

Bioinformatics and Data Science of Pandemics: Introduction

Alexander Schönhuth



Bielefeld University
May 10, 2021

Organization

ORGANIZATION

- ▶ Organization and introduction: *today*
- ▶ Full literature list available: *by May 17*
- ▶ How to present (brief): *May 17*
- ▶ How to write (brief): *shortly before presentations*
- ▶ **Presentations:** *from June 1:*
 - ▶ Each presentation 30-45 minutes (less in case of one presenter)
 - ▶ Up to two presentations per week, if that suits everyone's schedules
 - ▶ If desired/necessary, block seminar day possible as well
- ▶ **Technical Report:** *after presentation:*
 - ▶ Each report 8-15 pages
 - ▶ Optimally, report profits from feedback provided after presentation
 - ▶ Drafts can be submitted for discussion
 - ▶ Improving drafts based on feedback

ORGANIZATION

- ▶ Organization and introduction: *today*
- ▶ Full literature list available: *by May 17*
- ▶ How to present (brief): *May 17*
- ▶ How to write (brief): *shortly before presentations*
- ▶ **Presentations:** *from June 1:*
 - ▶ Each presentation 30-45 minutes (less in case of one presenter)
 - ▶ Up to two presentations per week, if that suits everyone's schedules
 - ▶ If desired/necessary, block seminar day possible as well
- ▶ **Technical Report:** *after presentation:*
 - ▶ Each report 8-15 pages
 - ▶ Optimally, report profits from feedback provided after presentation
 - ▶ Drafts can be submitted for discussion
 - ▶ Improving drafts based on feedback

ORGANIZATION

- ▶ Organization and introduction: *today*
- ▶ Full literature list available: *by May 17*
- ▶ How to present (brief): *May 17*
- ▶ How to write (brief): *shortly before presentations*
- ▶ **Presentations:** *from June 1:*
 - ▶ Each presentation 30-45 minutes (less in case of one presenter)
 - ▶ Up to two presentations per week, if that suits everyone's schedules
 - ▶ If desired/necessary, block seminar day possible as well
- ▶ **Technical Report:** *after presentation:*
 - ▶ Each report 8-15 pages
 - ▶ Optimally, report profits from feedback provided after presentation
 - ▶ Drafts can be submitted for discussion
 - ▶ Improving drafts based on feedback

Motivation

MOTIVATION

- ▶ The Covid-19 crisis is a “bioinformatics” crisis
- ▶ Have a (truly) scientific look at topics of current interest
- ▶ The following is a list of suggested topics
- ▶ If you wish to discuss a different topic, please suggest
- ▶ In any case: scientific papers that deal with topic required

MOTIVATION

- ▶ The Covid-19 crisis is a “bioinformatics” crisis
- ▶ Have a (truly) scientific look at topics of current interest
- ▶ The following is a list of suggested topics
- ▶ If you wish to discuss a different topic, please suggest
- ▶ In any case: scientific papers that deal with topic required

MOTIVATION

- ▶ The Covid-19 crisis is a “bioinformatics” crisis
- ▶ Have a (truly) scientific look at topics of current interest
- ▶ The following is a list of suggested topics
- ▶ If you wish to discuss a different topic, please suggest
- ▶ In any case: scientific papers that deal with topic required

Topics

TOPICS

- ▶ PCR Test Design
- ▶ The Origin of SARS-CoV-2
- ▶ Virus Genome Assembly / Evolutionary Tracking
- ▶ Infection Wave Dynamics
- ▶ Drug Repurposing
- ▶ Any further? ...

TOPICS

- ▶ PCR Test Design
- ▶ The Origin of SARS-CoV-2
- ▶ Virus Genome Assembly / Evolutionary Tracking
- ▶ Infection Wave Dynamics
- ▶ Drug Repurposing
- ▶ Any further? ...

TOPICS

- ▶ PCR Test Design
- ▶ The Origin of SARS-CoV-2
- ▶ Virus Genome Assembly / Evolutionary Tracking
- ▶ Infection Wave Dynamics
- ▶ Drug Repurposing
- ▶ Any further? ...

TOPICS

- ▶ PCR Test Design
- ▶ The Origin of SARS-CoV-2
- ▶ Virus Genome Assembly / Evolutionary Tracking
- ▶ Infection Wave Dynamics
- ▶ Drug Repurposing
- ▶ Any further? ...

TOPICS

- ▶ PCR Test Design
- ▶ The Origin of SARS-CoV-2
- ▶ Virus Genome Assembly / Evolutionary Tracking
- ▶ Infection Wave Dynamics
- ▶ Drug Repurposing
- ▶ Any further? ...

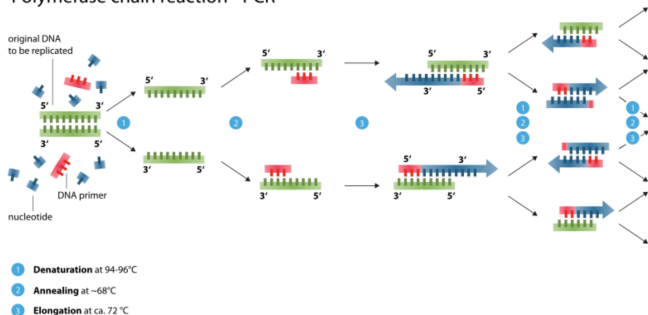
TOPICS

- ▶ PCR Test Design
- ▶ The Origin of SARS-CoV-2
- ▶ Virus Genome Assembly / Evolutionary Tracking
- ▶ Infection Wave Dynamics
- ▶ Drug Repurposing
- ▶ Any further? ...

PCR Test Design

PCR TEST DESIGN

Polymerase chain reaction - PCR

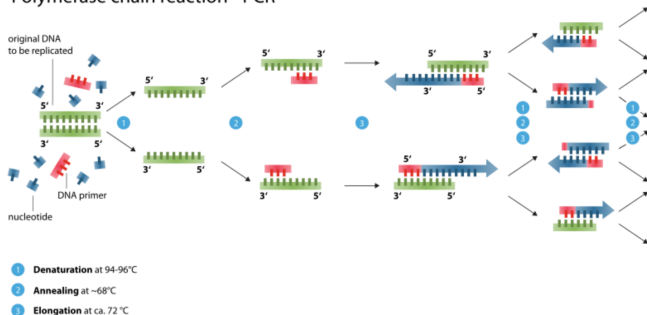


From: www.the-odin.com

- ▶ Primers need to anneal to target sequence
- ▶ Primers must not anneal against other sequences (related viruses)
- ▶ How to safely select such primers?

PCR TEST DESIGN

Polymerase chain reaction - PCR

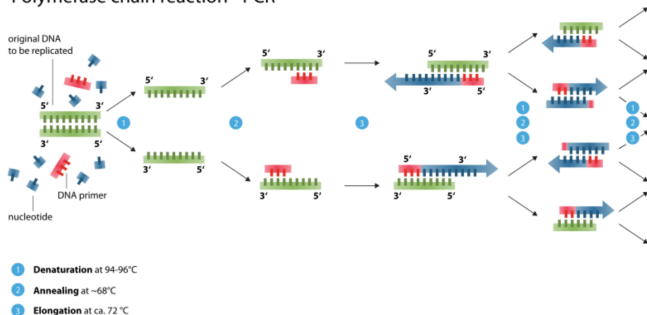


From: www.the-odin.com

- ▶ Primers need to anneal to target sequence
- ▶ Primers must not anneal against other sequences (related viruses)
- ▶ How to safely select such primers?

PCR TEST DESIGN

Polymerase chain reaction - PCR



From: www.the-odin.com

- ▶ Primers need to anneal to target sequence
- ▶ Primers must not anneal against other sequences (related viruses)
- ▶ How to safely select such primers?

PCR TEST DESIGN: PAPERS

- ▶ Tutorial (video):
https://www.youtube.com/watch?v=c-f1H07D_70
- ▶ Primer-BLAST: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-134>
- ▶ Primer3: <https://academic.oup.com/nar/article/40/15/e115/1223759>
- ▶ Drosten-Test: <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>
- ▶ Public Review: <https://cormandrostenreview.com/report/>

PCR TEST DESIGN: PAPERS

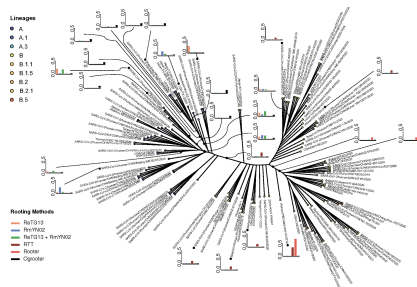
- ▶ **Tutorial (video):**
https://www.youtube.com/watch?v=c-f1H07D_70
- ▶ **Primer-BLAST:** <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-134>
- ▶ **Primer3:** <https://academic.oup.com/nar/article/40/15/e115/1223759>
- ▶ **Drosten-Test:** <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>
- ▶ **Public Review:** <https://cormandrostenreview.com/report/>

PCR TEST DESIGN: PAPERS

- ▶ **Tutorial (video):**
https://www.youtube.com/watch?v=c-f1H07D_70
- ▶ **Primer-BLAST:** <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-134>
- ▶ **Primer3:** <https://academic.oup.com/nar/article/40/15/e115/1223759>
- ▶ **Drosten-Test:** <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>
- ▶ **Public Review:** <https://cormandrostenreview.com/report/>

The Origin of SARS-CoV-2

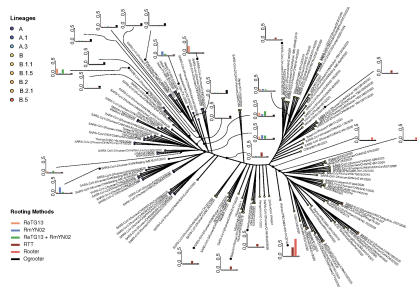
ORIGIN



From: www.the-odin.com

- ▶ What are the viruses that are related with SARS-CoV-2 and how are they related?
- ▶ What are the driving evolutionary forces having shaped them (mutation, recombination)?
- ▶ How can we determine all of this and how certain can we be about our conclusions?

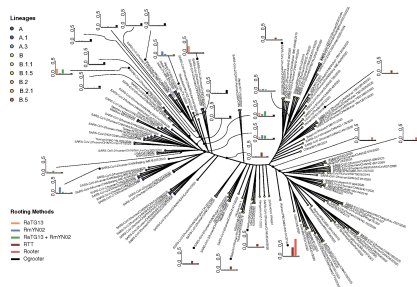
ORIGIN



From: www.the-odin.com

- ▶ What are the viruses that are related with SARS-CoV-2 and how are they related?
- ▶ What are the driving evolutionary forces having shaped them (mutation, recombination)?
- ▶ How can we determine all of this and how certain can we be about our conclusions?

ORIGIN



From: www.the-odin.com

- ▶ What are the viruses that are related with SARS-CoV-2 and how are they related?
- ▶ What are the driving evolutionary forces having shaped them (mutation, recombination)?
- ▶ How can we determine all of this and how certain can we be about our conclusions?

ORIGIN: PAPERS

- ▶ **Recent reviews:**
<https://doi.org/10.1007/s10311-020-01151-1>
- ▶ **Relation of virus lineage with other viruses:**
<https://doi.org/10.1038/s41564-020-0771-4>
- ▶ **Uncertainty of origin (Nielsen):**
<https://doi.org/10.1093/molbev/msaa316>
- ▶ **Latest insights from China:**
<https://doi.org/10.1101/2021.03.08.434390>

ORIGIN: PAPERS

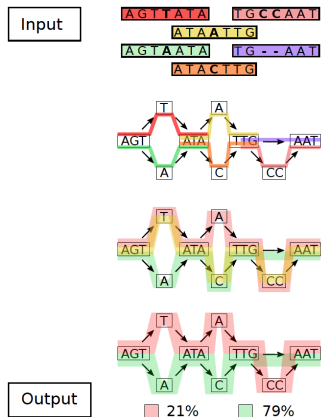
- ▶ **Recent reviews:**
<https://doi.org/10.1007/s10311-020-01151-1>
- ▶ **Relation of virus lineage with other viruses:**
<https://doi.org/10.1038/s41564-020-0771-4>
- ▶ **Uncertainty of origin (Nielsen):**
<https://doi.org/10.1093/molbev/msaa316>
- ▶ **Latest insights from China:**
<https://doi.org/10.1101/2021.03.08.434390>

ORIGIN: PAPERS

- ▶ **Recent reviews:**
<https://doi.org/10.1007/s10311-020-01151-1>
- ▶ **Relation of virus lineage with other viruses:**
<https://doi.org/10.1038/s41564-020-0771-4>
- ▶ **Uncertainty of origin (Nielsen):**
<https://doi.org/10.1093/molbev/msaa316>
- ▶ **Latest insights from China:**
<https://doi.org/10.1101/2021.03.08.434390>

Virus Genome Assembly Evolutionary Tracking

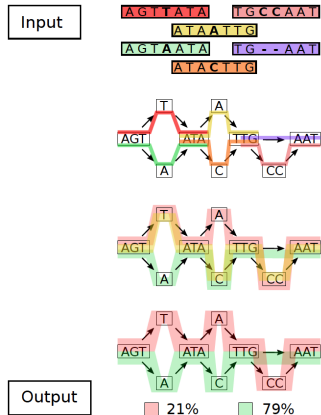
ASSEMBLY AND TRACKING



Constructing virus variation graph from input reads

- ▶ Viruses tend to infect hosts as ensemble of strains
- ▶ How to reconstruct all strains from an individual RNA/DNA sample?
- ▶ Once strains are reconstructed, how to put them into mutual context?
- ▶ Fast inspection of mutational status of infected people
- ▶ Overview of evolutionary status of pandemic
- ▶ Use graphs or trees

ASSEMBLY AND TRACKING



Constructing virus variation graph from input reads

- ▶ Viruses tend to infect hosts as ensemble of strains
- ▶ How to reconstruct all strains from an individual RNA/DNA sample?
- ▶ Once strains are reconstructed, how to put them into mutual context?
- ▶ Fast inspection of mutational status of infected people
- ▶ Overview of evolutionary status of pandemic
- ▶ Use graphs or trees

ASSEMBLY AND TRACKING: PAPERS

- ▶ Assembly:

- ▶ **Viral quasispecies assembly:** <https://genome.cshlp.org/content/27/5/835.full.html>
- ▶ Full-length assembly using variation graphs: <https://doi.org/10.1093/bioinformatics/btz443>

- ▶ Tracking:

- ▶ The Pantograph project (variation graphs): <https://graphgenome.org/>
- ▶ Phylogenetic trees ("nextstrain"): <https://doi.org/10.1093/bioinformatics/bty407>

ASSEMBLY AND TRACKING: PAPERS

- ▶ Assembly:

- ▶ **Viral quasispecies assembly:** <https://genome.cshlp.org/content/27/5/835.full.html>
- ▶ **Full-length assembly using variation graphs:** <https://doi.org/10.1093/bioinformatics/btz443>

- ▶ Tracking:

- ▶ The Pantograph project (variation graphs): <https://graphgenome.org/>
- ▶ Phylogenetic trees ("nextstrain"): <https://doi.org/10.1093/bioinformatics/bty407>

ASSEMBLY AND TRACKING: PAPERS

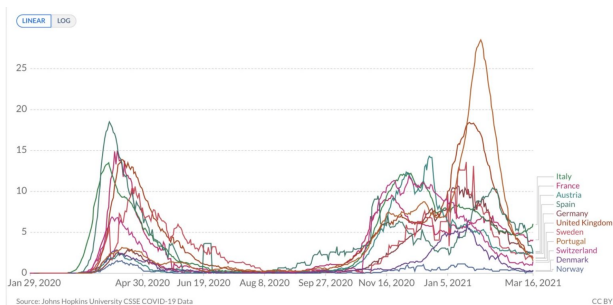
- ▶ Assembly:
 - ▶ **Viral quasispecies assembly:** <https://genome.cshlp.org/content/27/5/835.full.html>
 - ▶ **Full-length assembly using variation graphs:** <https://doi.org/10.1093/bioinformatics/btz443>
- ▶ Tracking:
 - ▶ **The Pantograph project (variation graphs):** <https://graphgenome.org/>
 - ▶ **Phylogenetic trees (“nextstrain”):** <https://doi.org/10.1093/bioinformatics/bty407>

ASSEMBLY AND TRACKING: PAPERS

- ▶ Assembly:
 - ▶ **Viral quasispecies assembly:** <https://genome.cshlp.org/content/27/5/835.full.html>
 - ▶ **Full-length assembly using variation graphs:** <https://doi.org/10.1093/bioinformatics/btz443>
- ▶ Tracking:
 - ▶ **The Pantograph project (variation graphs):** <https://graphgenome.org/>
 - ▶ **Phylogenetic trees (“nextstrain”):** <https://doi.org/10.1093/bioinformatics/bty407>

Infection Wave Dynamics

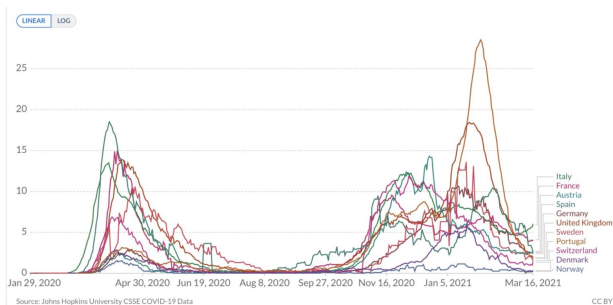
INFECTION WAVE DYNAMICS



Covid-19 death count curves per countries in Europe

- ▶ Counts of infected/died people take particular shapes
- ▶ Why is that? Are there reliable models?
- ▶ How to estimate infection rate? (R-value)
- ▶ Do the curves behave in known / controllable ways?

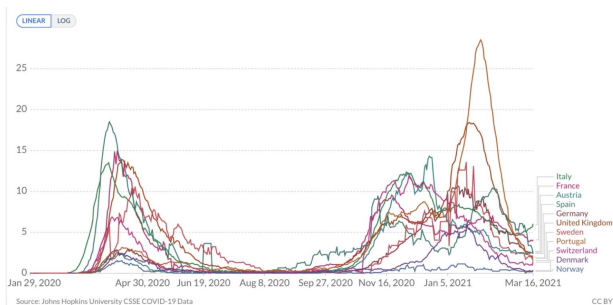
INFECTION WAVE DYNAMICS



Covid-19 death count curves per countries in Europe

- ▶ Counts of infected/died people take particular shapes
- ▶ Why is that? Are there reliable models?
- ▶ How to estimate infection rate? (R -value)
- ▶ Do the curves behave in known / controllable ways?

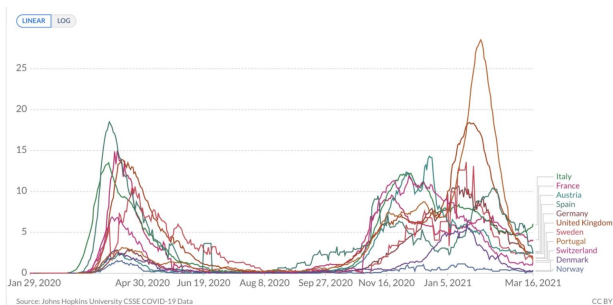
INFECTION WAVE DYNAMICS



Covid-19 death count curves per countries in Europe

- ▶ Counts of infected/died people take particular shapes
- ▶ Why is that? Are there reliable models?
- ▶ How to estimate infection rate? (R-value)
- ▶ Do the curves behave in known / controllable ways?

INFECTION WAVE DYNAMICS



Covid-19 death count curves per countries in Europe

- ▶ Counts of infected/died people take particular shapes
- ▶ Why is that? Are there reliable models?
- ▶ How to estimate infection rate? (R-value)
- ▶ Do the curves behave in known / controllable ways?

INFECTION WAVE DYNAMICS: PAPERS

- ▶ The R-Value:
 - ▶ Popular overview: <https://www.nature.com/articles/d41586-020-02009-w>
 - ▶ Early dynamics of COVID-19: [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
 - ▶ Stochastic non-linear dynamical systems: <https://doi.org/10.1093/biostatistics/kxs052>
- ▶ Curve Fitting:
 - ▶ Gompertz curves: <https://doi.org/10.1093/ptep/ptaa148>
 - ▶ Short-term prediction of Covid-19 spreading: <https://doi.org/10.1371/journal.pcbi.1008431>

INFECTION WAVE DYNAMICS: PAPERS

- ▶ The R-Value:
 - ▶ Popular overview: <https://www.nature.com/articles/d41586-020-02009-w>
 - ▶ Early dynamics of COVID-19: [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
 - ▶ Stochastic non-linear dynamical systems: <https://doi.org/10.1093/biostatistics/kxs052>
- ▶ Curve Fitting:
 - ▶ Gompertz curves: <https://doi.org/10.1093/ptep/ptaa148>
 - ▶ Short-term prediction of Covid-19 spreading: <https://doi.org/10.1371/journal.pcbi.1008431>

INFECTION WAVE DYNAMICS: PAPERS

- ▶ The R-Value:
 - ▶ Popular overview: <https://www.nature.com/articles/d41586-020-02009-w>
 - ▶ Early dynamics of COVID-19: [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
 - ▶ Stochastic non-linear dynamical systems: <https://doi.org/10.1093/biostatistics/kxs052>
- ▶ Curve Fitting:
 - ▶ Gompertz curves: <https://doi.org/10.1093/ptep/ptaa148>
 - ▶ Short-term prediction of Covid-19 spreading: <https://doi.org/10.1371/journal.pcbi.1008431>

INFECTION WAVE DYNAMICS: PAPERS

- ▶ The R-Value:
 - ▶ Popular overview: <https://www.nature.com/articles/d41586-020-02009-w>
 - ▶ Early dynamics of COVID-19: [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
 - ▶ Stochastic non-linear dynamical systems: <https://doi.org/10.1093/biostatistics/kxs052>
- ▶ Curve Fitting:
 - ▶ Gompertz curves: <https://doi.org/10.1093/ptep/ptaa148>
 - ▶ Short-term prediction of Covid-19 spreading: <https://doi.org/10.1371/journal.pcbi.1008431>

Infection Fatality Rate

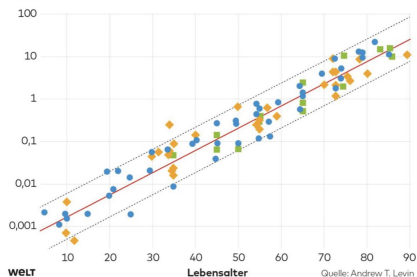
INFECTION FATALITY RATE

Infektionssterblichkeit nach Alter

(Regressionsanalyse)

- Vorhersage-Intervall
- Studien mit repräsentativen Proben und Antikörperbestimmung
- Studien aus Ländern mit wenigen Fällen und umfassender Kontaktverfolgung
- ◆ Sonstige Studien

Sterberate in Prozent



Infection Fatality Rate (IFR) per age / study

- ▶ When infected, how likely does one die?
- ▶ What role does age play?
- ▶ What are the techniques to provide reliable estimates?

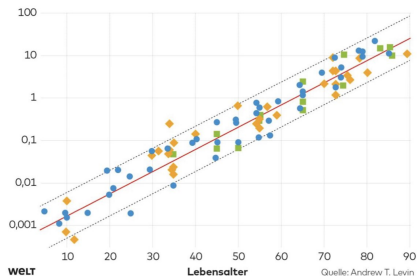
INFECTION FATALITY RATE

Infektionssterblichkeit nach Alter

(Regressionsanalyse)

- Vorhersage-Intervall
- Studien mit repräsentativen Proben und Antikörperbestimmung
- Studien aus Ländern mit wenigen Fällen und umfassender Kontaktverfolgung
- ◆ Sonstige Studien

Sterberate in Prozent



Infection Fatality Rate (IFR) per age / study

- ▶ When infected, how likely does one die?
- ▶ What role does age play?
- ▶ What are the techniques to provide reliable estimates?

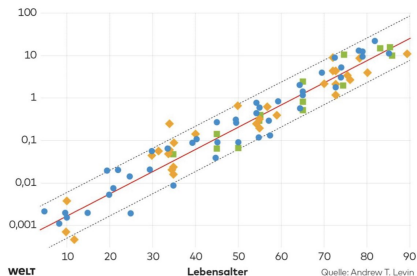
INFECTION FATALITY RATE

Infektionssterblichkeit nach Alter

(Regressionsanalyse)

- Vorhersage-Intervall
- Studien mit repräsentativen Proben und Antikörperbestimmung
- Studien aus Ländern mit wenigen Fällen und umfassender Kontaktverfolgung
- ◆ Sonstige Studien

Sterberate in Prozent



Infection Fatality Rate (IFR) per age / study

- ▶ When infected, how likely does one die?
- ▶ What role does age play?
- ▶ What are the techniques to provide reliable estimates?

IFR ESTIMATION: PAPERS

- ▶ **Heinsberg Study:**

<https://doi.org/10.1038/s41467-020-19509-y>

- ▶ **Geneva Study:**

[https://doi.org/10.1016/S0140-6736\(20\)31304-0](https://doi.org/10.1016/S0140-6736(20)31304-0)

- ▶ **Most recent meta-study (Ioannidis):**

<https://doi.org/10.1111/eci.13554>

IFR ESTIMATION: PAPERS

- ▶ **Heinsberg Study:**

<https://doi.org/10.1038/s41467-020-19509-y>

- ▶ **Geneva Study:**

[https://doi.org/10.1016/S0140-6736\(20\)31304-0](https://doi.org/10.1016/S0140-6736(20)31304-0)

- ▶ **Most recent meta-study (Ioannidis):**

<https://doi.org/10.1111/eci.13554>

IFR ESTIMATION: PAPERS

- ▶ Heinsberg Study:

<https://doi.org/10.1038/s41467-020-19509-y>

- ▶ Geneva Study:

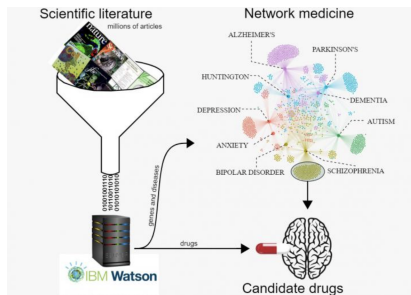
[https://doi.org/10.1016/S0140-6736\(20\)31304-0](https://doi.org/10.1016/S0140-6736(20)31304-0)

- ▶ Most recent meta-study (Ioannidis):

<https://doi.org/10.1111/eci.13554>

Drug Repurposing

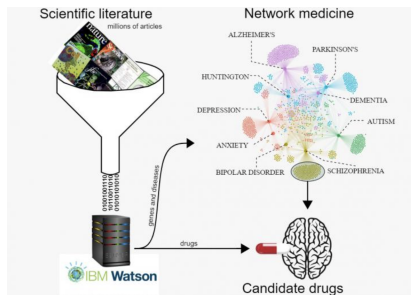
DRUG REPURPOSING



From: eurekaalert.org

- ▶ Drugs act against molecules (proteins) of particular shapes
- ▶ Different molecules may assume similar shapes
- ▶ Find such molecules in the context of particular disease

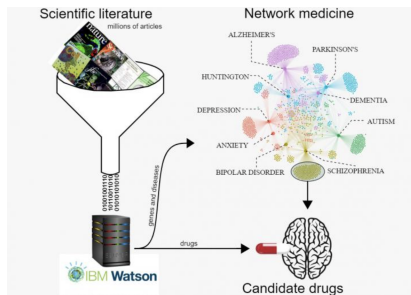
DRUG REPURPOSING



From: eurekaalert.org

- ▶ Drugs act against molecules (proteins) of particular shapes
- ▶ Different molecules may assume similar shapes
- ▶ Find such molecules in the context of particular disease

DRUG REPURPOSING



From: eurekaalert.org

- ▶ Drugs act against molecules (proteins) of particular shapes
- ▶ Different molecules may assume similar shapes
- ▶ Find such molecules in the context of particular disease

DRUG REPURPOSING: PAPERS

- ▶ **Review:** <https://doi.org/10.1002/med.21728>
- ▶ Early study (nice overview):
<https://doi.org/10.1038/s41421-020-0153-3>
- ▶ Recent study (Barabasi):
<https://doi.org/10.1073/pnas.2025581118>
- ▶ Knowledge graph completion:
<https://doi.org/10.1016/j.jbi.2021.103696>
- ▶ Deep generative graph model:
<https://arxiv.org/abs/2007.02338>

DRUG REPURPOSING: PAPERS

- ▶ Review: <https://doi.org/10.1002/med.21728>
- ▶ Early study (nice overview):
<https://doi.org/10.1038/s41421-020-0153-3>
- ▶ Recent study (Barabasi):
<https://doi.org/10.1073/pnas.2025581118>
- ▶ Knowledge graph completion:
<https://doi.org/10.1016/j.jbi.2021.103696>
- ▶ Deep generative graph model:
<https://arxiv.org/abs/2007.02338>

DRUG REPURPOSING: PAPERS

- ▶ Review: <https://doi.org/10.1002/med.21728>
- ▶ Early study (nice overview):
<https://doi.org/10.1038/s41421-020-0153-3>
- ▶ Recent study (Barabasi):
<https://doi.org/10.1073/pnas.2025581118>
- ▶ Knowledge graph completion:
<https://doi.org/10.1016/j.jbi.2021.103696>
- ▶ Deep generative graph model:
<https://arxiv.org/abs/2007.02338>

Thanks for your attention!