Determinants of SARS-CoV-2 transmission to guide vaccination strategy in an urban area

Sarah C. Brüningk[1,2,*], Juliane Klatt[1,2,*], Madlen Stange[2,3,5,*], Alfredo Mari[2,3,5], Myrta Brunner[3,4], Tim-Christoph Roloff[2,3,5], Helena M.B. Seth-Smith[2,3,5], Michael Schweitzer[3,5], Karoline Leuzinger[6,15], Kirstine K. Søgaard[3,5], Diana Albertos Torres[3,5], Alexander Gensch[3], Ann-Kathrin Schlotterbeck[3], Christian H. Nickel[7], Nicole Ritz[9], Ulrich Heininger[9], Julia Bielicki[9], Katharina Rentsch[10], Simon Fuchs[12], Roland Bingisser[7], Martin Siegemund[11], Hans Pargger[11], Diana Ciardo[13], Olivier Dubuis[13], Andreas Buser[14], Sarah Tschudin-Sutter[8], Manuel Battegay[8], Rita Schneider-Sliwa[4], Karsten M. Borgwardt[1,2,+], Hans H. Hirsch[6,8,15,+], Adrian Egli[3,5,+,†]

[1]*Machine Learning & Computational Biology, Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland*

[2]*Swiss Institute for Bioinformatics (SIB), Lausanne, Switzerland*

[3]*Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland*

[4]*Human Geography, Department of Environmental Sciences, University of Basel, Basel, Switzerland*

[5]*Clinical Bacteriology and Mycology, University Hospital Basel & University of Basel, Basel, Switzerland*

[6]*Clinical Virology, University Hospital Basel & University of Basel, Basel, Switzerland*

[7]*Emergency Department, University Hospital Basel, Basel, Switzerland*

[8]*Infectious Diseases and Hospital Epidemiology, University Hospital Basel & University of Basel,*

Basel, Switzerland

[9]*Paediatric Infectious Diseases and Vaccinology, University of Basel Children's Hospital Basel and University of Basel, Basel, Switzerland*

[10]*Laboratory Medicine, University Hospital Basel, Basel, Switzerland*

[11]*Intensive Care Medicine, University Hospital Basel, Basel, Switzerland*

[12]*Health Services for the Basel-City, Basel, Switzerland*

[13]*Viollier AG, Allschwil, Switzerland*

[14]*Regional Blood Transfusion Service, Swiss Red Cross, Basel, Switzerland*

[15]*Transplantation & Clinical Virology, Department Biomedicine, University of Basel, Basel, Switzerland*

**These authors contributed equally to this work*

+These authors share senior authorship

†*corresponding author*

Research in context

Evidence before this study: We searched Google Scholar, PubMed, and medRxiv for articles with the keywords 'SARS-CoV-2', 'transmission', 'phylogeny', 'city', and 'model' as of February 2021 and evaluated the relevant abstracts and study details where appropriate. The identified studies predominantly evaluated SARS-CoV-2 transmission within large metropolises with only a few studies reporting on European cities, including analyses of London and Geneva. We noted, that studies (even outside Europe) generally differentiated between either phylogenetic

analysis or a dynamic modelling approach and did not combine the two, despite the different aspects described. Moreover, the majority of studies evaluate publicly available data for modelling SARS-CoV-2 transmission and to estimate effective reproductive numbers. Such analyses naturally lack a direct relation of cases and relevant socioeconomic, or geographical information on a per case basis. To date, whole-genome-sequencing information has not been used to clearly identify inherently related cases for modelling transmission behaviour in any of the studies assessed. Finally, of the modelling studies reported, none provided a data-driven estimate of vaccine scenarios in a city, for which all model parameters were identified from the same source.

Added value of this study: In this comprehensive study we examine SARS-CoV-2 transmission patterns within a medium-sized European city. We benefit from a rich data set allowing us to combine phylogenetic clustering with compartmental modelling based exclusively on sequenced and inherently related cases that are mapped to a residential address, and hence coupled with socioeconomic and demographic information. This allows us to evaluate population groups driving SARS-CoV-2 transmission and to quantify relevant effective reproductive numbers. Moreover, we identify both traceable and cryptic transmission chains allowing us to suggest which measures (such as vaccination vs. effective contact tracing) would be most effective for each group.

Implications of all the available evidence: Depending on the specific characteristics of the vaccine, we estimate that vaccination of exclusively the senior population groups will reduce intensive-care-unit occupancy, but overall case number would more effectively be contained

by prioritising highly mobile population groups from a socioeconomically weaker background (i.e. transmission drivers), even in the context of comparably shallow socioeconomic gradients within a wealthy European urban area. We identified predominantly clustered transmission amongst more senior or more affluent and less mobile population groups, which implies that extensive testing strategies could be used to more effectively prevent SARS-CoV-2 transmission among these groups. By contrast, mobile, low-income population groups were characterized by cryptic transmission. These are very important findings which could be considered in future vaccine prioritization designs for comparable urban areas.

Background. Transmission chains within small urban areas (accommodating~30% of the European population) greatly contribute to case burden and economic impact during the ongoing COVID-19 pandemic, and should be a focus for preventive measures to achieve containment. Here, at very high spatio-temporal resolution, we analysed determinants of SARS-CoV-2 transmission in a European urban area, Basel-City (Switzerland).

Methods. We combined detailed epidemiological, intra-city mobility, and socioeconomic data-sets with whole-genome-sequencing during the first SARS-CoV-2 wave. For this, we succeeded in sequencing 44% of all reported cases from Basel-City and performed phylogenetic clustering and compartmental modelling based on the dominating viral variant (B.1-C15324T; 60% of cases) to identify drivers and patterns of transmission. Based on these results we simulated vaccination scenarios and corresponding healthcare-system burden (intensive-care-unit occupancy).

Findings. Transmissions were driven by socioeconomically weaker and highly mobile population groups with mostly cryptic transmissions, whereas amongst more senior population transmission was clustered. Simulated vaccination scenarios assuming 60-90% transmission reduction, and 70-90% reduction of severe cases showed that prioritizing mobile, socioeconomically weaker populations for vaccination would effectively reduce case numbers. However, long-term intensive-care-unit occupation would also be effectively reduced if senior population groups were prioritized, provided there were no changes in testing and prevention strategies.

Interpretation. Reducing SARS-CoV-2 transmission through vaccination strongly depends on the efficacy of the deployed vaccine. A combined strategy of protecting risk groups by extensive testing coupled with vaccination of the drivers of transmission (i.e. highly mobile groups) would be most effective at reducing the spread of SARS-CoV-2 within an urban area.

Funding: No dedicated funding was used for this work.

Introduction

Efforts to understand transmission of SARS-CoV-2 have been undertaken at different scales including at a global level[1–3], across continents (Europe and North America[4]), within countries (Austria[5], Brazil[6], France[7], Iceland[8], South Africa[9], Thailand[10]) and in large cities (Beijing[11], Boston[12], Houston[13], and New York City[14–16]). In Europe, ~30% of the population live in small urban areas ($10k - 300k$ inhabitants)[17], which accordingly play a major role in SARS-CoV-2 transmission yet have not been studied. Moreover, to date city-based studies of SARS-CoV-

2 transmission have very limited resolution in terms of the proportion of sequenced positive cases (incomplete transmission chains), have a paucity of socioeconomic or mobility data (incomplete determinants), or fail to combine analysis of transmission clusters with quantitative, descriptive models accounting for population mixing[18]. Of those studies describing the distribution of cases together with changes in mobility, none to date rigorously study socioeconomic differences between city quarters as determinants of transmission[13,15,16,19]. An integrated model considering all of these factors (including epidemiological, geographic, mobility, socioeconomic, transmission dynamics information) is anticipated to provide profound insights into the determinants underpinning transmission, that can be used to guide the delivery of vaccines. We here present such an integrated analysis for Basel-City which is part of a metropolitan area, a functional urban area, and a European cross-border area as outlined in the supplement. Basel-City is hence representative of other areas in the EU classified as such.

Local interventions are most effective in cutting transmission chains in families, and small community networks[20,21] that represent well-defined (phylogenetic or epidemiological) clusters. However, most infections are acquired from unknown sources and transmitted cryptically making it essential to identify the key determinants and transmission routes at the city-level to improve interventions and vaccination campaigns. To address this challenge, we have combined phylogenetic cluster analysis[22] and compartmental ordinary differential equation (ODE) modelling based on high density (81% of reported cases assessed), and high resolution (spatial:housing blocks, temporal: day-by-day) epidemiological, mobility, socioeconomic, and serology (estimate unreported cases) data-sets from the first COVID-19 wave (February-April

2020). Whole genome sequencing (WGS) of all included cases (44% of all cases successful) allowed the analysis be restricted to a single, dominant viral variant (B.1-C15324T, 60% of cases [23]). This ensured that our analyses focused on inherently related cases and enabled estimation of effective reproductive numbers for different socioeconomic and demographic population subgroups to provide the basis for vaccination scenario building.

Methods

Detailed methods are given in the supplements.

Included data. All analyses were based on PCR-positive (750/7073 tests) cases of residents of Basel-City between February 25th and April 22nd 2020, obtained from the University Hospital Basel, covering 81% of all reported cases in the relevant interval and location. All samples were subjected to WGS, 53% resulted in high quality genomes (i.e. 44% of all cases). Of these 247 (247/411, 60%) contained the monophyletic C15324T mutation in the B.1 lineage (B.1-C15324T) and were used for further analysis. Each case was linked to the patient's place of residence anonymized to one of 1,078 statistical housing-blocks. For each of these housing-blocks, where privacy legislation permitted, Basel-City's Cantonal Statistical Office provided socioeconomic indicators for the year of 2017 (most recent available). These included (i) living space (per capita in m^2), (ii) share of 1-person private households, (iii) median income (CHF), and (iv) population seniority (percentage of citizens aged over 64 years). According to these indicators, blocks were allocated to one of three socioeconomic tertiles (T1:≤33rd percentile, T2: 33rd to 66th

percentile, T3:>66th percentile, N/A: no available data or censored) where possible (e.g. Figure 2A). Generally, sparsely populated blocks displayed a maximum of three positive cases and had to be excluded from analysis. All following analyses with respect to socioeconomic factors were based on these partitions.

We determined SARS-CoV-2 antibody responses in 2,019 serum samples collected between 25th of February and 22nd of May 2020 to account for delayed seroconversion (see supplement for details). An estimated 1.9% (38/2,019) of the Basel-City population was infected with SARS-CoV-2, corresponding to 88% unreported cases for the sequenced B.1-C15324T variant (see supplement).

Finally, we included data on the number and age distribution of COVID-19 intensive-care unit (ICU) patients during the relevant period from University Hospital Basel, a tertiary hospital with a capacity of 44 ICU beds: 4.5% of reported SARS-CoV-2 positive cases were admitted to ICU and median length of ICU stay was 5.9 days (IQR, 1.5-12.9). 40% of these patients were younger than 64 years.

Phylogenetic inference and cluster analysis. SARS-CoV-2 genomes were phylogenetically analysed in a global context as described previously[23] (see supplements). Phylogenetic clusters were consolidated with epidemiological data (occupation in a health service job, resident of a care home, contact to positive cases, onset of symptoms, place of infection) to confirm the suitability of the divergence parameter chosen and then combined with ancillary geographic (quarter), and socioeconomic information as described above. We inferred statistical significance for

clusters in respective tertiles.

Mobility data. We employed the official traffic model provided by the traffic department of Basel-City [24] consisting of the 2016 average A-to-B traffic on a grid of ~1400 counting zones for foot, bike, public motorized and private motorized transport. We computed the spatial variation of mobility within and between each of the socioeconomic partitions (see supplement) resulting in a unity-normalized three-by-three mobility matrix M_{jk} representing relative within-tertile/inter-tertile mobility on/off its diagonal. Additionally, weekly averages of pass-by traffic for combined foot-bike traffic, as well as private motorized traffic were obtained together with weekly public-transport passenger loads (from SBB Swiss Federal Railways and Basel-Verkehrsbetriebe). These were combined in a weighted sum according to the relevant transport mode contribution, normalized and smoothed with a uni-variate spline to obtain the final temporal mobility variation, $\alpha_{mob}(t)$).

Dynamic changes in social interaction. SARS-CoV-2 transmission is contact-based. While the number of contacts is largely influenced by human mobility, the risk of a contact becoming a transmission event is further determined by the precautions taken by the individuals in contact (washing hands, wearing masks, distance keeping) resulting in an effective, time-dependent reproductive number $R_{eff}(t)$. We derive the relevant time-dependence of $R_{eff}(t)$ by applying a Kalman filter [25,26] to the piece-wise linearised time-series of daily confirmed B.1-C15324T cases. Assuming a multiplicative model, the time-dependence of transmission risk stemming from

social interaction $\alpha_{soc}(t)$, is obtained by point-wise division of the time-dependence of $R_{eff}(t)$ by $\alpha_{mob}(t)$ (Figure 2D).

SEIR-model. We used a compartmental two-arm susceptible-exposed-infected-recovered (SEIR) model [19,27,28] including sequenced and unsequencend/unreported cases that is outlined in Figure S1. The initial number of susceptibles was fixed to the relevant population. All other compartments were initialized as zeros, apart from a seed in E corresponding to the first reported cases. In summary, our model is based only on six free parameters: the reproductive number per tertile R_j (three parameters, range $[0,20]$), the initial number of exposed in a single tertile (range $[0,5]$), and the infectious times T_{infU} and T_{infP}(range $[1.5,12]$ days). We assumed a latency period T_{inc} of two days[29]. Model fitting and evaluation are described in the supplement. Results are shown as mean values over 50 bootstraps with bootstrap uncertainty bands. We compare effective reproductive numbers corresponding to the normalization of R by the effective mobility contribution ($\sum_k M_{jk}$):

$$R_{eff,j} = \frac{R_j}{\sum_k M_{jk}} \tag{1}$$

Significance levels are scored based on a comparison to 99 random partitions of the statistical blocks (see Table S1).

Scenario simulation. The impact of mobility relative to social interaction was analysed by recalculating the predicted epidemic trajectory under the constraint of constant ($\alpha_{mob}(t) = 1$) or fully restricted ($\alpha_{mob}(t) = 0$) mobility. These scenarios were compared to the baseline of ob-

served reduction in mobility.

Vaccination scenarios were simulated for 70-90% effective vaccines to prevent COVID-19[30–32], as well as 60% and 90% efficacies to prevent SARS-CoV-2 transmission. The effectively immunized fraction of the population was moved to the recovered compartment U_r. We accounted for a change in social interaction behaviour following vaccination by assigning a mean social interaction score of the vaccinated and not-vaccinated population amongst the initial susceptibles ($\alpha_{soc,vacc}(t) = 0.75$, $\alpha_{soc,novacc}(t) = 0.5$). Full mobility was modelled ($\alpha_{mob}(t) = 1$) after a single exposed individual was introduced into each of the tertiles. All scenarios were compared to the no-vaccine case (V0): vaccination of a fixed population fraction at random (V1); vaccination of different population groups (V2: exclusively from T1 median income, V3: exclusively from T3 seniority, V4: 50% from T1 median income, 50% from T3 seniority). In addition to case numbers, the time to reach 50% of ICU capacity was determined as quantifiable endpoint indicating healthcare system burden. In case of not-random vaccination, we adjusted the relevant fraction of ICU cases based on the represented proportions of senior population fractions within all susceptibles.

Results

SARS-CoV-2 spread and clustering. We observed 29 viral lineages in Basel-City (Figure S10), with 247 genomes (60.0%) belonging to the B.1-C15324T variant[23] (Figure 1A, B). Applying a genetic divergence threshold, a total of 128 phylogenetic clusters were determined across all samples, of which 70 belonged to lineage B.1-C15324T (Figure 1C). Mapping phylogenetic clus-

ters onto tertiles of socioeconomic and demographic determinants, we found that for median income, T1 contained the most and T3 the least clusters (Figure 1C). Most within- and among-tertile transmission clusters were spread randomly, except for significant within-tertile transmissions among high median income households (T3) (Figure 1D). Further, we observed that SARS-CoV-2 isolates were more likely to belong to the same phylogenetic cluster for blocks with either the highest living space per person, lowest share of 1-person households, or highest seniority (Figure 1E). This is also true among people living in the more affluent quarters Riehen, Bruderholz, Am Ring, and Iselin, who transmitted the virus in their social networks either in the same quarter or in the same socioeconomic rank (Figure S9). By contrast, positive cases that belong to lower socioeconomic/demographic tertiles (either lower income, less living space, or younger age) are less likely to be members of the same tight phylogenetic cluster, indicating cryptic transmission predominates among lower socioeconomic and younger demographic groups.

Spatio-temporal variation of mobility and social interaction patterns. Figure 2A and Figures S2-S4 show Basel-City's partition and the corresponding mobility graph. Importantly, the statistical blocks per tertile do not form a single, geographically connected, entity. We observe that mobility varies by transport modality and tertile (Figure 2A inset). For example, for low-median income (T1) the share of private motorized traffic and mobility is more pronounced than in the tertiles of higher median income (T2 and T3).

Figure 2B shows for each partition the summed edge weights of the mobility graph accounting

for the mobility contribution to the final effective reproductive number. We observe that low and median income populations are more mobile than their wealthier counterparts. Moreover, there is little mobility within areas with a low share of 1-person households, a result of the predominantly peripheral location of the relevant statistical blocks (see Figure S6). For living space per person or percentage of senior citizens, mobility was comparable between tertiles with a trend towards higher mobility within the younger population groups.

Dynamic changes in mobility were assessed by agglomerating normalized traffic counts for public and private transport modalities (Figure 2C). There was a clear drop in mobility for both public and private transport modes around the onset of the national lockdown date (12th March 2020). The decrease was more pronounced for public transport, resulting in a weighted average mobility drop of approximately 50% (Figure 2D). Figure 2D also shows the dynamic change in social interaction contribution to B.1-C15324T case numbers. Despite noticeable fluctuation, social interaction contribution decreased on average over time. This data also reflects variation in case reporting which affected the estimated effective reproductive number. Importantly, since the B.1-C15324T variant was eventually eradicated despite non-zero mobility, a final social interaction contribution of zero was expected.

Spatio-temporal spread of the epidemic and its socioeconomic determinants. Unreported cases appeared to be a driving force of the transmission (88% for the sequenced B.1-C15324T variant). Figures 3A-C (Figures S5-S7A-C) show the SEIR-model fit to data for each median income tertile. The corresponding dynamic change of the effective reproductive number (R_{eff}) is

given in Figures 3D-F (Figures S5-S7D-F). Independent of the underlying partition, the model provided adequate fits (RMSE < 4.5) and we observed a drop in R_{eff} following the dynamic changes in mobility and social interaction. Importantly, there was a significant difference (2% achieved significance) in R_{eff} between statistical blocks of the highest and lowest median income. For all socioeconomic partitions these differences are summarized as histograms in Figures 3G-J. Here, we found that blocks with higher living space per person (1% achieved significance), or higher median income (2% achieved significance), or or lower share of 1-person households (2% achieved significance) had a significantly lower R_{eff} (< 1.7) compared to the maximum R_{eff} observed in the relevant partition (Table S1). A partitioning based on the share of senior residents did not result in significant differences in R_{eff} (45% achieved significance). Differences in R_{eff} are due to two factors: the effective mobility contribution (Figure 2B), and the modelled reproductive number (R, eq.(8)). In particular, the tertile with the lowest share of 1-person households (T1) showed less mobility compared to T2 and T3, leading to significant differences in R_{eff}. By contrast, mobility in the T1 and T3 tertiles of living space per person were similar (Figure 2B), yet differences in R_{eff} were significant, indicating that the transmission was not dominated by mobility alone.

Impact of mobility changes and modelling of vaccination scenarios. We simulated the developments of the first wave of the epidemic under the assumption of different mobility scenarios and modelled two future vaccination strategies. Figure 4A (Figures S5-S7G) displays the results for mobility scenarios as observed with up to 50% mobility reduction (scenario MO),

100% mobility (scenario M1), and no mobility (scenario M2). Peak case numbers (April 12th) would have been approximately three times higher in the case of no reduction in mobility (M1). However, the decrease in peak case numbers assuming zero mobility (M2) as modelled, was not as pronounced. Mobility reduction hence played a vital role for the containment of SARS-CoV-2.

Figure 4B (Figures S5-S7H) shows the results for an outbreak scenario (denoted as V1) in which a specific fraction of the population (33% or 66%) received a vaccine that provided either 60% or 90% vaccine transmission reduction, with 90% severe COVID-19 case prevention. As expected, we observe that higher vaccine efficacy or higher population fraction vaccinated reduces the slope and plateau of the epidemic curve. Scenarios for less effective vaccines are shown in Figure S11. Effectiveness to prevent severe COVID-19 solely affects the rate of severe cases and hence ICU occupancy and the time point of reaching 50% ICU occupancy (Figure 4B vs. Figure S11E). It should be noted, that vaccination of the population at random is an artificial scenario, applied here only to demonstrate the impact of vaccination efficacy relative to the population fraction vaccinated. This scenario serves as a baseline comparison for two more realistic vaccination strategies given in Figure 4C. Here, as a proof of concept, vaccines that provide 90% transmission reduction, and 90% reduction of severe cases are delivered to 23% of the population are presented. For scenario V2, vaccination is prioritized in what we determined as determinants of SARS-CoV-2 transmission – individuals with low income, who have fewer options to socially distance (e.g. by working from home) and hence were more likely to be exposed to and/or transmit the virus (reflected by a higher R_{eff}). With this strategy, the slope

of the epidemic curve would be reduced compared to randomly vaccinating the same number of subjects from the whole population. Figure 4D describes the corresponding development of ICU occupancy for the scenarios modelled, revealing that scenario V2 leads to a delay of approximately 11 days to reach the 50% ICU capacity mark as compared to scenario V1. In scenario V3, resembling the approach by several countries, priority was given to the population group with the highest share of senior residents, which had lower mobility than the rest of the population (Figure 2B) but constitutes 60% of ICU cases. We observe that scenario V3 resulted in a marginally steeper epidemic curve (Figure 4C) and would yield 50% ICU capacity at a similar time as a random vaccination strategy (Figure 4D). However, the total number of cases at this time would be approximately double in scenario V3 compared to V1 (Figure 4C), whereas the overall peak ICU occupancy would be lowest in V3 (Figure 4D). These simulations suggest that - in case of vaccines reducing SARS-CoV-2 transmission - vaccination of population groups driving transmission are most effective in reducing the slope of the epidemic curve, whereas vaccination of high-risk groups reduces healthcare system burden in the long term. The presented effects strongly depend on the specific vaccine characteristics and population fraction vaccinated. In Figure 4E, F we finally show the potential effects of a mixed vaccination scenario giving equal priority to senior and highly mobile population groups as a representative example. This mixed mode provides a possible compromise with lower case numbers compared to V1, but also delayed and reduced ICU occupancy relative to V1.

Discussion

This analysis evaluates complementary aspects of the spread of SARS-CoV-2 within a medium-sized European urban area, including local transmission analysed by phylogenetic tree inference and clustering, and the overall spread described by a compartmental SEIR-model enabling simulation of vaccination strategies. The main strength of this study lies in the high degree of diverse and detailed data included as well as the complementary models and analyses.

Patterns of SARS-CoV-2 transmission have previously been discussed from different angles either via network and transmission modelling [33], by statistical evaluation [34], or by phylogenetic clustering based on genomic sequencing data[35]. Independent of model choice, the importance of socioeconomic factors has been suggested previously[18,33]. However, analyses focused on metropolitan areas only may be biased towards their underlying socioeconomic and demographic characteristics making it important to also quantify SARS-CoV-2 transmission in other urban areas, such as in a city context, as well as in countries across all continents and stages of economic potential. Modelling studies provide the foundation of scenario predictions and estimation of effective reproductive numbers but require balancing the trade-off between detail described and the number of data points available. Accordingly, many published models rely on publicly available case numbers without being able to directly relate socioeconomic parameters and specific geographic locations per case, and/or are only performed for large metropolitan areas[19]. This has led to biases in evaluations and the neglect of a considerable population share living outside major cities (53%/41% in Europe/worldwide)[36,37]. In this study we chose Basel-City as a case study of a European urban area. Despite the strong economic

status of Switzerland, implying potentially less extreme socioeconomic gradients than other countries, we demonstrate that socioeconomic background impacts the probability to acquire and transmit SARS-CoV-2. It would be expected that in cities containing more pronounced socioeconomic gradients, these disparities would emphasize transmission patterns of SARS-CoV-2 between and within socioeconomic groups.

The success of our analyses is based on the optimized choice of evaluation and model, high data-density and -quality, rather than large absolute case numbers. It is difficult to distinguish the spread of competing viral variants within the same population and to account for new introductions in a model since classical ODE or agent-based models are relying on the assumption of uninterrupted transmission chains. Sequencing information is essential to reliably inform on such transmission patterns. Yet given the cost of such analyses, WGS covering entire epidemic waves is often unfeasible. We included 81% of all reported cases in the study time frame and geographic area, which allowed us to restrict our analysis to a subset of 247 phylogenetically related cases consisting of a single SARS-CoV-2 variant. Given the ~200k inhabitants of Basel-City, this implies one of the largest per-capita sequencing densities of reported studies to date. We deliberately choose a simple way to incorporate socioeconomic, demographic and mobility information into our compartmental model since more complex network approaches would be unfeasible for limited case numbers. The use of mobility and socioeconomic data in our models is unique since we include regularly collected data analysed by the statistical office of Basel-City, providing a high spatial and temporal resolution network of the inner city mobility patterns. In contrast to mobile phone data[16,19,33,38] our data is not subjected to privacy

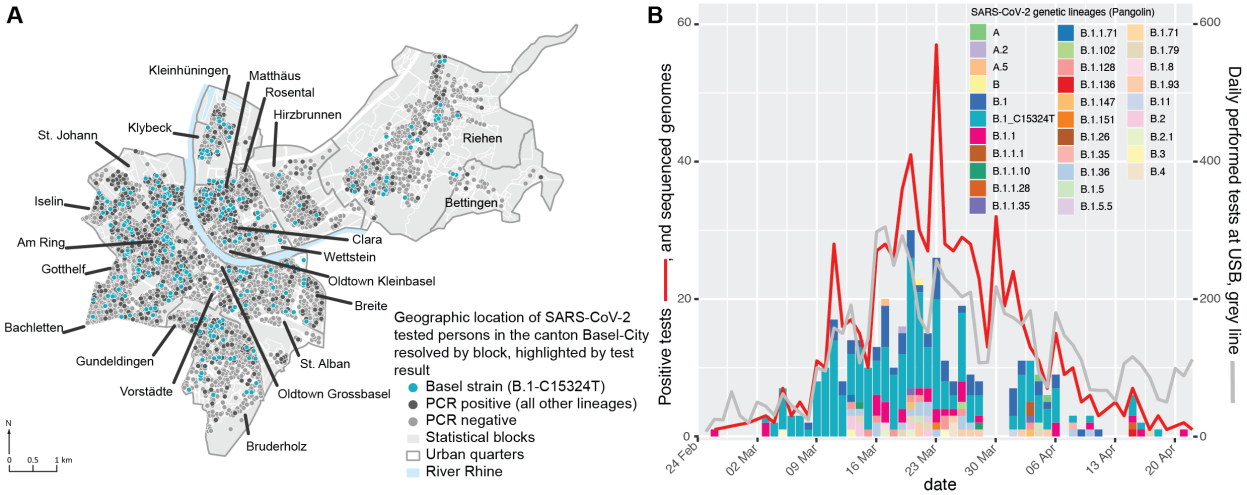
legislation and is hence expected to be more readily available for other medium-sized urban areas around the world, making our analysis transferable. We do not hold information on the duration and specific location of individuals, but a continuum estimate of population mixing that aligns well with the concept of a compartmental ODE model. Mobility and the reduction thereof have been suggested as a proxy to evaluate the reduction of the spread of SARS-CoV-2 [6,38,39]; however, there has also been a change in hygiene practices and social interaction behaviour. In our SEIR-model we separate these two contributions allowing for an easier translation of our model for scenario building. We further addressed unreported cases which were the driving force of infection outside the observed clusters. Our 77% estimate of unreported cases overall (not limited to B.1-C15324T) falls within the range of previous reports within Europe[8,40]. The SEIR-model evaluates general trends of transmission, such as effective reproductive numbers, and vaccine scenario building. To complement this, we employ phylogenetic analysis to identify transmission clusters. This comprehensive evaluation showed that socioeconomic brackets characterized by low median income and smaller living space per person, were associated with significantly larger effective reproductive numbers. In line with previous results[33], we suggest that population groups from a weaker socioeconomic background are more mobile and at higher risk for SARS-CoV-2 infection/transmission originating from multiple sources via cryptic transmission. This aligns with the possibility that low socioeconomic status may relate to jobs requiring higher personal contact, and unavoidable mobility[41], which has been shown to increase the risk of infection by 76%[42]. By contrast, phylogenetic clusters were predominantly discovered within higher socioeconomic, or more senior groups, implying a spread within the

same social network. It is likely that those individuals are retired, or have had the ability to work from home, a pattern that has been observed also in other cities[43]. Effective contact tracing and testing strategies could be most efficient for these groups, which were not driving SARS-CoV-2 transmission.

These results should be accounted for during vaccine prioritization depending on the relevant vaccine characteristics and ICU capacity available. Our simulation framework provides flexibility to model various scenarios and vaccine efficacies but did not account for fatalities due to limited case numbers during the studied period in Basel-City. In case of a combined effect of vaccines to both prevent SARS-CoV-2 transmission and protect against COVID-19[44–46], vaccination of individuals driving transmission in addition to the protection of high-risk groups would be ideal to arrive at a combined concept of protection from and containment of SARS-CoV-2. Vaccinating high-risk groups reduces the number of hospitalized and ICU patients in the short term, the spread of the pandemic would however, be more effectively contained by vaccinating the transmission drivers. By restricting vaccination to only risk groups, a larger fraction of the general population will be exposed to SARS-CoV-2 implying that contact and travel restrictions would remain necessary to contain transmission. Such measures come at great economic cost. Based on our results it would be recommended to follow a combined strategy to employ extensive testing where transmission chains are traceable, e.g. among less mobile population groups, and to combine this with a vaccination strategy aiming to prevent cryptic transmission.

In conclusion, high-resolution city-level epidemiological studies are essential for understanding factors affecting pandemic transmission chains and thereby supporting tailored public health

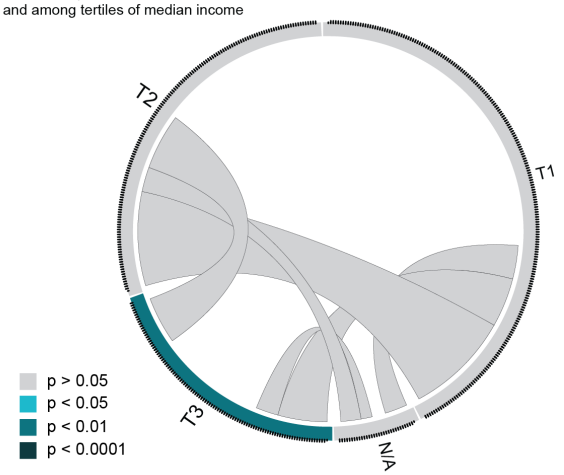
information campaigns and vaccination distribution strategies at the municipal level. We here provided an example of such an analysis within a representative medium-sized European city at the core of the Greater Basel area and part of the Upper Rhine Region Metropolitan Economy, which suggests that the findings and modelling approaches presented may be readily translated to other such areas.



C Table.Number of phylogentic clusters in tertiles of median income within lineage B.1-C15324T (N = 70) and among all intodruced lineages (N = 128). Please note that clusters can be shared among tertiles and therefore to not add up to N.

	N clusters in B.1-C15324T	N clusters in Basel-City (all strains)	N infected persons in B.1-C15324T	N infected persons Basel-City (all strains)
T1	51	82	94	141
T2	38	62	63	96
T3	22	36	36	56
w/o adress	13	15	15	35
w/o cluster			39	83

D Results for permutation test of shared clusters within and among tertiles of median income



E Table. P-values as determined by permutation analysis for within-tertile shared clusters within B.1-C15324T by different socioeconomic factors. All among-tertile transmissions are non-significant indicating that no significant amounf of clusters are shared between different socioeconomic groups.

tertile	1	2	3	N/A
Living space per person	0.240	0.300	0.009	0.999
1-person households	0.001	0.287	0.295	0.999
Median income	0.222	0.458	0.002	0.458
Seniority	0.223	0.068	0.012	1.000

Figure 1: SARS-CoV-2 transmission in and among socioeconomic and demographic groups during the first COVID-19 wave in Basel-City. A) Spatial positive/negative case distribution throughout the city with the most dominant SARS-CoV-2 variant (B.1-C15324T), the focus of this study, highlighted in turquoise. B) Epidemiological curve for Basel-City and distribution of phylogenetic lineages (pangolin nomenclature) from 25th of February to 22nd of April 2020. C) Summary for inferred phylogenetic clusters within (i) all lineages and (ii) the major variant B.1-C15324T in tertiles of median income. High number of infected people within a tertile with a low number of clusters indicates presence of large transmission clusters whereas large number of clusters and low number of people infected within a tertile indicates random infections and cryptic transmission. D) Visualisation of a significance test for transmission within (indicated on circle edges) and among (intra circle connections) tertiles of median income. E) Results of a significance test for transmission between tertiles of different socioeconomic factors. T1: low, T2: intermediate, T3: high, N/A: no available data or censored for privacy reasons.

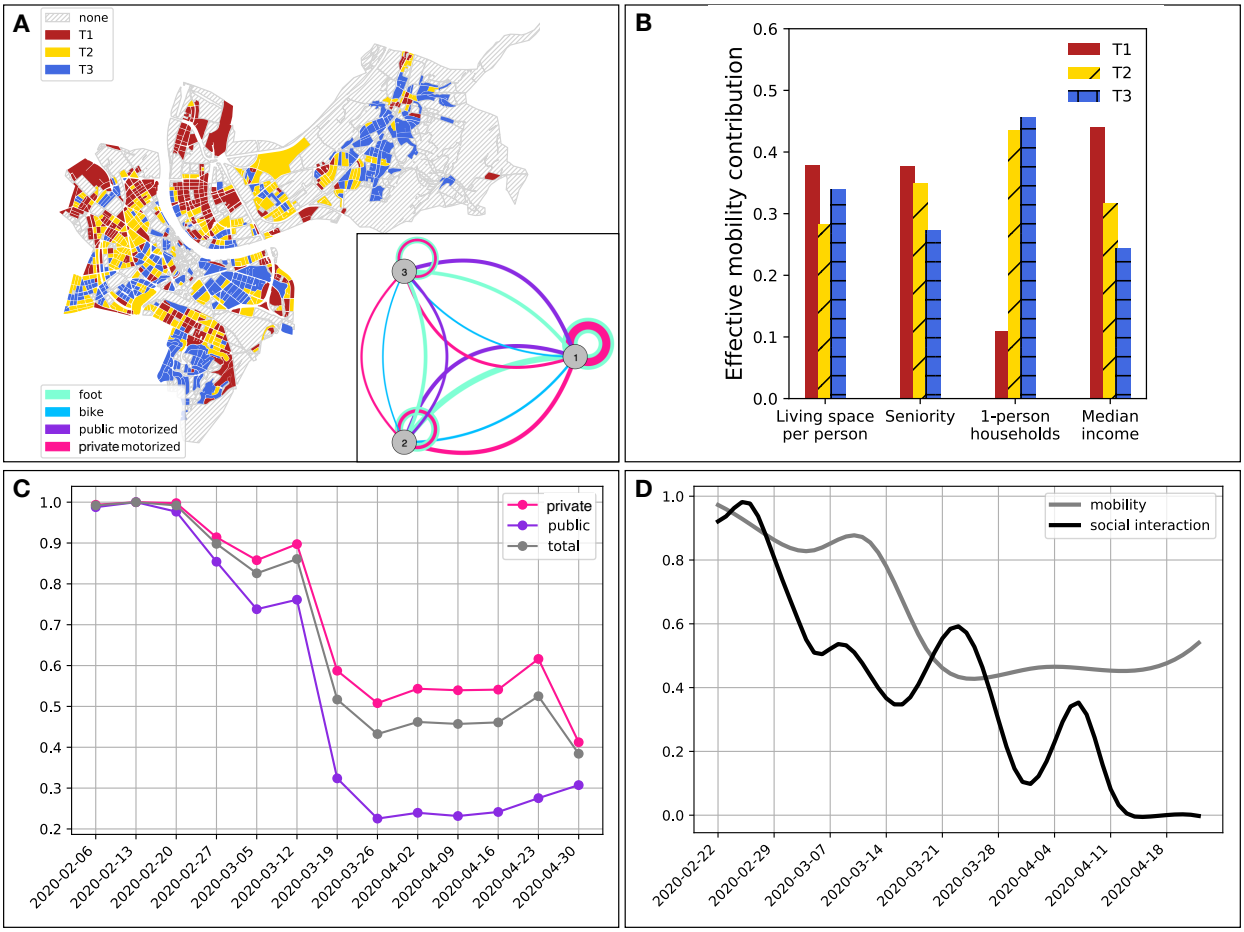


Figure 2: Spatio-temporal variation of mobility patterns within the Canton of Basel-City. A) Basel-City and its delineation with respect to statistical blocks colored according to the partition into tertiles T1, T2, and T3 of increasing median income. Inset: resulting mobility-graph, with nodes representing tertiles and edge-widths representing the strength of effective connectedness through mobility by means of various modes of transport, as computed from the traffic-model provided by the traffic department of the Canton of Basel-City. B) Relative mean contribution of mobility to a socioeconomic tertile's effective reproductive number associated with the major variant B.1-C15324T. C) Normalized temporal development of private and public transport as well as their weighted sum during the first wave of the pandemic in Basel-City. D) Smoothed relative temporal development of social interaction and mobility contribution to the effective reproductive number associated with the major variant B.1-C15324T.

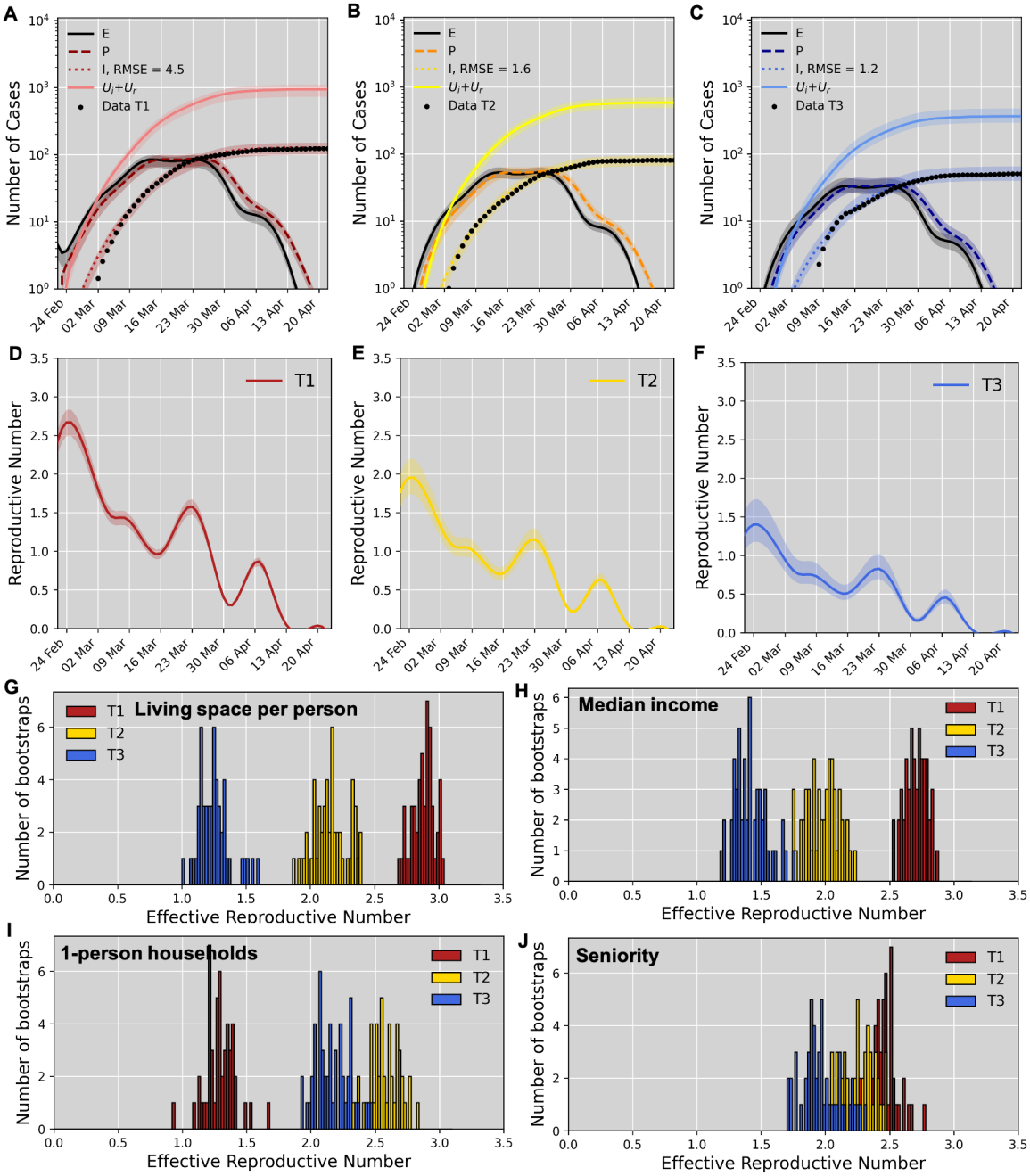


Figure 3: Model fit to the case number time-series. A-C) Fit results for a partition based on median income. Data points are shown together with model predictions based on undisturbed data (solid lines), and fifty bootstraps from disturbed data (bands) for the different tertiles T1 (low, A), T2 (intermediate, B) and T3 (high median income, C). D-F) The dynamic variation of the effective reproductive number for each of the tertiles shown in A-C. G-J) Histograms over all bootstraps for the effective, pre-lockdown reproductive number for each socioeconomic partition. Results are shown for partitions based on living space per person (G), median income (H), share of 1-person households (I), and share of senior residents (J).

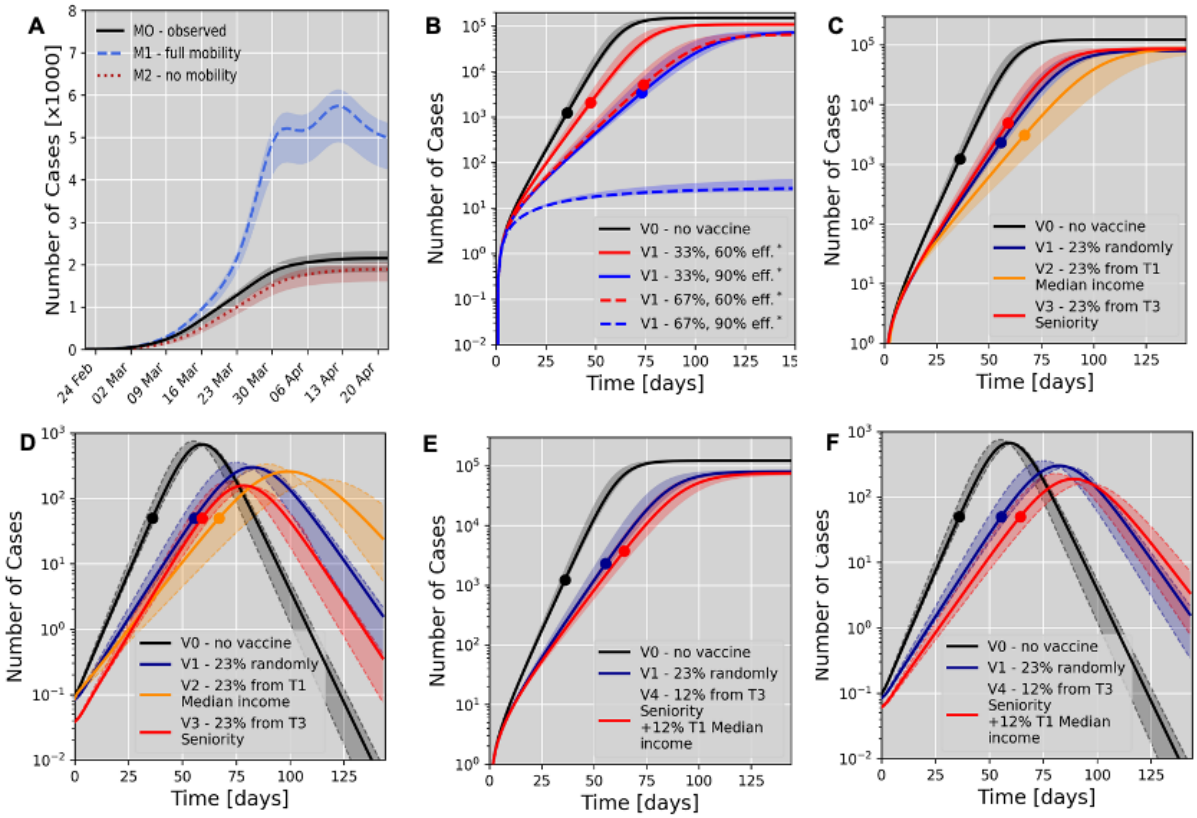


Figure 4: Scenario simulations for a partition based on median income. A) Influence of the mobility pattern on the total number of infected cases during the first wave (sum of reported and unreported cases) modelling either no change in mobility (no lockdown scenario, M1), or full shut-down of all inner city mobility (M2). For comparison the observed scenario (MO) is shown. B) Simulation of future vaccination effects if a specific percentage of all citizens was randomly selected for vaccination at given efficacy (V1). We compare this to the scenario of no vaccine (V0). * Here vaccination of different population fractions at either 60% or 90% efficacy to prevent SARS-CoV-2 transmission were modelled. C) Simulation of future vaccination effects based on a partition according to median income. Scenario V2 models vaccination of 23% of all citizens selected from the tertile with the lowest median income (T1). Scenario V3 models vaccination of 23% of all citizens selected from the tertile with the highest share of senior residents (T3). In C and E we model 90% vaccine efficacy (transmission and severe COVID-19) and compare with scenarios V0 and V1. Dots indicate the time of reaching a 50% ICU occupancy. D) Temporal evolution of ICU occupancy for the scenarios modelled in C. E) Simulation of a mixed vaccination strategy giving equal priority to senior citizens and mobile population groups. F) Temporal evolution of ICU occupancy for the scenarios modelled in E).

Acknowledgements We greatly appreciate the input and data received from Construction- and Traffic department Canton Basel-City, Baselland Transport AG, Basler Verkehrs-Betriebe, Autobus AG Liestal, SBB Federal Railways, Statistical Office of the Canton of Basel-City, and want to specifically thank Björn Lietzke, Lukas Mohler, and Madeleine Imhof (all Statistical Office of the Canton of Basel-City), from Construction- and Traffic Department of the Canton of Basel-City Michael Redle and Kathrin Grotrian, Matthias Hofmann (Basler Verkehrs-Betrieb), Roman Stingelin (Autobus AG), Nadine Ruch (SBB AG) and Stefan Burtschi (Baselland Transport AG) for their support. We thank Christine Kiessling, Magdalena Schneider, Elisabeth Schultheiss, Clarisse Straub, and Rosa-Maria Vesco (University Hospital Basel) for excellent technical assistance with next generation sequencing. Computations were performed at sciCORE (http://scicore.unibas.ch) scientific computing facility at the University of Basel. Data exchange was organized via the BioMedIT node between the University of Basel and ETH Zurich, Department of Biosystem Science and Engineering. We thank all authors, who have shared their genomic data on GISAID, especially the Stadler Lab from ETH Zurich for sharing Swiss sequences. A full table (csv) outlining the originating and submitting labs is included as a supplementary file. This study was supported by the Alfried Krupp Prize for Young University Teachers of the Alfried Krupp von Bohlen und Halbach-Stiftung (KB). We finally thank Dr. A. Jermy (Geminate Science Consulting) for his critical review of the manuscript.

Author contributions AE, HHH, KB devised the project. SCB and JK developed and performed the mathematical modelling. MSt performed and interpreted phylogenetic analyses. SCB, JK and MSt led the writing and revising of the report. AM provided the genome assembly pipeline. TCR and HSS prepared viral RNA for sequencing. MyB and RSS provided geographical expertise. RSS provided detailed background on the representative status of Basel-City. KKS collected clinical data. KR, DAT, AG, AKS,

MSch analysed serology samples. DC and OD provided serology samples from Viollier AG. KL provided virological expertise. AB provided serology samples from the blood transfusion service. JB, STS and SF provided public health and epidemiological expertise. HP, MSi, CHN, RB, MB provided clinical expertise and valuable discussion on the results. NR, UH, JB provided epidemiological expertise. All author reviewed and edited the manuscript.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Adrian Egli, adrian.egli@usb.ch, Clinical Bacteriology and Mycology, University Hospital Basel, Petersgraben, CH-4031 Basel, Switzerland, Phone: +41 61 556 57 49.

Data Availability The SEIR-model code used for this submission will be available on https://github.com/BorgwardtLab. Code that was used for phylogenetic inference and calculation of significance of clusters in specified groups is available at https://github.com/appliedmicrobiologyresearch. SARS-CoV-2 whole genomes from Basel-City are available at gisaid.com and at European Nucleotide Archive (ENA) under accession number PRJEB39887.

Ethical statement Ethical approval was given by the local ethical committee *Ethik Kommission Nordwest und Zentralschweiz* (EKNZ No. 2020-00769, to be found at https://ongoingprojects. swissethics.ch) and the project was registered at clinicaltrial.gov under NCT04351503.

References

1. Yuan, F., Wang, L., Fang, Y. & Wang, L. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. *Transboundary and emerging diseases* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33207070.

2. Nabil, B., Sabrina, B. & Abdelhakim, B. Transmission route and introduction of pandemic SARS-CoV-2 between China, Italy, and Spain. *Journal of medical virology* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32697346 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7404595.

3. Hadfield, J. *et al.* NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018). URL https://pubmed.ncbi.nlm.nih.gov/29790939/.

4. Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science (New York, N.Y.)* **370**, 564–570 (2020). URL https://science.sciencemag.org/content/370/6516/564.full.

5. Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Science translational medicine* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33229462.

6. Candido, D. S. *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science (New York, N.Y.)* **369**, 1255–1260 (2020).

URL http://www.ncbi.nlm.nih.gov/pubmed/32703910 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7402630.

7. Salje, H. *et al.* Estimating the burden of SARS-CoV-2 in France. *Science (New York, N.Y.)* **369**, 208–211 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32404476 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7223792.

8. Gudbjartsson, D. F. *et al.* Spread of SARS-CoV-2 in the Icelandic Population. *New England Journal of Medicine* **382**, 2302–2315 (2020).

9. Post, L. A. *et al.* A SARS-CoV-2 Surveillance System in Sub-Saharan Africa: Modeling Study for Persistence and Transmission to Inform Policy. *Journal of medical Internet research* **22**, e24248 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33211026 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7683024.

10. Puenpa, J. *et al.* Molecular epidemiology of the first wave of severe acute respiratory syndrome coronavirus 2 infection in Thailand in 2020. *Scientific Reports* **10** (2020). URL https://pubmed.ncbi.nlm.nih.gov/33024144/.

11. Du, P. *et al.* Genomic surveillance of COVID-19 cases in Beijing. *Nature communications* **11**, 5503 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33127911 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7603498.

12. Lemieux, J. *et al.* Phylogenetic analysis of SARS-CoV-2 in the Boston area highlights the role of recurrent importation and superspreading events. *medRxiv : the preprint server for health sciences* **12**, 15

(2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32869040 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7457619.

13. Long, S. W. *et al.* Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. *medRxiv* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33024977 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7536878.

14. Maurano, M. T. *et al.* Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City region. *Genome research* **30**, 1781–1788 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33093069.

15. Bushman, D. *et al.* Detection and Genetic Characterization of Community-Based SARS-CoV-2 Infections - New York City, March 2020. *MMWR. Morbidity and mortality weekly report* **69**, 918–922 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32678072 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7366849.

16. Kissler, S. M. *et al.* Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nature communications* **11**, 4674 (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/32938924 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7494926.

17. EUROSTAT. Population Data Collection for European Local Administrative Units from 1960 onward (2011). URL

https://ec.europa.eu/eurostat/web/nuts/local-administrative-units.

18. De Ridder, D. *et al.* Socioeconomically disadvantaged neighborhoods face increased persistence of sars-cov-2 clusters. *Frontiers in Public Health* **8**, 1091 (2021). URL https://www.frontiersin.org/article/10.3389/fpubh.2020.626090.

19. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* (2020). URL http://www.nature.com/articles/s41586-020-2923-3.

20. Koo, J. R. *et al.* Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *The Lancet Infectious Diseases* **20**, 678–688 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S1473309920301626.

21. Mansilla Domínguez, J. M. *et al.* Risk Perception of COVID-19 Community Transmission among the Spanish Population. *International journal of environmental research and public health* **17** (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33276532.

22. Ragonnet-Cronin, M. *et al.* Automated analysis of phylogenetic clusters. *BMC Bioinformatics* **14**, 317 (2013).

23. Stange, M. *et al.* SARS-CoV-2 outbreak in a tri-national urban area is dominated by a B.1 lineage variant linked to a mass gathering event. *medRxiv; accepted for publication in PLOS Pathogens* 2020.09.01.20186155 (2021). URL https://doi.org/10.1101/2020.09.01.20186155.

24. Bau- und Verkehrsdepartement Basel-Stadt Mobilität / Mobilitätsstrategie. Gesamtverkehrsmodell der Region Basel, Basismodell: “Ist-Zustand 2016” (2020).

25. Kalman, R. E. *et al.* Contributions to the theory of optimal control. *Bol. soc. mat. mexicana* **5**, 102–119 (1960).

26. Welch, G., Bishop, G. *et al.* An introduction to the kalman filter (1995).

27. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020). URL https://www.sciencemag.org/lookup/doi/10.1126/science.aba9757.

28. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020). URL https://www.sciencemag.org/lookup/doi/10.1126/science.abb3221.

29. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**, 672–675 (2020). URL https://www.nature.com/articles/s41591-020-0869-5.

30. AstraZeneca. AZD1222 vaccine met primary efficacy endpoint in preventing COVID-19 (2020).

31. Mahase, E. Covid-19: Moderna vaccine is nearly 95% effective, trial involving high risk and elderly people shows. *BMJ: British Medical Journal (Online)* **371** (2020).

32. Dagan, N. *et al.* Bnt162b2 mrna covid-19 vaccine in a nation-wide mass vaccination setting. *New England Journal of Medicine* **0**, null (0). URL https://doi.org/10.1056/NEJMoa2101765. https://doi.org/10.1056/NEJMoa2101765.

33. Jay, J. *et al.* Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. *Nature human behaviour* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33144713.

34. De Ridder, D. *et al.* Socioeconomically Disadvantaged Neighborhoods Face Increased Persistence of SARS-CoV-2 Clusters. *Frontiers in Public Health* **8** (2021). URL https://www.frontiersin.org/articles/10.3389/fpubh.2020.626090/full.

35. Bluhm, A. *et al.* SARS-CoV-2 transmission routes from genetic data: A Danish case study. *PLOS ONE* **15**, e0241405 (2020). URL https://dx.plos.org/10.1371/journal.pone.0241405.

36. United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision* (United Nations, New York, 2019). URL https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

37. European Commission. GDP per capita, consumption per capita and price level indices (2021). URL https://ec.europa.eu/eurostat/statistics-explained/index.php/GDP

38. Rader, B. *et al.* Crowding and the shape of COVID-19 epidemics. *Nature Medicine* **26**, 1829–1834 (2020). URL nature.com/articles/s41591-020-1104-0#citeas.

39. Douglas, J. *et al.* Phylodynamics reveals the role of human travel and contact tracing in controlling COVID-19 in four island nations. *medRxiv* (2020).

40. Burki, T. Mass testing for COVID-19. *The Lancet Microbe* **1**, e317 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S2666524720302056.

41. Reeves, R. & Rothwell, J. Class and covid: How the less affluent face double risks. brookings (march 27) (2020).

42. Rodríguez-Barranco, M. *et al.* The spread of SARS-CoV-2 in Spain: Hygiene habits, sociodemographic profile, mobility patterns and comorbidities. *Environmental Research* **192**, 110223 (2021).

43. Wilson, C. These graphs show how covid-19 is ravaging new york city's low-income neighborhoods. *Time* (2020).

44. Swan, D. A. *et al.* Vaccines that prevent sars-cov-2 transmission may prevent or dampen a spring wave of covid-19 cases and deaths in 2021. *medRxiv* (2020). URL https://www.medrxiv.org/content/early/2020/12/14/2020.12.13.20248120. https://www.medrxiv.org/content/early/2020/12/14/2020.12.13.20248120.full

45. Moghadas, S. M. *et al.* The impact of vaccination on COVID-19 outbreaks in the United States. *medRxiv : the preprint server for health sciences* (2020). URL http://www.ncbi.nlm.nih.gov/pubmed/33269359 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7709178.

46. Anderson, R. M., Vegvari, C., Truscott, J. & Collyer, B. S. Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination. *The Lancet* **396**, 1614–1616 (2020). URL https://linkinghub.elsevier.com/retrieve/pii/S0140673620323187.