cmvdrg - An R package for Human Cytomegalovirus antiviral Drug Resistance Genotyping

Authors: Oscar J Charles[1], Cristina Venturini[1], Judith Breuer[1,2 *]

Affiliations:

[1]Division of Infection and Immunity, University College London, The Cruciform Building, Gower St, London, WC1E 6BT

[2]Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

*Corresponding author. Email: j.breuer@ucl.ac.uk

One Sentence Summary:

Currently data regarding Human Cytomegalovirus resistant mutations are contained in unconnected literature sources, here we present an exhaustive open source database and analysis tool for the community.

Keywords:

Human cytomegalovirus, drug resistance, database, software

Abstract:

The prevention and treatment of HCMV infection is based on the utilization of antiviral therapies as HCMV lacks an effective vaccine. The rise of drug resistance is therefore an increasing patient threat. We identified the need for an open source and comprehensive HCMV resistance mutations database, to support the research community in this area. Here we present “Cytomegalovirus Drug Resistance Genotyping” (cmvdrg), a freely available database contained within an easily accessible R package, which provides a succinct extraction of literature material in the form of a text file database. Additionally, cmvdrg includes methods for calling resistance in common sequencing files and an optional user-friendly web interface.

Availability The cmvdrg package is freely available under the GNU GPL v3 license at https://github.com/ucl-pathgenomics/cmvdrg,

Main text:

Introduction

The prevention and treatment of Human Cytomegalovirus (HCMV) is essential in management of solid organ transplant (SOT), hematopoietic stem cell transplant (HSCT) and other immunocompromised hosts.[1–3]. Although a handful of treatment options are available such as Ganciclovir, Cidofovir and Mirabavir, emergence of HCMV resistance is now found for all current drugs and pose a significant threat due to aggressive disease course and a greater mortality risk[4,5]. Early detection of resistance is important to inform alternative treatments.

Testing clinically for drug resistance mutations depends on previously phenotypically characterised resistant variants and may be reported to differing degrees of confidence. Phenotyping follows two channels[6]. 1 – Mutations identified with a genotypic procedure where resistance sites are associated with treatment failure or increased viral load in HCMV infected cohorts. Potential new resistant mutants are then conveyed with variable levels of statistical rigor or anecdotally. 2 – Mutations identified with in vitro studies by the classic plaque reduction assay, or Bacterial Artificial Chromosome (BAC) reporter assays e.g. SEAP[7]. In these experiments, SNPs are marker transferred to a well characterised lab strain, typically AD169 or TOWNE. The effects of drug concentration can be compared to the original control strain and IC_{50} values obtained; it is typical to report results as a fold change to the control strain. There is currently no consensus cut-off ratio for defining “resistant” or “susceptible” phenotype.

As the cost of sequencing has rapidly decreased, the number of samples sequenced for research studies and clinical diagnostics has increased proportionately becoming commonplace in some institutes[8,9]. Sanger sequencing or PCR amplicons are used in clinical settings and focus on the DNA polymerase gene (UL54) and protein kinase gene (UL97)[4] and are only able to detect drug resistance mutations at 20% or higher[10,11]. The complete picture is however more complex. With Maribavir, Tomeglovir and Letermovir garnering more interest clinically, respective resistant mutants for these drugs occur frequently in genes not sequenced in many diagnostic labs[12], namely UL27, UL51, UL56 and UL89.

Next generation sequencing (NGS), including target enrichment and deep sequencing, confers improved sensitivity compared to conventional methods, makes use of information in the full HCMV genome and allows us to look at mutations at frequency as low as 1%. Detection of low level mutations may eventually become established for clinical use, allowing us to identify resistance mutations early on before rising to fixation and track them over time in longitudinal samples. Accumulation of multiple resistance variants has been also identified as a potential biomarker of poor outcome in immunosuppressed paediatric patients[11,13,14] (Colston & Venturini, 2020, paper in preparation).

Recent reviews have put focus on the field to develop tools and practices that promote the acquisition of well-validated data.[15] For the use in clinical diagnostics, rapid and accurate resistance detection is necessary and user-friendly tools are needed to allow clinicians and clinical scientists to analyse genetic HCMV data.

To the best of our knowledge, currently published tools or processes only allow analysis of FASTA sequences and/or focus only well known resistance genes, such as UL97 and UL54[16,17].

Here we present a new open source tool to analyse NGS HCMV data, optimised for clinical research, able to detect low frequency resistance mutations using an up-to-date database.

Methods and Implementation

Creating the database from literature

Data regarding Gene-Mutation to Resistance phenotype was extracted in a semi-automated fashion from a comprehensive pubmed search with key terms, further filtering used regular expressions to detect the occurrence of Mutations. Journal article data was manually extracted, review articles have been highly valuable in this approach however care was taken to find the original article for mutations in all possible cases. Where original articles & review values were not aligned, the “Status” field was set to “Obsolete” Indicating they are not considered in analysis. Certain co-occurring mutations are known to recover a susceptible phenotype in HCMV, conversely different co-occurring mutations may amplify resistance[18,19]. We recorded these data points when presented in the literature, and are flagged in cmvdrg.

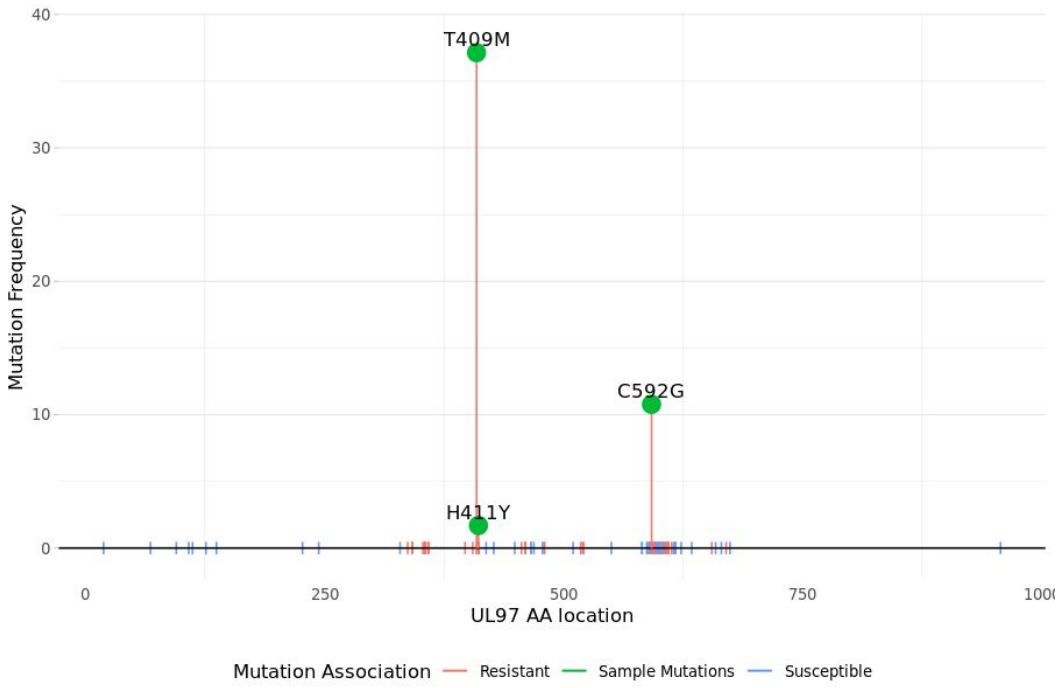

Figure1: A lollipop plot of all mutations in UL54, along with sample variants.

How to use the package

The package can be install in R with the following commands:

1) Install devtools if needed with: install.packages("devtools")
2) Install cmvdrg by running: devtools::install_github("ucl-pathgenomics/cmvdrg"). This command will take care of installing other dependencies.

Inputs for the package are Variant Call Format (VCF >= version4.0) and default Varscan2[20] output (tab delimited), for detection of low frequency mutations. Where the data has been mapped to the HCMV reference strain Merlin (NCBI Reference Sequence: NC_006273.2). Annotation and resistance calling are covered by internal R[21] functions making use of the VariantAnnotation[22] package, and dependencies. The main return object from the "call_resistance" function is a table containing the relationship between identified resistance mutants and any database entries with conveying information.

FASTA files are also accepted by cmvdrg to allow consensus analysis, both for the whole genome and specific genes (e.g. UL54 and UL97). These data do not need to be aligned or assembled to a specific reference allowing for de novo assemblies, those that are of different length compared to Merlin are first aligned using MAFFT with "—add –keeplength" parameters[23], SNPs are extracted from the FASTA alignment in VCF relative to Merlin using snp-sites[24].

Outputs

A concise per drug clinical output is included, which includes:
Resistance phenotype

- High level (any mutation confers a fold change above 15)
- Moderate level (the maximal fold change returned is between 5-15)
- Low resistance (the maximum fold change returned is between 2-5)
- Resistant (only anecdotal resistant data was returned)
- Susceptible (less than 2, or recorded as such)
- NA (no mutations returned)

Evidence strength in 3 categories.

- Good, in vitro
- Weaker, anecdotal
- NA (no evidence)

Cutoff values are in line with recommendations from 'The Third International Consensus Guidelines on the Management of Cytomegalovirus in Solid-organ Transplantation'[12]. We also included a method to return the full data for resistance mutations identified.

Clinical Overview

	Ganciclovir	Cidofovir	Foscarnet	Brincidofovir	Letermovir	Tomeglovir	GW275175	Maribavir	Cyclopropavir
Resistance Phenotype	Low level	Low level	Low level	NA	NA	NA	NA	High level	Low level
Evidence Strength	good, in vitro	good, in vitro	good, in vitro	No Evidence	No Evidence	No Evidence	No Evidence	good, in vitro	good, in vitro

Figure2: Example table showing clinical output produced by cmvdrg

Functionality is extended with a web interface, a hosted version of which is available at http://51.11.13.133:3838/cmvdrg/ ,specific terms of use for the hosted instance are included with the shiny application. The application was developed using the shiny framework.[25] We include example files for whole genome analysis (VCF, varscan tab, FASTA), and specific gene regions FASTA.

The lab has funding to maintain and update the database until 2023. As an open source community resource anyone can maintain the data, or use it in its current form.

Conclusion

The R package cmvdrg offers a set of freely available tools, an intuitive web application and open source database to characterise resistance in HCMV sequencing data, thereby supporting research into HCMV infections by way of a set of community accessible tools.

Funding

This work was supported directly by UKRI MRC grant MR/N013867/1. The lab receives current funding from the Wellcome trust.

Conflicts of interest

None declared.

Acknowledgements

We would like to thank Steven A Kemp and Salvatore Camiolo for their testing and informative suggestions.

References

1. Limaye, A. P. Ganciclovir-Resistant Cytomegalovirus in Organ Transplant Recipients. *Clin. Infect. Dis.* **35**, 866–872 (2002).
2. Ariza-Heredia, E. J., Nesher, L. & Chemaly, R. F. Cytomegalovirus diseases after hematopoietic stem cell transplantation: A mini-review. *Cancer Lett.* **342**, 1–8 (2014).

3. Emery, V. Investigation of CMV disease in immunocompromised patients. *J. Clin. Pathol.* **54**, 84–88 (2001).
4. Lurain, N. S. & Chou, S. Antiviral Drug Resistance of Human Cytomegalovirus. *Clin. Microbiol. Rev.* **23**, 689–712 (2010).
5. Komatsu, T. E., Pikis, A., Naeger, L. K. & Harrington, P. R. Resistance of human cytomegalovirus to ganciclovir/valganciclovir: A comprehensive review of putative resistance pathways. *Antiviral Res.* **101**, 12–25 (2014).
6. Drew, W. L. Cytomegalovirus Resistance Testing: Pitfalls and Problems for the Clinician. *Clin. Infect. Dis.* **50**, 733–736 (2010).
7. Chou, S., Van Wechel, L. C., Lichy, H. M. & Marousek, G. I. Phenotyping of Cytomegalovirus Drug Resistance Mutations by Using Recombinant Viruses Incorporating a Reporter Gene. *Antimicrob. Agents Chemother.* **49**, 2710–2715 (2005).
8. Russell, D. A. Sequencing, Assembling, and Finishing Complete Bacteriophage Genomes. in *Bacteriophages: Methods and Protocols, Volume 3* (eds. Clokie, M. R. J., Kropinski, A. M. & Lavigne, R.) 109–125 (Springer, 2018). doi:10.1007/978-1-4939-7343-9_9.
9. Gu, W., Miller, S. & Chiu, C. Y. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu. Rev. Pathol.* **14**, 319–338 (2019).
10. Sahoo, M. K. *et al.* Detection of Cytomegalovirus Drug Resistance Mutations by Next-Generation Sequencing. *J. Clin. Microbiol.* **51**, 3700–3710 (2013).
11. Houldcroft, C. J. *et al.* Detection of Low Frequency Multi-Drug Resistance and Novel Putative Maribavir Resistance in Immunocompromised Pediatric Patients with Cytomegalovirus. *Front. Microbiol.* **7**, (2016).
12. Kotton, C. N. *et al.* The Third International Consensus Guidelines on the Management of Cytomegalovirus in Solid-organ Transplantation. *Transplantation* **102**, 900–931 (2018).
13. Garrigue, I. *et al.* Contribution of next generation sequencing to early detection of

cytomegalovirus UL97 emerging mutants and viral subpopulations analysis in kidney transplant recipients. *J. Clin. Virol.* **80**, 74–81 (2016).

14. Guermouche, H. *et al.* Characterization of the dynamics of human cytomegalovirus resistance to antiviral drugs by ultra-deep sequencing. *Antiviral Res.* **173**, 104647 (2020).
15. Chou, S. Advances in the genotypic diagnosis of cytomegalovirus antiviral drug resistance. *Antiviral Res.* **176**, 104711 (2020).
16. Chevillotte, M. *et al.* A new tool linking human cytomegalovirus drug resistance mutations to resistance phenotypes. *Antiviral Res.* **85**, 318–327 (2010).
17. López-Aladid, R. *et al.* Improvement in detecting cytomegalovirus drug resistance mutations in solid organ transplant recipients with suspected resistance using next generation sequencing. *PLOS ONE* **14**, e0219701 (2019).
18. Chou, S. Rapid In Vitro Evolution of Human Cytomegalovirus UL56 Mutations That Confer Letermovir Resistance. *Antimicrob. Agents Chemother.* **59**, 6588–6593 (2015).
19. Chou, S., Ercolani, R. J. & Lanier, E. R. Novel Cytomegalovirus UL54 DNA Polymerase Gene Mutations Selected In Vitro That Confer Brincidofovir Resistance. *Antimicrob. Agents Chemother.* **60**, 3845–3848 (2016).
20. Koboldt, D. C. *et al.* VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
21. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* (R Core Team, 2014).
22. Obenchain, V. *et al.* VariantAnnotation : a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* **30**, 2076–2078 (2014).
23. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
24. Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA

alignments. *Microb. Genomics* **2**, (2016).

25. Chang, W., Cheng, J., Yihui, X. & Jonathan, M. *shiny: Web Application Framework for R. R package version 1.3.2*.