Posted on Authorea 9 Mar 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.164684924.40795659/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

The Globally search for a Regular Expression and Print matching lines (GREP) strategy: an innovative reanalysis strategy combining bibliographic monitoring with fast GREP directly applied to a massive genomic database to rapidly improve diagnosis

Frédéric Tran Mau Them[1], Alexis Overs[1], ange-line bruel[1], Romain Duquet[1], Mylene Thareau[1], Anne-Sophie Denommé-Pichon[1], Antonio Vitobello[1], Arthur Sorlin[1], Hana Safraou[1], Sophie Nambot[1], Julian Delanne[1], Sebastien Moutton[1], Caroline RACINE[1], Camille Engel[1], Melchior D'agay[1], Daphné Lehalle[1], Alice Goldenberg[2], Marjolaine Willems[3], christine Coubes[3], David Geneviève[3], Alain Verloes[4], Yline CAPRI[5], Laurence Perrin[5], Marie-Line Jacquemont[6], Laetitia Lambert[7], Elodie Lacaze[8], Julien Thevenon[1], Nadine Hanna[9], Van-Gils Julien[10], Charlotte Dubucs[11], Varoona Bizaoui[12], Marion Gerard[12], James Lespinasse[13], Sandra Mercier[14], Anne-Marie Guerrot[2], Isabelle Maystadt[15], Emilie Tisserant[1], Laurence Faivre[1], Christophe Philippe[1], Yannis Duffourd[1], and Christel Thauvin-Robinet[1]

[1]Centre Hospitalier Universitaire Dijon Bourgogne
[2]Centre Hospitalier Universitaire de Rouen
[3]Centre Hospitalier Regional Universitaire de Montpellier
[4]Hopital Universitaire Robert-Debre Departement de genetique
[5]Hopital Universitaire Robert-Debre Unite fonctionnelle de genetique clinique
[6]Centre Hospitalier Universitaire de La Reunion Sites Sud Saint-Pierre
[7]Centre hospitalier regional universitaire de Nancy
[8]Groupe Hospitalier du Havre
[9]Hopital Bichat - Claude-Bernard
[10]Centre Hospitalier Universitaire de Bordeaux
[11]Centre Hospitalier Universitaire de Toulouse
[12]Centre Hospitalier Universitaire de Caen
[13]Centre Hospitalier Metropole Savoie
[14]Centre Hospitalier Universitaire de Nantes
[15]Institut de Pathologie et de Genetique asbl

March 9, 2022

Abstract

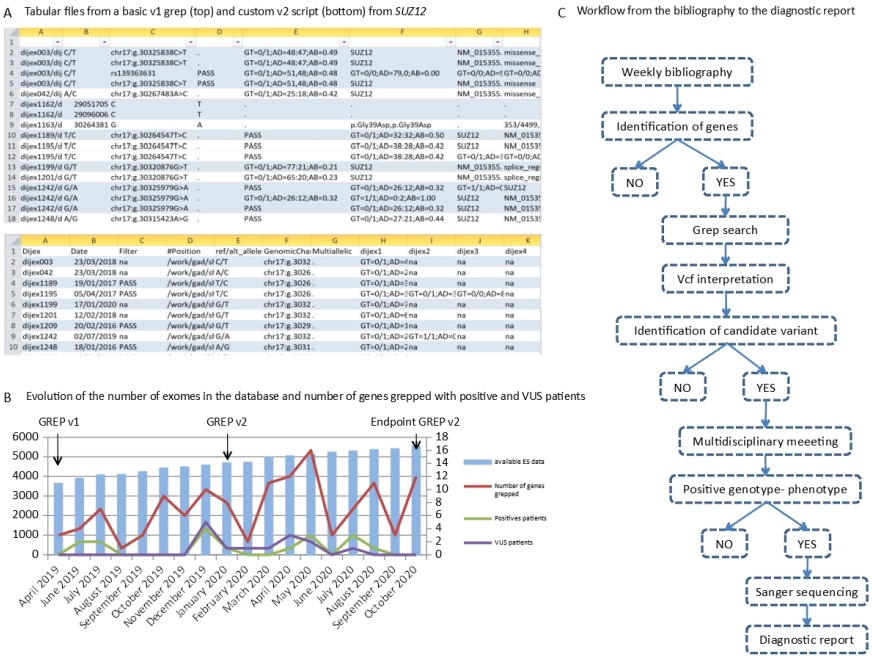
Purpose: Exome sequencing has a diagnostic yield ranging from 25% to 70% in rare diseases and regularly implicates genes in novel disorders. Prospective data reanalysis has demonstrated strong efficacy in improving diagnosis, but poses organizational difficulties for clinical laboratories. We applied a reanalysis strategy based on intensive prospective bibliographic monitoring, and directly applied the Globally search for a Regular Expression and Print matching lines (GREP) command-line to a massive ES database. **Methods:** For 18 months, we submitted daily the same 5 keywords of interest ((*intellectual disability*, (

Posted on Authorea 9 Mar 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.164684924.40795059/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

neuro)developmental delay, (neuro)developmental disorder)) to PubMed, to identify recently published, novel disease-gene associations, or new phenotypes in genes already implicated in human pathology. We used the Linux GREP command-line and an in-house script, to collect all variants in these genes from our 5459 exome database. **Results**: We *grepped* 128 genes and collected 56 candidate variants in 53 individuals. We confirmed causal diagnosis for 19/128 genes (15%) in 21 individuals, and identified variants of unknown significance for 19/128 genes (15%) in 23 individuals. Altogether, we confirmed pathogenicity in 21/2875 undiagnosed affected probands (0.7%). **Conclusion**: The GREP command-line is efficient, and less tedious than complete periodical reanalysis. It is an interesting reanalysis strategy to improve diagnosis.

Hosted file

`GREP_final.docx` available at https://authorea.com/users/464395/articles/559305-the-globally-search-for-a-regular-expression-and-print-matching-lines-grep-strategy-an-innovative-reanalysis-strategy-combining-bibliographic-monitoring-with-fast-grep-directly-applied-to-a-massive-genomic-database-to-rapidly-improve-diagnosis



Posted on Authorea 9 Mar 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.164684924.40795659/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Workflow and version of the GREP reanalysis strategy with number and class of identified variants, genes and individuals in the expanding exome database