Combining big data science with clinics: Novel approach for understanding human sex development and its variants

Daniel Rodríguez Gutiérrez[1], Isabel von der Decken[1], Patrick Sproll[1], Brian Stevenson[3], Mariarosaria Lang-Muritano[4], Daniel Konrad[4], Dagmar L'Allemand[5], Serge Nef[2], Anna Biason-Lauber[1]

[1]Endocrinology division, Section of Medicine, University of Fribourg, Fribourg, Switzerland
[2]Department of genetic and molecular medicine, University of Geneva, Geneva, Switzerland
[3]Vital-IT Group, University of Lausanne, Lausanne, Switzerland
[4]Endocrinology / Diabetology, Kinderspital Zürich, UZH, Zurich Switzerland
[5]Pediatric Endocrinology / Diabetology Kantonspital St. Gallen, Switzerland

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

ESPE 2021
The 59th Annual Meeting Online
22-26 September 2021

Introduction

Clinical genetics become essential to provide a definitive diagnosis for rare syndromes. It also can have an influence on disease prognosis, treatment and prevention in reproduction and fertility . Next-generation sequencing (NGS) makes now possible to analyze thousands of genes at an unprecedented speed. However, one major challenge is to process the huge amount the data and correlate it with relevant medical information. This is particularly relevant in rare diseases where, after filtering and discarding known disease causative genes, researchers still have to deal with approx. 1,000 potentially relevant variants for each patient. In more than 50% of DSD patients no known gene variants are identified, suggesting the existence of a number of unknown sex-developmental genes. With this work, we aim to develop a methodology to add contextual information to NGS data, allowing us to highlight potential gene candidates for further studies.

Methods

Information from Single-cell RNA seq of gonadal cells was used to add gonadal context to the gene candidates. We can use this context to filter variants from undiagnosed patients and prioritize genes for further studies.

Three steps approach:

a) Cell-context Scoring

Single-cell RNA seq of gonadal cells and unrelated tissues was obtained forn publicly available sources as Li et al, Cell Stem Cell, 2017. Data were normalized and ranked based on the expression level (R) giving higher score to genes with higher expression in gonadal cells and lower expression in unrelated tissues. We combined these contextual rankings with gene interaction databases to create a contextual network of the gonadal expression landscape (cell-context network).

b) Clinical data processing

We focused on patients with 46,XY DSD. Our DSD cohort consists of 114 46,XY DSD patients and 43 relatives. After filtering the data and separate variants in known DSD genes, 84 patients (74%) remained without definitive diagnosis.

c) In-depth analysis

Interactions between gene candidates and sex development-related genes was performed via gene ontology enrichment using Toppcluster. The direct interaction networks (Cytoscape, ClueGO, and Pathlinker) were focused on activation/inhibition interactions. Nodes were weighed based on the sum of the R score and the number of direct interactions with known sex development-related genes (I). This score was also normalized as a final score (F).

Networks were examined using DAVID Bioinformatics Resources 6.8. Among the list of gene candidates, we selected those with the highest F score (F ≥ mean+2SD) for in-depth analysis, including literature search and variant modelling (YASARA/Alphafold2).

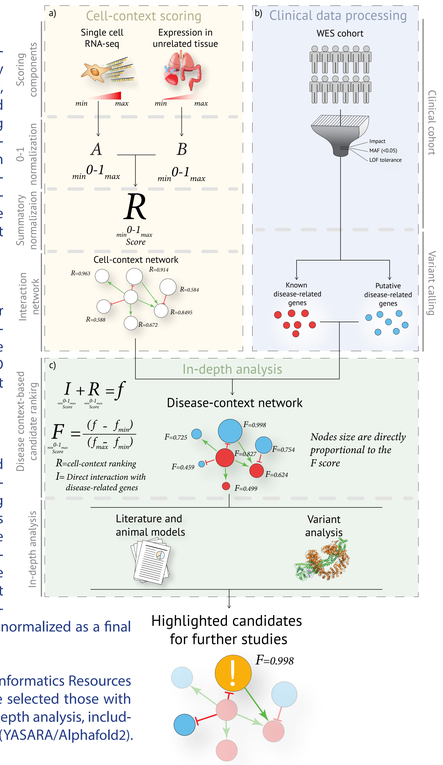


Figure 1: Project pipeline representation

Results

Cell-context Scoring

Genes ranked higher than the mean were considered "relevant" for the cell type function. The list of "relevant genes" includes known SD-genes and novel candidates. Contextual gene ranking of the three cell types in the male gonad. All the analyzed known SD- genes were ranked as relevant genes for at least one of the cell types.

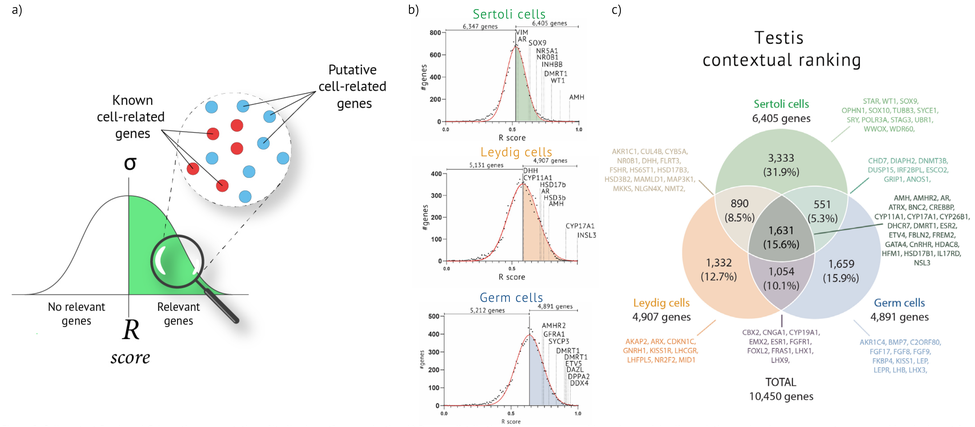


Figure 2: Cell-context Scoring. a) Schematic representation of "cell-relevant" gene selection. b) R score distribution of genes expressed in each gonadal cell population, colored area represent genes considered relevant for the analysis. c) Venn diagram of relevant gene clusters in the male gonad. Known sex development-related genes in each cluster are highlighted.

Data integration and contextual network generation

For each cell type, we generated an interaction network combining the corresponding contextual ranking and the list of sex development-related genes (in red) and novel variants (in blue) from the clinical database. The nodes are weighed based on its F score. Consequently, candidates for in-depth analysis (in yellow) were selected among the genes with higher F score and more direct interactions with known sex development-related genes.

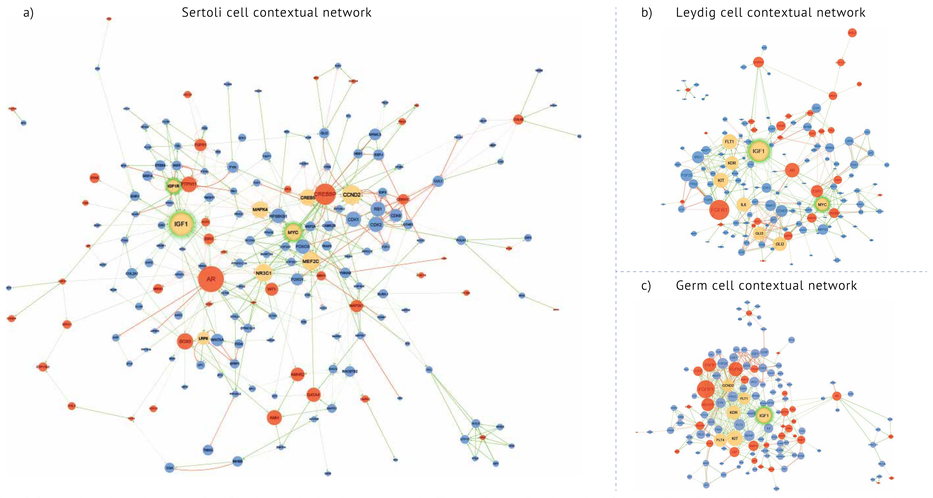
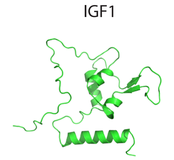


Figure 3: Contextual gene interaction networks for a) Sertoli cells, b) Leydig cells, and c) germ cells. The network shows direct interactions of novel candidates (in blue) with known sex development-related genes (in red). Activations are represented as green arrows and inhibitions as red lines. Node size is correlated with the F score of the gene.

Some examples of gene candidates selected for in-depth analysis:



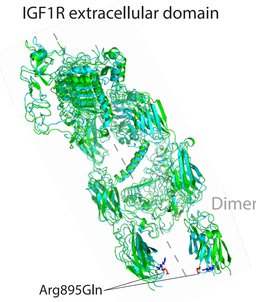
IGF1:

IGF1 activates, through IGF1R, the MAPK and the PI3K/AKT pathways and stimulates growth and survival. In immature SC it seems to play in FSH action. In mice, Igf1 deficiency shows postnatal lethality, growth retardation, infertility, and developmental defects in major organs. In patients IGF1 mutations are known to produce intrauterine growth delay, but no gonadal abnormalities were reported.

IGF1R:

Pathogenic homozygous mutations in the IGF1R gene in humans cause diminished autophosphorylation of IGF1R. Mouse XY gonads lacking both insulin and IGF1 receptors resemble ovaries and are indistinguishable from XX gonads, with complete absence SCs and LCs.

The IGF1R variant observed in our patient (p.Arg895Gln9, in red) is localized at the extracellular domain of the protein, close to the transmembrane domain. Since IGF1R activity depends on the physical proximity of both intracellular kinase domains of the dimer, regulated by the conformational change of the transmembrane domains, mutations close to the transmembrane domain could affect the IGF1R activity independent from the ligand.



MYC:

MYC (also known as C-Myc) plays a role in cell cycle progression, cellular transformation and apoptosis. In humans, upregulation of C-Myc expression induce proliferation of cultured adult human SCs, via WNT/b-catenin signaling. Knockdown of C-Myc affect WNT/β-catenin stimulation of SCs proliferation. Myc mouse models reported Leydig cell hyperplasia, reduced female fertility, and decreased litter size. An FSH-dependent C-Myc regulation had been linked to the control of the SCs cycle. In Drosophila, Daughterless (da), a gene with multiple similarities to C-Myc, has important functions in embryonic development, being necessary for sex determination.

Two patients from our cohort carry variants in MYC. One patient presents with 46,XY DSD and female phenotype and the other with undescended testes and hypospadias. There is no available crystal structure for MYC and homology modelling did not result in an adequate model.

Conclusions

- NGS applicability can be boosted by adding contextual information to the raw data.
- SC-RNA seq and clinical WES data can be combined to increase data value.
- We can prioritize over thousands of potential variants, based on cell and disease context.
- This methodology can be applied to virtually any cell type and human disease, helping to find novel genes and pathways underlying the mechanism of disease.

Contact

Address: Ch. du Musée 5 CH-1700 Fribourg, Switzerland
E-mail: daniel.rodriguez@unifr.ch
Phone: +41 (0)26 300 86 21
Website: http://www.unifr.ch/scimed

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG
UNIL | Université de Lausanne
UNIVERSITÉ DE GENÈVE
FNSNF