Solve-RD: systematic pan-European data sharing and collaborative analysis to solve rare diseases

ZUREK, Birte, ELLWANGER, Kornelia, VISSERS, Lisenka E. L. M., SCHÜLE, Rebecca, SYNOFZIK, Matthis, TÖPF, Ana, DE VOER, Richarda M., LAURIE, Steven, MATALONGA, Leslie, GILISSEN, Christian, OSSOWSKI, Stephan, 'T HOEN, Peter A. C., VITOBELLO, Antonio, SCHULZE-HENTRICH, Julia M., RIESS, Olaf, BRUNNER, Han G., BROOKES, Anthony J., RATH, Ana, BONNE, Gisèle, GUMUS, Gulcin, VERLOES, Alain, HOOGERBRUGGE, Nicoline, EVANGELISTA, Teresinha, HARMUTH, Tina, SWERTZ, Morris, SPALDING, Dylan, HOISCHEN, Alexander, BELTRAN, Sergi, GRAESSNER, Holm, HAACK, Tobias B., ZUREK, Birte, ELLWANGER, Kornelia, DEMIDOV, German, STURM, Marc, KESSLER, Christoph, WAYAND, Melanie, WILKE, Carlo, TRASCHÜTZ, Andreas, SCHÖLS, Ludger, HENGEL, Holger, HEUTINK, Peter, BRUNNER, Han, SCHEFFER, Hans, STEYAERT, Wouter, SABLAUSKAS, Karolis, DE VOER, Richarda M., KAMSTEEG, Erik-Jan, VAN DE WARRENBURG, Bart, VAN OS, Nienke, TE PASKE, Iris, JANSSEN, Erik, DE BOER, Elke, STEEHOUWER, Marloes, YALDIZ, Burcu, KLEEFSTRA, Tjitske, VEAL, Colin, GIBSON, Spencer, WADSLEY, Marc, MEHTARIZADEH, Mehdi, RIAZ, Umar, WARREN, Greg, DIZJIKAN, Farid Yavari, SHORTER, Thomas, STRAUB, Volker, BETTOLO, Chiara Marini, SPECHT, Sabine, CLAYTON-SMITH, Jill, BANKA, Siddharth, ALEXANDER, Elizabeth, JACKSON, Adam, FAIVRE, Laurence, THAUVIN, Christel, VITOBELLO, Antonio, DENOMMÉ-PICHON, Anne-Sophie, DUFFOURD, Yannis, TISSERANT, Emilie, BRUEL, Ange-Line, PEYRON, Christine, PÉLISSIER, Aurore, BELTRAN, Sergi, GUT, Ivo Glynne, LAURIE, Steven, PISCIA, Davide, MATALONGA, Leslie, PAPAKONSTANTINOU, Anastasios, BULLICH, Gemma, CORVO, Alberto, GARCIA, Carles, FERNANDEZ-CALLEJO, Marcos, HERNÁNDEZ, Carles, PICÓ, Daniel, PARAMONOV, Ida, LOCHMÜLLER, Hanns, GUMUS, Gulcin, BROS-FACER, Virginie, HANAUER, Marc, OLRY, Annie, LAGORCE, David, HAVRYLENKO, Svitlana, IZEM, Katia, RIGOUR, Fanny, STEVANIN, Giovanni, DURR, Alexandra, DAVOINE, Claire-Sophie, GUILLOT-NOEL, Léna, HEINZMANN, Anna, COARELLI, Giulia, ALLAMAND, Valérie, NELSON, Isabelle, YAOU, Rabah Ben, METAY, Corinne, EYMARD, Bruno, COHEN, Enzo, ATALAIA, Antonio, STOJKOVIC, Tanya, MACEK, Milan, TURNOVEC, Marek, THOMASOVÁ, Dana, KREMLIKOVÁ, Radka Pourová, FRANKOVÁ, Vera, HAVLOVICOVÁ, Markéta, KREMLIK, Vlastimil, PARKINSON, Helen, KEANE, Thomas, SENF, Alexander, ROBINSON, Peter, DANIS, Daniel, ROBERT, Glenn, COSTA, Alessia, PATCH, Christine, HANNA, Mike, HOULDEN, Henry, REILLY, Mary,

Alexander, BELTRAN, Sergi, GRAESSNER, Holm, HAACK, Tobias B., ZUREK, Birte, ELLWANGER, Kornelia, DEMIDOV, German, STURM, Marc, KESSLER, Christoph, WAYAND, Melanie, WILKE, Carlo, TRASCHÜTZ, Andreas, SCHÖLS, Ludger, HENGEL, Holger, HEUTINK, Peter, BRUNNER, Han, SCHEFFER, Hans, STEYAERT, Wouter, SABLAUSKAS, Karolis, DE VOER, Richarda M., KAMSTEEG, Erik-Jan, VAN DE WARRENBURG, Bart, VAN OS, Nienke, TE PASKE, Iris, JANSSEN, Erik, DE BOER, Elke, STEEHOUWER, Marloes, YALDIZ, Burcu, KLEEFSTRA, Tjitske, VEAL, Colin, GIBSON, Spencer, WADSLEY, Marc, MEHTARIZADEH, Mehdi, RIAZ, Umar, WARREN, Greg, DIZJIKAN, Farid Yavari, SHORTER, Thomas, STRAUB, Volker, BETTOLO, Chiara Marini, SPECHT, Sabine, CLAYTON-SMITH, Jill, BANKA, Siddharth, ALEXANDER, Elizabeth, JACKSON, Adam, FAIVRE, Laurence, THAUVIN, Christel, VITOBELLO, Antonio, DENOMMÉ-PICHON, Anne-Sophie, DUFFOURD, Yannis, TISSERANT, Emilie, BRUEL, Ange-Line, PEYRON, Christine, PÉLISSIER, Aurore, BELTRAN, Sergi, GUT, Ivo Glynne, LAURIE, Steven, PISCIA, Davide, MATALONGA, Leslie, PAPAKONSTANTINOU, Anastasios, BULLICH, Gemma, CORVO, Alberto, GARCIA, Carles, FERNANDEZ-CALLEJO, Marcos, HERNÁNDEZ, Carles, PICÓ, Daniel, PARAMONOV, Ida, LOCHMÜLLER, Hanns, GUMUS, Gulcin, BROS-FACER, Virginie, HANAUER, Marc, OLRY, Annie, LAGORCE, David, HAVRYLENKO, Svitlana, IZEM, Katia, RIGOUR, Fanny, STEVANIN, Giovanni, DURR, Alexandra, DAVOINE, Claire-Sophie, GUILLOT-NOEL, Léna, HEINZMANN, Anna, COARELLI, Giulia, ALLAMAND, Valérie, NELSON, Isabelle, YAOU, Rabah Ben, METAY, Corinne, EYMARD, Bruno, COHEN, Enzo, ATALAIA, Antonio, STOJKOVIC, Tanya, MACEK, Milan, TURNOVEC, Marek, THOMASOVÁ, Dana, KREMLIKOVÁ, Radka Pourová, FRANKOVÁ, Vera, HAVLOVICOVÁ, Markéta, KREMLIK, Vlastimil, PARKINSON, Helen, KEANE, Thomas, SENF, Alexander, ROBINSON, Peter, DANIS, Daniel, ROBERT, Glenn, COSTA, Alessia, PATCH, Christine, HANNA, Mike, HOULDEN, Henry, REILLY, Mary, VANDROVCOVA, Jana, MUNTONI, Francesco, ZAHARIEVA, Irina, SARKOZY, Anna, TIMMERMAN, Vincent, BAETS, Jonathan, VAN DE VONDEL, Liedewei, BEIJER, Danique, DE JONGHE, Peter, NIGRO, Vincenzo, BANFI, Sandro, TORELLA, Annalaura, MUSACCHIA, Francesco, PILUSO, Giulio, FERLINI, Alessandra, SELVATICI, Rita, ROSSI, Rachele, NERI, Marcella, ARETZ, Stefan, SPIER, Isabel, SOMMER, Anna Katharina, PETERS, Sophia, OLIVEIRA, Carla, PELAEZ, Jose Garcia, MATOS, Ana Rita, JOSÉ, Celina São, FERREIRA, Marta, GULLO, Irene, FERNANDES, Susana, GARRIDO, Luzia, FERREIRA, Pedro, CARNEIRO, Fátima, SWERTZ, Morris A., JOHANSSON, Lennart, VAN DER VELDE, Joeri K., VAN DER VRIES, Gerben, NEERINCX, Pieter B., ROELOFS-PRINS, Dieuwke, KÖHLER, Sebastian, METCALFE, Alison, VERLOES, Alain, DRUNAT, Séverine, ROORYCK, Caroline, TRIMOUILLE, Aurelien, CASTELLO, Raffaele, MORLEO, Manuela, PINELLI, Michele, VARAVALLO, Alessandra, DE LA PAZ, Manuel Posada, SÁNCHEZ, Eva Bermejo, MARTÍN, Estrella López, DELGADO, Beatriz Martínez, DE LA ROSA, F. Javier Alonso García, CIOLFI, Andrea, DALLAPICCOLA, Bruno, PIZZI, Simone, RADIO, Francesca Clementina, TARTAGLIA, Marco, RENIERI, Alessandra, BENETTI, Elisa, BALICZA, Peter, MOLNAR, Maria Judit, MAVER, Ales, PETERLIN, Borut, MÜNCHAU, Alexander, LOHMANN, Katja, HERZOG, Rebecca, PAULY, Martje, MACAYA, Alfons, MARCÉ-GRAU, Anna, OSORIO, Andres Nascimiento, DE BENITO, Daniel Natera, LOCHMÜLLER, Hanns, THOMPSON, Rachel, POLAVARAPU, Kiran, BEESON, David, COSSINS, Judith, CRUZ, Pedro M. Rodriguez, HACKMAN, Peter, JOHARI, Mridul, SAVARESE, Marco, UDD, Bjarne, HORVATH, Rita, CAPELLA, Gabriel, VALLE, Laura, HOLINSKI-FEDER, Elke, LANER, Andreas, STEINKE-LANGE, Verena, SCHRÖCK, Evelin and RUMP, Andreas (2021). Solve-RD: systematic pan-

Supplementary Information

Zurek B, Ellwanger K, Vissers L, Schüle R, Synofzik M, Töpf A, et al. Solve-RD: systematic Pan- European data sharing and collaborative analysis to solve Rare Diseases.

Details on Solve-RD Working groups

The "SNV/indel" WG (WG1) is using the GPAP APIs (Application Programming Interfaces) with different combinations of filters to programmatically re-analyse the exome/genome-phenome data. The initial approach to discover the most obvious cases ("low-hanging fruit") has contributed to solve 120 cases (1). This WG has also combined and annotated all the gVCFs with a common pipeline to make them available through the Sandbox for further analyses. gVCF files, together with PhenoPackets, are used as input for Exomiser (2, 3), LIRICAL (4) and CAPICE (5).

The "Copy Number Variation (CNV)" WG (WG2) has already processed all existing WES data. Subclusters for experiments batched by the enrichment kit used for library preparation were generated by ClusterWES (6). Each of these 85 clusters was then used for CNV detection by four tools: VarGenius (7), ClinCNV (https://github.com/imgag/ClinCNV), Conifer (8) and ExomeDepth (9).

The "Runs of Homozygosity (ROH) and relatedness" WG3 has performed quality control tests on the dataset using several methods in parallel, including kinship analyses and the estimation of consanguinity based on the computation of ROH as recently described (10).

The "*de novo* mutations" WG4 have analysed trios to systematically identify *de novo* mutations using existing and novel tools. Finally, the "Meta-analysis" WG5 aims to

identify genes with an excess of variants i.e. ‘variant burden’ in certain patient cohorts, with the first candidate genes already being evaluated.

Supplementary References

1. Matalonga L, Hernández-Ferrer C, Piscia D, Solve-RD SNV-indel working group, Vissers LELM, Schüle R, et al. Diagnosis of rare disease patients through programmatic reanalysis of genome-phenome data. Manuscript submitted to EJHG.
2. Smedley D, Jacobsen JO, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10(12):2004-15.
3. Robinson PN, Köhler S, Oellrich A, Wang K, Mungall CJ, Lewis SE, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res. 2014;24(2):340-8.
4. Robinson PN, Ravanmehr V, Jacobsen JOB, Danis D, Zhang XA, Carmody LC, et al. Interpretable Clinical Genomics with a Likelihood Ratio Paradigm. Am J Hum Genet. 2020;107(3):403-17.
5. Li S, van der Velde KJ, de Ridder D, van Dijk ADJ, Soudis D, Zwerwer LR, et al. CAPICE: a computational method for Consequence-Agnostic Pathogenicity Interpretation of Clinical Exome variations. Genome Med. 2020;12(1):75.
6. Johansson L. Manuscript in preparation2020.
7. Musacchia F, Ciolfi A, Mutarelli M, Bruselles A, Castello R, Pinelli M, et al. VarGenius executes cohort-level DNA-seq variant calling and annotation and allows to manage the resulting data through a PostgreSQL database. BMC Bioinformatics. 2018;19(1):477.
8. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012;22(8):1525-32.
9. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics. 2012;28(21):2747-54.
10. Matalonga L, Laurie S, Papakonstantinou A, Piscia D, Mereu E, Bullich G, et al. Improved Diagnosis of Rare Disease Patients through Systematic Detection of Runs of Homozygosity. J Mol Diagn. 2020.