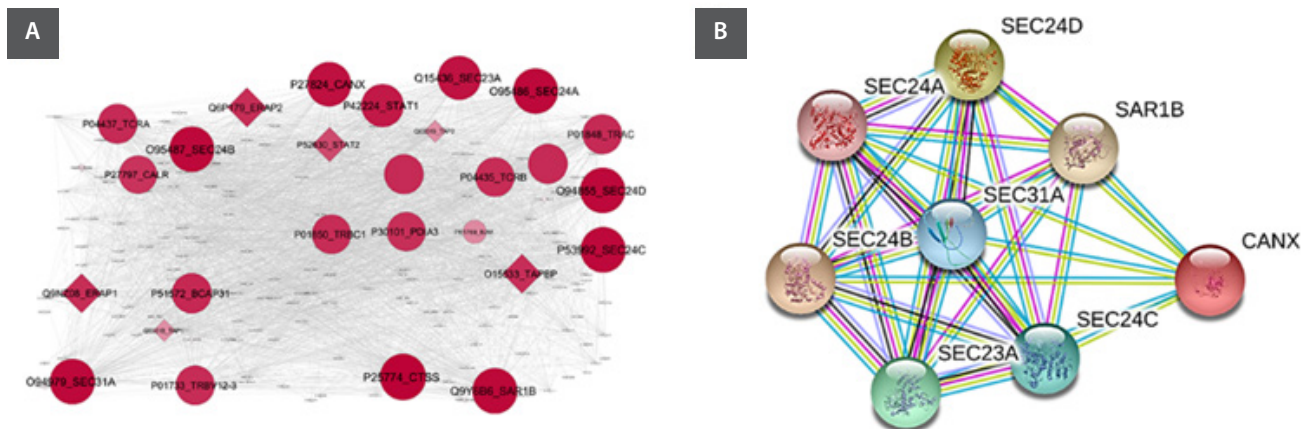
Complex interaction network constructed using CHAT Cytoscape app and Reactome.
(A). The main activity centers are represented by a color gradient. Top 20 highly connected nodes (hubs) were selected from them for subsequent analysis with STRING. STRING predicted protein-protein interactions of top 20 hubs.
(B). Color lines connecting nodes represent the level of evidence (for the meaning of colors, visit https://string-db.org/). The result of functional enrichment can be read in Table 2.

Biologically relevant proteins are part of immune system and zinc ion binding processes. For a wider interpretation of the data in a biological context, the list of proteins was analyzed and enriched using the Cytoscape program and its application CHAT (using the Reactome data as the source). The supplementary Figure 1 (Available at: https://zenodo.org/record/546267) shows a general view of the web explorer of Reactome base, which was used before inputting the proteins in Cytoscape. The principal overrepresented pathways were transduction of signals and the immune system processes. Cytoscape built a network of 3,963 nodes (the complete list can be viewed in the supplementary dataset Available at: https://zenodo.org/record/546267).

The major centers of activity (hubs) in the interaction network are represented by nodes in red (Figure 3A). Top 20 hubs are presented in Table 1. In order to better understand the role of these hubs in RAS, their functional interactions were categorized using STRING. The constructed network represented a clear simplification of that viewed in Cytoscape (Figure 3B). Eight of the 20 proteins were shown to be connected. Table 2 shows that functional enrichment from 20 biologically relevant proteins corresponded to immune system process (13 out of 20 proteins), zinc ion binding (7/20) and protein in endoplasmic reticulum (8/20).

DISCUSSION.

This study summarized molecules and biological functions associated with RAS. Overall, the results suggest that immune system and zinc ion binding function have an important role in the etiopathogenesis of RAS.

A disease is a consequence of different disturbances in intercellular complex networks.[9] Linking genes and proteins with diseases in which they are involved is at the heart of molecular medicine.[16] With this motivation, this analysis has attempted to link a large quantity of genes, proteins and biological processes to RAS, an oral disease where the etiopathogenic process presents several questions.

RAS is an epidemiologically relevant disease affecting some 10-20% of the general population with a predilection for young patients (between 20 and 30 years).[3,17] Therapeutic alternatives on which the dentists rely in order to treat RAS are focused on reducing pain symptoms.[7] Biomedical

literature on RAS shows that this disease is associated with complex biological processes. The focus of this work allows initiating the exploration of RAS without knowing exactly the genes involved. The selection of published information is relatively simple. In the entire process the manual preparation of data is minimal.

One of the initial limitations of this study lies in the definition of RAS. Besides its clinical variants, RAS can be divided between simple and complex aphthosis. Additionally, complex aphthosis can be divided into primary and secondary groups. In the primary group it is not possible to identify the causes of the case and this remains as idiopathic. Nevertheless, in the second group different causes can be listed, like hematinic deficiencies, inflammatory processes and immunodeficiencies.[18] This precise definition is found often absent in the sources used and an initial filter in this sense is difficult to apply as of now. However, with cells being considered as interaction networks,[19] it appears difficult to assume that local or systemic conditions that accompany oral ulcerations do not share different pathways and molecules.

PANTHER classification for RAS proteins showed that most representative categories pointed to signal transduction, cellular communication, cytokines, receptors and cell membrane as chief components. This general scenario approximates the etiopathogenic process habitually described for RAS: a process initiated by stimulation of keratinocytes of oral mucosa by an unrecognized antigen, leading to a stimulation of subpopulations of T lymphocytes, liberation of cytokines, different interleukins, which precede neutrophils arrival.[20,21]

The previous description is also found in tune with Reactome results (main overrepresented pathways were transduction of signals and immune system) and subsequent analysis.

STRING enrichment for Cytoscape/CHAT major centers of activity resulted in immune system network process. There is a large amount of literature demonstrating this association. However, an unexpected result was the appearance of molecular zinc ion binding function. In patients with zinc deficiency, an overproduction of pro-inflammatory cytokines such as TNF-α takes place.[22] This cytokine has shown to be essential in the RAS pathogenic process.[23] Exploring this function, there is

previous evidence presenting an association between levels of zinc and RAS episodes. It has been reported that zinc deficiency is found in 30-40% patients with RAS[24,25] and the treatment with zinc has beneficial effects, reducing the frequency of episodes.[25,26]

However, therapeutic evidence does not clearly describe sample size determination or randomization. In addition, a Cochrane systematic review (25 trials were included) concluded that no single treatment was found to be effective and remain inconclusive in regard to the best systemic intervention for RAS.[21] This emphasizes the need for more well-designed clinical cohort-based studies.

CONCLUSION.

Molecules and processes associated with immune system are recurrent in all published information on RAS. Molecular zinc ion binding function could be an area for exploring more specific and effective therapeutic interventions.

The integrated knowledge which allows to "include several trees to see the forest", has the potential to present basic domains for building an RAS pathogenic therapeutic map. As the database increases with new biomedical information, new systematic analyses are necessary. Future studies could verify and monitor experimentally and clinically the role of the molecules listed in this research.

REFERENCES.

1. Edgar NR, Saleh D, Miller RA. Recurrent Aphthous Stomatitis: A Review. J Clin Aesthet Dermatol. 2017;10(3):26–36.
2. Akintoye SO, Greenberg MS. Recurrent aphthous stomatitis. Dent Clin North Am. 2014;58(2):281–97.
3. Rivera C, Jones-Herrera C, Vargas P, Venegas B, Droguett D. Oral diseases: a 14-year experience of a Chilean institution with a systematic review from eight countries. Med Oral Patol Oral Cir Bucal. 2017;22(3):e297–e306.
4. Preeti L, Magesh K, Rajkumar K, Karthik R. Recurrent aphthous stomatitis. J Oral Maxillofac Pathol. 2011;15(3):252–6.
5. Jin LJ, Lamster IB, Greenspan JS, Pitts NB, Scully C, Warnakulasuriya S. Global burden of oral diseases: emerging concepts, management and interplay with systemic health. Oral Dis. 2016;22(7):609–19.
6. Riera Matute G, Riera Alonso E. [Recurrent aphthous stomatitis in Rheumatology]. Reumatol Clin. 2011;7(5):323–8.
7. Dan S, Jinwei Z, Qiang Z, Jianwei S, Weijun Z. Exploring the Molecular Mechanism and Biomarker of Recurrent Aphthous Stomatitis Based on Gene Expression Microarray. Clin Lab. 2017;63(2):249–53.
8. Donnard E, Barbosa-Silva A, Guedes RL, Fernandes GR, Velloso H, Kohn MJ, Andrade-Navarro MA, Ortega JM. Preimplantation development regulatory pathway construction through a text-mining approach. BMC Genomics. 2011;12(Suppl 4):S3.
9. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56–68.
10. Fontaine JF, Priller F, Barbosa-Silva A, Andrade-Navarro MA. Génie: literature-based gene prioritization at multi genomic scale. Nucleic Acids Res. 2011;39(suppl 2):W455–61.
11. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 2017;45(D1):D183–9.
12. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013;8(8):1551–66.
13. Muetze T, Goenawan IH, Wiencko HL, Bernal-Llinares M, Bryan K, Lynn DJ. Contextual Hub Analysis Tool (CHAT): A Cytoscape app for identifying contextually relevant hubs in biological networks. F1000Res. 2016;5
14. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H, D'Eustachio P. The Reactome pathway Knowledgebase. Nucleic Acids Res. 2016;44(D1):D481–7.
15. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(D1):D447–52.
16. Pletscher-Frankild S, Pallejà A, Tsafou K, Binder JX, Jensen LJ. DISEASES: text mining and data integration of disease-gene associations. Methods. 2015;74(1):83–9.
17. Chattopadhyay A, Chatterjee S. Risk indicators for recurrent aphthous ulcers among adults in the US. Community Dent Oral Epidemiol. 2007;35(2):152–9.
18. Baccaglini L, Lalla RV, Bruce AJ, Sartori-Valinotti JC, Latortue MC, Carrozzo M, Rogers RS 3rd. Urban legends: recurrent aphthous stomatitis. Oral Dis. 2011;17(8):755–70.
19. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. Cell. 2011;144(6):986–98.
20. Avci E, Akarslan ZZ, Erten H, Coskun-Cevher S. Oxidative stress and cellular immunity in patients with recurrent aphthous ulcers. Braz J Med Biol Res. 2014;47(5):355–60.
21. Brocklehurst P, Tickle M, Glenny AM, Lewis MA, Pemberton MN, Taylor J, Walsh T, Riley P, Yates JM. Systemic interventions for recurrent aphthous stomatitis (mouth ulcers). Cochrane Database Syst Rev. 2012;(9):CD005411.
22. Prasad AS. Zinc in human health: effect of zinc on immune cells. Mol Med. 2008;14(5-6):353–7.
23. Slebioda Z, Szponar E, Kowalska A. Etiopathogenesis of recurrent aphthous stomatitis and the role of immunologic aspects: literature review. Arch Immunol Ther Exp. 2014;62(3):205–15.
24. Ozler GS. Zinc deficiency in patients with recurrent aphthous stomatitis: a pilot study. J Laryngol Otol. 2014;128(6):531–3.
25. Orbak R, Cicek Y, Tezel A, Dogru Y. Effects of zinc treatment in patients with recurrent aphthous stomatitis. Dent Mater J. 2003;22(1):21–9.
26. Endre L. Recurrent aphthous ulceration with zinc deficiency and cellular immune deficiency. Oral Surg Oral Med Oral Pathol. 1991;72(5):559–61.