1. RESEARCH STRATEGY

A. Significance:

Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) emerged in China threatening public health and global economies (*1, 2*). Like most other emerging pathogens (*3*), it originated in animal reservoir hosts, initially thought to be carnivores (*4*), and later shown by our group to be bats (*5*). Bats harbor a high diversity of β-coronaviruses, including those related to Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV) (*6-9*) and the newly emerged Swine Acute Diarrhea Syndrome coronavirus (SADS-CoV) (*10*), and may be the progenitor hosts of all *Coronaviridae* (*5, 11-15*). SARS-CoV uses the angiotensin-converting enzyme 2 (ACE2) receptor to gain entry to human cells (*16*). In 2012, we isolated and characterized two bat SARS-related coronaviruses (SARSr-CoVs) in China that use the ACE2 receptor and are closely related to SARS-CoV (*17*). Since then, under an R01 awarded in 2014, we have discovered >50 bat SARSr-CoVs in southern China. Some of these strains can bind to and infect human cells, cause SARS-like clinical signs in a humanized mouse model, and evade therapeutic and vaccine candidates against SARS-CoV (*18*). The *Rhinolophus* spp. bat hosts of these viruses are abundant across southern China, where hunting and consumption of wildlife is common and human population growth high, and where we have now identified serological evidence of exposure to SARSr-CoVs and other bat CoVs (*19*). Thus, there is significant potential for future spillover of SARSr-CoVs, and of their subsequent spread. *Yet salient questions remain: Are there specific bat communities and sites that harbor CoV strains with higher risk for bat-to-human spillover? Which human behaviors drive risk of bat SARSr-CoV exposure that could lead to infection? Does human exposure to these viruses cause SARS-like or other illness? Can we characterize viral strain diversity, bat traits and human behaviors to assess risk of potential future CoV spillover?* **This R01 renewal** proposal aims to address these critical issues by conducting: **1)** focused sampling of bats in southern China to **identify viral strains with high predicted risk of spillover**; **2)** community-based, and clinic-based syndromic, sampling of people to **identify spillover, and assess behavioral risk factors and evidence of illness**; and **3)** conduct *in vitro* and *in vivo* viral characterization and analyze epidemiological data **to identify hotspots of future CoV spillover risk**.

B. Progress report: R01 AI110964, Daszak PI, Project Period: 06/01/2014 – 05/31/2019

The aims of our previous R01 were to: 1) Assess bat SARSr-CoV spillover potential at high risk human-wildlife interfaces, e.g. the wildlife trade, as reported for the 2003 outbreak (*4, 20*); 2) Analyze how viral diversity and phylogeny relates to host range and risk of emergence; and 3) Use binding assays, cell culture and mouse models to test the propensity of different SARSr-CoVs to infect humans. We made significant discoveries leading to **18 published peer-reviewed papers (*18, 19, 21-33*), including two papers in *Nature* (*10, 34*), and a review in *Cell* (*35*)** (see Progress Report Publication List). These findings include:

Diversity and distribution of bat β- and SARSr-CoVs in Southern China.

We sampled and PCR-screened >16,000 individual bats from 6 families (16 genera) in southern China, finding 9 species positive (5,730 individuals screened) for SARSr-CoVs **(Table 1, Fig. 1)**. We identified 178 novel β-CoVs, of which 172 were novel (52 novel SARSr-CoVs). This includes members of a new β-CoV clade, "lineage E" (*26*), and diverse HKU3-related CoVs (179 sequences) within a 'sister' clade to the SARS-CoV lineage.

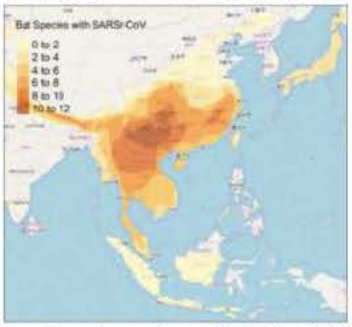

Fig. 1 (left): Bat sampling at 47 sites in China under our previous R01. Yellow = sampling effort, red = CoV +ve bats. **Fig. 2 (right)**: Map of bat species found positive for SARSr-CoVs in our previous R01, highlighting S. China (particularly Yunnan Province) as a center of diversity for SARSr-CoV reservoir host species.

We found 6.7% mean PCR prevalence of SARSr-CoVs across bat hosts, with a small number of *Rhinolophus* spp. horseshoe bats having significantly higher PCR prevalence than other species sampled **(Table 1)**. These bats are widely distributed, diverse, abundant, and roost and feed close to people and livestock, suggesting high potential for future SARSr-CoVs spillover. Distribution data for SARSr-CoV bat hosts suggest viral strain diversity is likely highest in southern China, particularly Yunnan Province **(Fig. 2)**.

Table 1: Species found PCR-positive for SARSr-CoVs in our R01, with sample sizes and prevalence estimates.

Bat Species	Individuals tested	# positive	SARSr-CoV mean prev.	SARSr-CoV prev. range
Rhinolophus sinicus	1,328	113	8.5%	7.1 – 10.1%
R. macrotis	70	3	4.3%	0.9 - 12%
R. ferrumequinum	406	12	3.0%	1.5 – 5.1%
R. spp.	331	10	3.0%	1.5 – 5.5%
R. affinis	792	7	0.9%	0.4 – 1.8%
R. pusillus	1,023	8	0.8%	0.3 – 1.5%
Aselliscus stoliczkanus	269	2	0.7%	0.1 – 2.7%
Hipposideros pratti	323	2	0.6%	0.1 – 2.2%
H. armiger	1,188	1	0.1%	0.0 – 0.5%

We identified one cave system (the "Jinning Cave") in Yunnan Province that harbors *Rhinolophus* spp. bats with diverse SARSr-CoVs, including some with S proteins able to use human ACE2 as entry receptors. Bats in this cave carried SARSr-CoVs with **all unique genetic elements of the SARS-CoV outbreak virus**, suggesting that this site may be a potential public health risk (*29*).

We used a novel phylogeographic analysis, Maximum Clade Credibility (MCC) tree, to reconstruct the geographic areas of evolutionary origin for β-CoVs that we sequenced. Results suggest that: 1) Guangdong Province is the ancestral center of diversity of β-CoVs (data not shown); 2) Guizhou is the likely origin of the HKU3-related clade (lineage 1); and 3) Guangdong and Guangxi are the likely ancestral origins of the SARS-CoV outbreak sequences (lineage 2) **(Fig. 3)**. Despite our intensive sampling at some sites, around half of the 20 *Rhinolophus* spp. we identified were captured at sample sizes below the minimum required to detect SARSr-CoVs at prevalences we found (n=110, power 80%), and 5 others were SARSr-CoV negative in our study. To estimate sampling gaps, we used a viral 'mark-recapture' approach we previously published (*36, 37*). Results suggest we are approaching saturation of CoV strain discovery at some sites, whereas other sites contain rich pools of SARSr-CoVs that remain undiscovered **(Fig. 4)**. **In the current proposal,** we have used these analyses to estimate geographic and species-specific sampling targets **to more effectively identify new strains and CoV lineages needed to support experimental infection studies and risk assessment.**

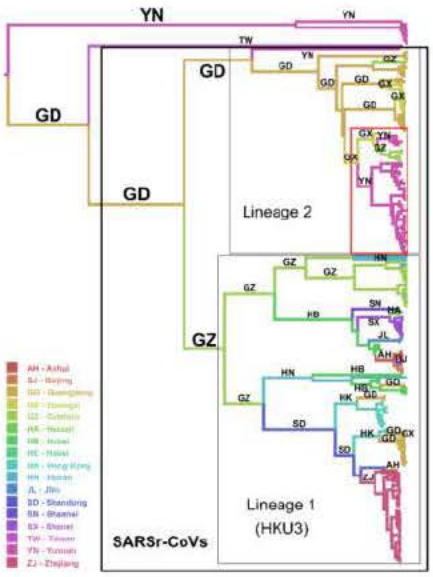

Fig.3 (left): MCC phylogeny of lineage B β-CoVs, including SARSr-CoVs (black box). Lineage 1 includes HKU3-related CoVs, lineage 2 includes SARS-CoV outbreak strains and close relatives (red box). Branches colored according to province of inferred ancestral origin (Guangdong GD, Yunnan YN, Guizhou, GZ).

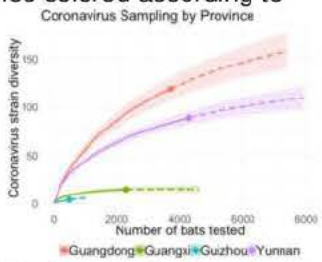

Fig. 4 (right): Estimates of SARSr-CoV strain diversity in the bats we sampled (strain defined as >10% sequence divergence in RdRp gene). GD and YN harbor highest CoV diversity, but discovery has not yet saturated. We estimate proposed additional sampling of 5,000 bats will identify >80% of remaining β-CoV strains in bat hosts from these regions.

In vitro & in vivo *characterization of SARSr-CoV potential for human infection*

We conducted *in vitro* and *in vivo* experiments to characterize the pathogenic potential of novel SARSr-CoVs. We isolated three SARSr-CoVs from bat feces: WIV1, WIV16 and Rs4874, with S protein sequences that diverged from SARS-CoV by 3% to 7% (*17, 22, 29*). We conducted full-length genome sequencing of 12 other novel SARSr-CoVs from the Jinning Cave, some highly similar to outbreak SARS-CoV in the most variable genes: N-terminal domain and receptor binding domain (RBD) of the S gene, ORF8 and ORF3 (*29*). Using our reverse genetics system, we constructed chimeric viruses with SARSr-CoV WIV1 backbone and the S gene of different variants, including WIV1-Rs4231S and WIV1-Rs7327S. All 3 SARSr-CoV isolates and the two chimeric viruses replicated efficiently in Vero E6 cells and in HeLa cells expressing hACE2, but not in HeLa cells that don't express ACE2 (*17, 22, 29*) **(Fig. 5a)**. In collaboration with Ralph Baric (UNC), we used the SARS-CoV reverse genetics system (*38*) to generate a chimeric virus with a mouse-adapted SARS-CoV

backbone expressing SHC014 S protein with 10% sequence divergence from SARS-CoV S. This chimera replicated in primary human airway epithelium, using the human ACE2 receptor to enter into cells (*18*) (**Fig. 5b**). Thus, **SARSr-CoVs with diverse variants of SL-CoV S protein without deletions in their RBD can use human ACE2 as receptor for cell entry.**

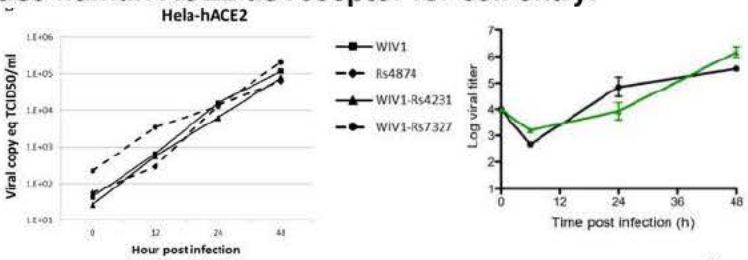

Fig. 5a (left): RT-PCR shows that bat SARSr-CoVs WIV1, Rs4874, and chimeras WIV1-Rs4231S, WIV1-Rs7327S grow in HeLa cells expressing human ACE2. **Fig. 5b (right):** Viral replication of SARS-CoV Urbani (black) and SARS-SHC014S (green) primary air-liquid interface human airway epithelial cell cultures at an MOI of 0.01.

We infected transgenic mice expressing hACE2 with 10^5 pfu of full-length recombinant WIV1 and three chimeric viruses (WIV1 backbone with SHC014S, WIV16S and Rs4231S). hACE2 transgenic mice challenged with rWIV1-SHC014S experienced ~20% body weight loss by 6dpi; rWIV1 and rWIV-4231S produced less body weight loss, and rWIV1-WIV16S led to no body weight loss **(Fig. 6a)**. At 2 and 4 dpi, viral loads in lung tissues of mice challenged with all three chimeras reached > 10^6 genome copies/g, significantly higher than rWIV1 infection **(Fig. 6b)**. This demonstrates that pathogenicity of SARSr-CoVs in humanized mice differs with divergent S proteins, **confirming the value of this model in assessing novel SARSr-CoV pathogenicity**.

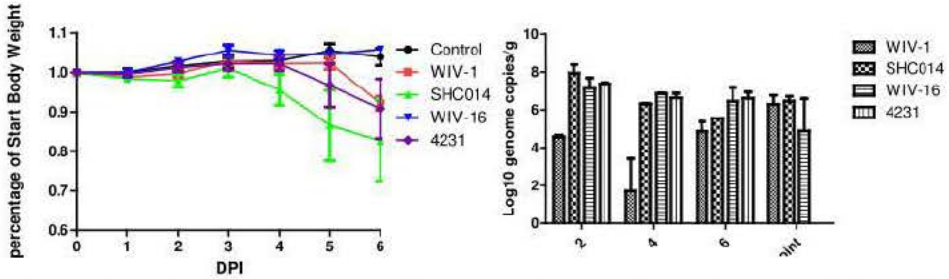

Fig. 6: *In vivo* infection of SARSr-CoVs in hACE2 transgenic mice. **6a (left)** Body weight change after infection; **6b (right)** Viral load in lung tissues.

Infection of rWIV1-SHC014S caused mild SARS-like clinical signs in the transgenic hACE2 mouse model **that weren't reduced by immune-therapeutic monoclonals that attenuate SARS-CoV pathogenecity.** Vaccination against SARS-CoV did not reduce severity of clinical signs in mice subsequently infected with rSARS-SHC014S (*18*). We found 2/4 broad human mAbs against SARS-CoV RBD cross-neutralized WIV1, but none could efficiently neutralize SHC014 which is less similar to SARS-CoV in the RBD (*39*). We repeated this virus characterization approach with chimeras using HKU3r-CoV S proteins that are ~25% divergent from SARS-CoV S, **and found that they are unable to use the ACE2 receptor**. Additionally, we were unable to culture HKU3r-CoVs in Vero E6 cells, or human cell lines. **The ability of HKU3r-CoVs to infect people, and their receptor binding target, remain unknown.**

This work has three implications for our R01 renewal: 1) some SARSr-CoVs currently circulating in bats in southern China **are likely able to infect and replicate within people**; 2) clinical outcomes of infection may include SARS-like illness that is currently **not treatable with mAb nor preventable with experimental vaccines**; 3) SARSr-CoV ability to bind human ACE2 is lost with S protein divergence between 10% (SHC014) and 25% (HKU3r-CoVs). **Although no viruses within this range have so far been described, these strains likely use hACE2 but could escape existing vaccines and immunotherapeutics and represent significant public health threats**. In our R01 renewal proposal, we will actively seek to identify viruses with this level of S protein divergence, characterize their binding targets *in vitro*, and their capacity to produce SARS-like disease that evades immunotherapy and vaccination *in vivo*.

Discovery of a novel bat-origin α-CoV associated with pig die-offs

Coronaviruses have a well-described propensity to jump the species barrier and cause new outbreaks (*40*). In 2016-17, we analyzed fecal samples from pigs at 5 farms in Guangdong Province (GD) affected by a fatal diarrheal disease. We discovered an α-CoV closely related to HKU2, and used PCR, serological and pathological data, followed by infection experiments to demonstrate that this novel virus, Swine Acute Diarrheal Syndrome coronavirus (SADS-CoV), caused the death of more than 20,000 pigs at these farms (*10*). We

identified SADSr-CoVs in *Rhinolophus* spp. bats in GD, and analyzed >30 full-length genomes to provide phylogenetic evidence that SADS-CoV originated in these bats (**Fig. 7**).

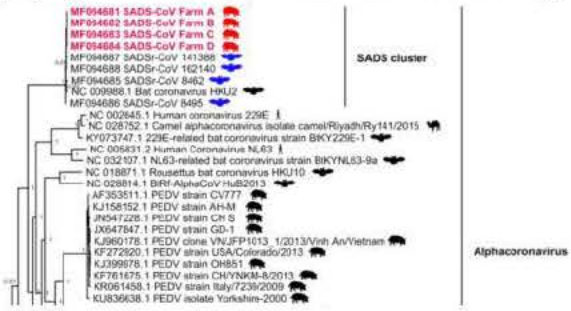

Fig. 7: Bayesian phylogenetic tree of the full-length genome sequences of SADS-CoV (red), bat SADSr-CoVs (blue), and related α-coronaviruses. Host species represented by symbol.

SADS-CoV replicates in Vero cells (*10, 41, 42*), but its capacity to replicate in human cell lines, and its zoonotic potential remains unknown. We developed a novel Luciferase Immunoprecipitation Systems (LIPS) antibody assay for SADS-CoV and found no evidence of spillover to pig farm workers at affected farms (0/33 people seropositive) (*10*). In the current proposed work we will include SADS-CoV diagnostic reagents in our serological panel to opportunistically screen human samples for evidence of spillover into people exposed to bats in southern China

Mapping bat viral emergence risk

We analyzed host and viral data for all known mammalian viruses and used a generalized additive models to correct for underlying sampling and reporting biases (*34*). This approach allowed us to predict the relative number of yet-to-be-described or 'missing' viruses that a species likely harbors. For China, there are distinct hotspots of unknown bat viral diversity in Yunnan Province **(Fig. 8)**.

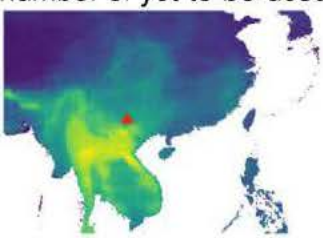

Fig. 8: Spatial distribution of predicted 'missing' or as-yet undiscovered viruses, from (*34*). Yellow = highest diversity, red triangle = Jinning Cave, Yunnan (*29*).

In a separate paper, we found that bat host diversity and climatic variability are correlates of viral diversity within bats, and that human population density, bushmeat hunting, and livestock production are correlates of the risk of transmission for viruses that spillover (*26*). The risk of spillover and spread differ spatially, suggesting that locations where bat viruses are most diverse may not be the most strategic sites for public health intervention (*21*). Work in the current proposal will improve on both approaches to identify hotspots of CoV emergence risk, by **using data from the high-risk locations and interfaces identified in our previous R01**, including better characterization of SARSr-CoV diversity in bats, and the potential of these viruses to cause infection.

Human risk behavior, the wildlife trade, and evidence of bat SARSr-CoV spillover.

Qualitative Study: Our previous R01 hypothesis was that SARSr-CoV spillover would most likely occur through the trade in bats for food, via the same market chains that to the emergence of SARS (*20*). To test this, we conducted an exploratory study using standardized one-on-one semi-structured ethnographic interviews and observational data in southern China among 88 people involved in trading wild bats, to assess local social and cultural norms and individual attitudes underlying contact with bats (publication in prep.). Our results suggest that in the 11 years since the emergence of SARS, there have been substantial changes to the wildlife trade: 1) Former wildlife markets are now predominantly selling captive-bred species (poultry, livestock, farmed wildlife); and 2) few bats are now sold through markets. We identified other risk factors for spillover, including people living near to bat roosts, and those visiting bat caves for hunting or recreation.

Human Questionnaire & Sero-surveillance: We used qualitative study findings to develop a human behavioral risk questionnaire on the type and frequency of animal contact, wildlife observed in daily life, and unusual illnesses reported over the past 12 months. We conducted a cross-sectional study among populations that live near bat caves or roosts where we had detected bat SARSr-CoVs. Study participants provided biological samples, and bats were concurrently captured and sampled. Questionnaires and biological samples (oropharynx swab, serum, plasma) were collected from 1,585 participants from 7 sites in Yunnan, Guangxi, and Guangdong provinces **(Fig. 9)**.

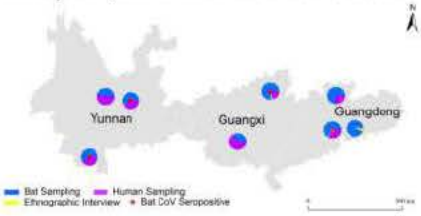

Fig. 9: Concurrent sampling in bats and target human population in communities in Yunnan, Guangxi, and Guangdong provinces. Pie-charts indicate sampling effort (bat sampling = blue, Human questionnaire and sampling = purple, Ethnographic interview = yellow, bat CoV seropositive = red)

Serological Evidence of Exposure: We developed serological assays for HKU9 CoVs (β), SARSr-CoV Rp3 (β), HKU10 CoV (α), and MERS-CoV (β) and used ELISA and Western blot to test serum samples collected in 2016/17. **We found 7 individuals (7/733, 0.95%) living within a 6 km radius of the Jinning Cave, and 6/209 people (2.87%) at one site, with evidence of exposure to bat SARSr-CoVs (Table 2; Fig. 8).**

Site	# tested	Bat CoV + (%)	SARSr-CoV Rp3 + (%)	HKU10 + (%)	HKU9 + (%)	MERS-CoV+ (%)
Jinning, Yunnan	209	6 (2.87)	6 (2.87)	-	-	-
Mengla, Yunnan	168	1 (0.6)	1 (0.6)	-	-	-
Jinghong, Yunnan	212	-	-	-	-	-
Lufeng, Yunnan	144	-	-	-	-	-
Guangdong	420	-	-	-	-	-
Guangxi	412	2 (0.48)	-	2 (0.48)	-	-

Table 2 (left): ELISA and Western blot confirmed testing of human sera for antibodies to 4 bat CoVs.

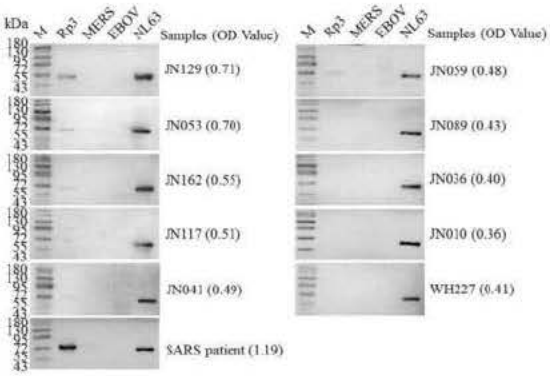

Fig. 10: Western blot reactivity of human sera to SARSr-CoV Rp3 NP.

We found evidence among human populations in Guangxi Province of people with prior exposure to the bat α-CoV HKU10 (2/412, 0.48%). This is of potential public health interest because HKU10 is known to be able to jump host species within bats, and therefore may have high propensity for emergence (*43*).

Behavioral Risk Factors: Questionnaire response and demographic data suggest specific risk factors included type of occupation, keeping of pets and visiting wet markets. Seropositive individuals were mostly farmers/peasants (8/9), living in Yunnan province (7/9), 41-60 yrs old (7/9), with domestic contact with rodents (6/9), and are male (6/9). However, these characteristics were common and occurred in ~40% of survey respondents. Although these results are preliminary and don't provide detailed information on routes of exposure, they suggest that further refining use serological tests coupled with qualitative and questionnaire data will identify likely routes of exposure to novel CoVs in China. In Aim 2 of this R01 renewal proposal, we identify strategies to better target at-risk people, and conduct focused **questionnaires and serosurveys to produce statistically significant findings.**

Analysis of self-reported illness: We analyzed data on self-reported symptoms of fever with cough and shortness of breath or difficulty breathing (severe acute respiratory illness - SARI), and fever with muscle aches, cough, or sore throat (influenza-like illness - ILI) from study participants. We used a least absolute shrinkage and selection operator (LASSO) regression to identify associations between ILI and/or SARI symptoms and contact with animals in the last year. Salient predictors or combination of predictors were, in descending order of odds ratio: 1) having eaten raw or undercooked carnivores; 2) having slaughtered poultry and being a resident of Guangxi province; 3) having had contact with cats or bats in the house; and 4) being a male who has raised poultry **(Fig. 11).**

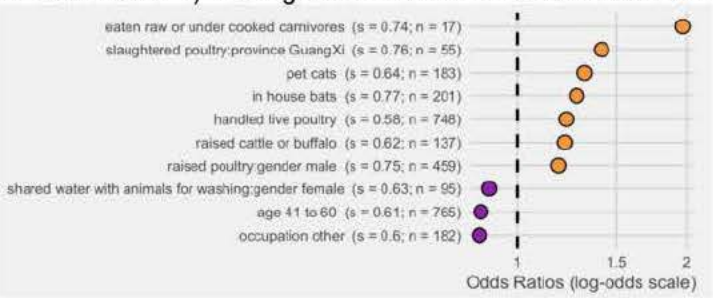

Fig. 11: Predictors of self-reported ILI and/or SARI in prior 12 months (s = bootstrap support; n = number +ve out of 1,585 respondents). Orange circles = odds ratios > 1 (positively associated with the outcome); purple = odds ratios <1 (negatively associated with the outcome).

2. APPROACH

Rationale and Innovation : Our previous R01 work demonstrates that bats in southern China harbor an extraordinary diversity of SARSr-CoVs, some of which are able to use human ACE2 to enter into human cells, can infect humanized mouse models to cause SARS-like illness, and evade available therapies or vaccines. We found that the bat hosts of SARSr-CoVs appear to no longer be traded in wildlife markets, and that people living close to bat habitats are the primary risk groups for spillover. At one of these habitats, we found diverse SARSr-CoVs containing every genetic element of the wild-type SARS-CoV genome, and serological evidence of human exposure among people living nearby. The proposed work in this renewal R01 builds on these

findings:. **In Aim 1, we will conduct targeted bat sampling at sites where we predict that undiscovered high risk SARSr-CoV strains exist**. Bat sampling will be targeted geographically and by host species to test predictions about evolutionary diversity of SARSr-CoV. We will analyze RdRp and S protein sequences to test their capacity for spillover to people in Aim 3. **In Aim 2,** will conduct focused, targeted human surveys and sampling **to identify key risk factors for SARSr-CoV spillover and evidence of illness**. To maximize our opportunity of capturing human exposure to bat CoVs, we will conduct community-based surveillance in regions with high SARSr-CoV prevalence and diversity, and individuals having contact with bats. We will assess bat-CoV seropositive status against a small number of questions about human-wildlife contact and exposure. We will conduct clinic-based syndromic surveillance close to these sites to identify patients presenting with influenza-like illness and severe acute respiratory illness, assess their exposure to bats via a questionnaire, and test samples for PCR- and serological evidence of SARSr-CoV infection. We will conduct follow-up sampling to capture patients who had not yet seroconverted at the time of clinic visit. **In Aim 3,** we will characterize the propensity of novel SARSr-CoVs to infect people *in vitro* using primary human airway epithelial cells and *in vivo* using the transgenic hACE2 mouse model. We will use mAband vaccine treatments to test our hypothesis that SARSr-CoVs with 10-25% divergence in S protein sequences from SARS-CoV are **likely able to infect human cells, and to evade mAb therapeutics and vaccines**. We will then map the geographic distribution of their bat hosts and other ecological risk factors to **identify the key 'hotspots' of risk for future spillover.** Our SARSr-CoV program serves as a model platform to integrate virologic, molecular and ecologic factors contributing to CoV emergence while informing high impact strategies to intervene and prevent future pandemics. This includes providing critical reagents, therapeutic interventions and recombinant viruses for future SARSr-CoV pandemic and public health preparedness.

Research team and management: We have reinforced our original collaboration between EcoHealth Alliance (EHA), a global leader in field investigations of emerging viruses from wildlife and modeling/analysis of viral risk, and Wuhan Institute of Virology, a global leader in bat viral investigations **(Fig. 12)**. First, we have included senior behavioral risk scientists Co-I Francisco (EHA) and Ren (Inst. Pathogen Biol., Beijing) to oversee human survey and sampling work in Aim 2. Second, Prof. Linfa Wang (Duke-NUS), a world leader in understanding the role of bats as hosts of emerging viruses, will act as a consultant by advising and assisting in the development of PCR and serological tests and virus characterization. Prof. Wang has developed a unique array of bat immunological reagents that enrich the serological arms of the proposal. Third, Prof Ralph Baric (UNC) will use his expertise in CoV characterization to conduct primary human epithelial airway cell infections to identify high risk strains that are poised for human emergence. He will oversee and participate in animal experiments in Aim 3. This expanded team will work on a more focused set of goals, based on the results of our previous R01. PI Daszak has collaborated with all partners for between 3 and 15 yrs and will host monthly calls, annual in-person meetings, conduct quarterly adaptive management to refine research lines of work.

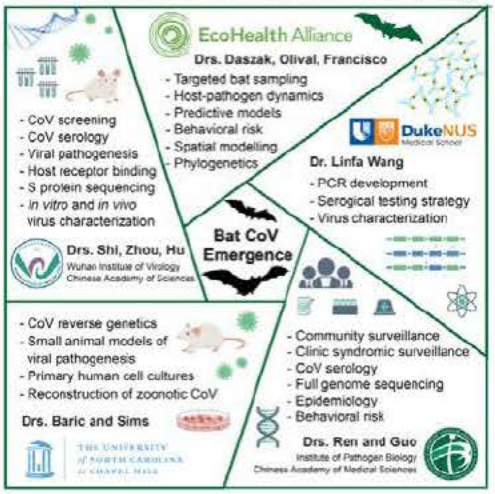

Fig. 12: Interdisciplinary team & roles in the proposed R01 renewal work.

Aim 1: Characterize the diversity and distribution of high spillover-risk SARSr-CoVs in bats in southern China

1.1 Rationale/Innovation: Our previous R01 work identified diverse SARSr-CoVs with high propensity for human infection (*18, 19, 29*). Characterization of these suggests SARSr-CoVs that are up to 10%, but not 25% different in the spike glycoprotein use human and bat ACE2 receptors for docking and entry. Uneven sampling gaps (e.g., no strains were found with 10-25% spike variation) prevent a thorough understanding of the transition point where the most divergent strains lose human ACE2 receptor usage. Our viral discovery curves **(Fig. 4)** suggest further sampling will reveal a rich diversity of as-yet-undiscovered SARSr-CoVs. In this aim we will use phylogenetic and viral discovery analysis to specifically target bat species and regions that are undersampled to allow sufficient power to identify and characterize missing SARSr-CoV strain diversity. Our

previous work also suggests that SARSr-CoVs with S proteins that are ~10% divergent from SARS-CoV resist neutralization by therapeutic mAbs and escape SARS-CoV vaccines (*17, 18, 23*), suggesting that . However, viruses with 10-25% divergence in S proteins may bind to human cell receptors, but completely evade therapeutic and vaccine effects, and could therefore be a higher risk for public health. We will sequence the S proteins of novel SARSr-CoVs to prioritize viruses for experimental work in Aim 3 to test this hypothesis.

1.2 General Approach: We will use sampling, testing, and CoV sequence data from our previous R01 to pinpoint sites and host species needing additional sampling. We will work in 4 provinces (Yunnan, Guangxi, Guizhou and Guangdong) that we have identified phylogenetically as having the highest diversity of as-yet-undiscovered SARSr-CoVs and with competent natural hosts. Precise sampling site locations will be refined in Y1. We will target at least 5,000 individual bats over 5 years from 15 currently undersampled species of *Rhinolophus* bats, which we calculate will allow us to almost fully characterize the expected natural diversity of SARSr- and other β-CoVs in the region. Bats will be captured, sampled, and released at the site. Specimens will be transported in liquid N_2 to Wuhan Inst. Virology (WIV) for PCR screening, and positive samples selected for further molecular characterization and S Protein sequencing. EHA will lead the study design, field sampling, and data analysis for this Aim; and WIV will lead the testing and viral sequencing.

1.3 Sampling and testing of bats: 1.3.a Site selection & sample sizes: In Y1 we will use our bat host and viral trait modeling, phylogeographic analyses of RdRp and S Protein sequences, and geographic and host species-based viral discovery curve analyses to identify SARSr-CoV diversity hotspot regions for bat sampling. We will sample at 8 new sites in four provinces. We will use cave site data (*44*), and demographic information to identify two sites in each of Yunnan, Guangxi, Guangdong, and Guizhou where humans likely have contact with bats. In Yunnan, we will identify two unsampled caves close to, but distinct from, the Jinning cave (*29*). This will provide adequate coverage of lineage 1 and 2 SARSr-CoVs, including a rich source of new HKU3r-CoVs, which have unknown potential for zoonotic spillover. Sampling will begin towards the end of Y1. We will use survey data from our previous R01 and host-specific viral accumulation curve data to target an additional 10 under-sampled *Rhinolophus* spp., 5 that were SARSr-CoV negative in our study, and a small number of related bat genera (including *Hipposideros* spp. and *Aselliscus* spp.) we previously found PCR positive for SARSr-CoVs **(Table 1)**. We will sample at least 5,000 bats from these 4 provinces (~1250 per province). Given ~5-12% prevalence of SARSr-CoVs in *Rhinolophus* spp. at our previous sites, **this sample size would give us 425 (±175) positive individual bats, and ~125 novel strains.**

1.3.b CoV screening, isolation: Viral RNA will be extracted from bat fecal pellets/anal swabs with High Pure Viral RNA Kit (Roche). RNA will be aliquoted, and stored at -80C. One-step hemi-nested RT-PCR (Invitrogen) will be used to detect the presence of CoV sequences using primers that target a 440-nt fragment in the RNA-dependent RNA polymerase gene (RdRp) of all known α- and β-CoVs (*45*). PCR products will be gel purified and sequenced with an ABI Prism 3730 DNA analyzer. We will attempt isolation on samples with diverse and interesting novel CoVs, using Vero E6 cells and bat primary cell culture.

1.3.c Sequencing S proteins: For all novel SARSr-CoV strains, we will sequence the complete S gene by amplifying overlapping fragments using degenerate primers as shown previously (*17, 29*). Full-length genomes of selected SARSr-CoV strains (representative across subclades) will be sequenced via high throughput sequencing method followed by genome walking. The sequencing libraries will be constructed using NEBNext Ultra II DNA Library Prep Kit for Illumina and sequenced on a MiSeq sequencer. PCR will be performed to fill gaps in the genome. The full length S gene sequences, including the amount of variation in the S receptor binding residues that bind the ACE2 receptor, will be used to select strains for Aim 3 experiments.

1.3.d Host ACE2 receptors: We will sequence host ACE2 receptors of different bat species or different bat populations from a single species (e.g. *Rhinolophus sinicus*) to identify relative importance of different hosts of high risk SARSr-CoVs and the *intra*specific scale of host-CoV coevolution. Of particular interest is homology across 18 bat and human orthologue ACE2 contact interface residues that engage the SARS RBD as a potential indicator of SARSr-CoV cross species transmission potential and growth in human cells (*46*).

1.4 Analyses: 1.4.a Bat-CoV evolution and distribution: We will use Bayesian and Maximum Likelihood phylogenetic analyses of RdRp, Spike, and full genome sequence (when available) data to reconstruct the evolutionary history of the novel bat SARSr-CoVs we identify. We will rerun MCC analyses **(Fig. 3)** to

reconstruct β-CoV evolutionary origins using our expanded dataset. Ecological niche models will be used to predict spatial distribution of PCR-positive species, identify target sites for additional bat sampling, and analyze overlap with people. We will use our viral discovery/accumulation curve approach **(Fig. 4)** to monitor progress towards discovery of >80% of predicted diversity of SARSr-CoVs (lineage 1 & 2) within species and sites, halting sampling when this target is reached, and freeing up resources for work other work (*37, 47*).

1.4.b Viral strain prioritization: Of the expected 100-200 novel SARSr-CoV strains, we will down-select to prioritize for further characterization based on S genes that are: i) different from SHC014, WIV1, SARS-CoV with diversity ranges of 10-25%; ii) have virus S RBD that could use human/bat receptors; iii) have recombinant chimeric spikes indicative of gene flow between clade I and II strains; iv) have bat ACE2 receptors that might select for spike RBDs that can use human receptors for entry (15/18 conserved residues in human/bat ACE2 molecules that bind SARSs-CoV S RBD domains are likely more efficient receptors than 3/18 conserved sites). Using structural models based on the SARS S glycoprotein, the extent and location of antigenic variation will be annotated onto the structure, further resolving the locations of highly evolving and conserved structural motifs/epitopes that function in protective immunity (*48-51*).

1.5 Potential problems/alternative approaches: Permission to sample bats in sites or provinces we select. We have a >15-year track record of successful field work in southern China and have worked with local authorities and communities to ensure access. We have existing permissions in place which will be renewed at the start of this project. Due to the abundance of distinct caves in the region, if access to a site is withdrawn, we will rapidly identify a suitable, complementary site with similar bat species composition and abundance. **We may not identify β-CoVs in our sample bat species due to seasonality of viral shedding.** The sampling regions we have selected are subtropical, and our previous data, and even published studies in temperate regions (*52*), do not suggest a strong pattern of seasonality in SARSr-CoV shedding. Nonetheless to account for this we will conduct sampling evenly on quarterly basis within each province.

Aim 2: Community and clinic-based surveillance to capture SARSr-CoV spillover, routes of exposure and potential public health consequences.

2.1 Rationale: Our previous R01 study identified serological evidence of exposure to SARSr-CoVs in certain communities in S. China **(Table 2, Fig. 10)** (*19*). However, the low seroprevalence (0.6%-2.7% at positive sites) suggests we need a larger sample size, and a focused, targeted study to maximize the likelihood of identifying seropositive cases. In Aim 2, we will use combined biological-behavioral risk surveillance in targeted populations within the community and clinical settings to 1) identify risk factors correlated with seropositivity (exposure to) and PCR positive status (infection with) bat SARSr-CoVs; and 2) assess possible health effects of SARSr-CoVs infection in people. Obtaining this information could be a significant step in understanding the likelihood of recent 'hidden' spillover events and their public health impacts, as well as the risk of future emergence of SARS-like diseases. It will also support the development of risk-mitigation strategies by public health authorities within China and other countries with bats that harbor these viruses (e.g. most of SE Asia).

2.2 General Approach/Innovation: We will use a dual study design to gain in depth understanding of exposure and risk factors for SARSr-CoV spillover **(Fig. 13)**. We will conduct community-based surveillance, with more focused questionnaires and biological sampling to determine the seroprevalence of SARSr-CoVs in at-risk human populations, and to identify risk-factors for SARS-CoV spillover in these communities. We will conduct clinic-based syndromic surveillance and biological sampling at sites that include the community-based surveillance sites within their catchment. This will include follow-up sampling to capture seroconversion of recently infected people and the full course of symptoms. We will also use PCR to test for present of active SARSr-CoV replication. Both community-based and clinic-based syndromic surveillance programs are case-control studies designed with the sample sizes necessary to statistically quantify (with a power of 80%) risk factors and health impacts for SARSr-CoV spillover, linked to serological status and symptoms.

2.3 Target population & sample size: We will target sites in the same four provinces, and close to those for bat sampling, based on: 1) sites of likely divergence of SARSr-CoVs; 2) a high diversity of SARSr-CoVs within the S protein sequence divergence of 5-25%; and 3) high rates of human-wildlife interactions. Community-based surveillance will be conducted at 2 sites in each of the 4 provinces, a total of 8 sites. From our previous work we anticipate that 10-30% of the community population will have had exposure to bats allowing us to

capture highly exposed and non-exposure individuals at each site. Individuals living or working around bat roosts, who hunt wildlife, work with wildlife or livestock farming, transportation, selling, or slaughtering wildlife in the surveyed areas will be targeted so that they make up ≥30% of the sampled population in each community. We will stratify sampling to ensure appropriate representation of sex, demographic, and socio-economic factors at each community site. We will initiate active clinic-based syndromic surveillance at 2 county-level hospitals and 1 provincial-level hospital in each of the 4 provinces, in total 12 hospital sites, all within the catchment areas for bat sampling, and which are used by people in our community-based surveillance. Patients ≥ 12 years old presenting at the health facility who meet the syndromic and clinical case definitions for SARI and ILI will be recruited into the study. We will enroll a total of at least 2,750 individuals for clinical studies, which accounts for up to 40% loss from follow-up. Study data will be pooled across sites, as clinical patients are limited by the number of individuals presenting at hospitals. For community-based surveillance, we will enroll 1,650 individuals per province, pooled across two sites for each province, allowing us to make province-level comparisons of differing effects. Estimating 5% overall seroprevalence in these high-exposure populations, these sample sizes are sufficient to estimate effect sizes of behavioral risk factors of 2X or greater with 80% power.

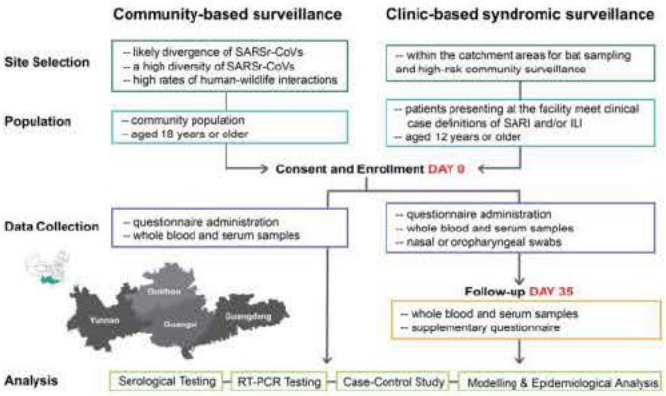

Fig. 13: Human survey and sampling study design overview

2.4 Data & sample collection: At both community and clinical settings, following enrollment with signed consent form, biological specimens (two whole blood samples, one max. 500 µL; two 500 µL serum samples) will be collected from all eligible participants, and a questionnaire will be administered. We will investigate five risk factors, so as to maximize the power of the analyses, all related to high risk wildlife exposure, based on continuing analysis of our previous work, and will include: 1) occupation; 2) observed or reported interactions with bats in/around house; 3) proximity to nearby bat roosts; 4) working or regular visit to animal markets; 5) self-reported ILI/SARI. An additional two nasal or oropharyngeal swabs will be collected from patients enrolled in the clinic-based syndromic study. With permission from each clinic, and consent from participants, we will review clinical records to collect data on medical history, clinical syndromes, and patient etiology.

2.5 Clinic enrollment and follow-up: We will recruit inpatients and outpatients after initial screening to meet the clinical case definition of 1) severe/acute respiratory illness (SARI/ARI) of unknown etiology; or 2) Influenza-like illness (ILI) of unknown etiology. Once enrolled, biological samples will be collected and a questionnaire administered by trained hospital staff that speak appropriate local dialects. Samples will be taken concurrently when collecting samples for normative diagnostics. For inpatients, samples will be collected within 10 days of reported onset of illness to increase the chance of PCR CoV detection (*53*). We will follow up 35 days after enrollment to collect an additional two 500 µL serum samples conduct a standardized questionnaire supplement to collect additional data on the course of symptoms in the interim period. 35 days gives adequate time for development of IgG, which occurs <28 days after onset of symptoms for SARS patients (*54*).

2.6: Laboratory analysis: 2.6.a Serological testing: In our previous R01, we expressed his-tagged nucleocapsid protein (NP) of SARSr-CoV Rp3 in *E.coli* and developed a SARSr-CoV specific ELISA for serosurveillance using the purified Rp3 NP. The specificity of this assay was evaluated using polyclonal antibodies against HKU1, OC43, 229E, NL63, MERS-CoV and EBOV and no significant cross-activity was detected (*19*). **While this shows it is a specific test for Rp3, it suggests that if we can expand our serology tests to cover other bat CoVs, we may identify many more seropositive individuals.** In this renewal, we will therefore use two serological testing approaches. First, we will expand test all human sera collected from both community- and clinic- based sampling for a panel of bat CoVs that will include SARS-CoV (outbreak strain), a range of lineage 2 SARSr-CoVs (including WIV1 and SHC014), lineage 1 HKU3r-CoVs, MERS-CoV, and the α-CoVs SADS-CoV and HKU10. We previously found serological evidence of human exposure to HKU10 (*19*), but HKU10 is known to jump from one host bat species to another (*43*) and is

therefore likely to have infected people more widely. Incoporating serological testing for SHC014 is also likely to yield higher seroprevalence because it is readily divergent from SARS-CoV wildtype and therefore unlikely to have been picked up in our earlier testing. It is possible that non-neutralizing cross reactive epitopes exist that afford an accurate measure of cross reactivity between clade1 and clade 2 strains which would allow us to target exposure to strains of 15-25% divergence from SARS-CoV. Additionally, we will test samples for antibodies to common human CoVs (HCoV NL63, OC43 – see potential pitfalls/solutions below). Secondly, we recognize that CoVs have a high propensity to recombine. To serologically target 'novel' recombinant virus exposure, we will conduct 1) ELISA screening with SARSr-CoV S or RBD; 2) confirm these results by Western blot; then 3) use NP based ELISA and LIPS assays with a diversity of SARSr-CoV NP. For NP ELISA, micro-titer plates will be coated with 100 ng/well of recombinant batCoV NP and incubated with human sera in duplicates followed by detection with HRP labeled goat anti-human IgG antibodies. For confirmation, all ELISA positive samples will be subjected to Western blot at a dilution of 1:100 by using batCoV-NP as antigen. We will use an S protein-based ELISA to distinguish the lineage of SARSr-CoV. As new SARSr-CoVs are discovered, we will rapidly design specific Luciferase immunoprecipitation system (LIPS) assays targeting the S1 genes of bat-CoV strains for follow-up serological surveillance, as per our previous work (*10*).

2.6.b RT-PCR testing. Specimens will be screened using RT-PCR for the RdRp gene (See section 1.3.b for details). Positive samples will be subjected to full genome sequencing and RT-PCR amplification of the S glycoprotein gene. Samples from the clinic-based syndromic surveillance will also be tested using RT-PCR for Influenza A & B, HCoV NL63, OC43, HKU1, SARS-CoV & 229E as rule-outs, and SADS-CoV and HKU10 as an opportunistic survey for potential spillover these CoVs as proposed above.

2.7 Epidemiological analysis: We will conduct a case-control study to identify risk factors for SARSr-CoVs spillover. "Cases" are defined as participants whose samples tested positive for SARSr-CoVs by serological tests. "Controls" will be selected from the pool of participants admitted to the studies but testing negative. We will use nearest neighbor matching to pair cases demographically with controls at a 1-to-3 ratio or greater. We will use generalized linear models to analyze correlation between serological/PCR status and risk factors including: Activities with likely exposure to 1) bats; 2) livestock; and 3) locations of residence and work. We will use the same procedure to determine how clinical presentation differs between SARSr-COVs-exposed and unexposed enrollees, in the time course of illness, severity of symptoms, and type of symptoms.

2.8 Potential problems/alternative approaches: Rarity of spillover events means it may be difficult to identify sufficient seropositives to statistically analyze risk behavior or illness. First, we are now targeting our community-based surveillance to subpopulations with high-levels of bat exposure, at sites selected for diverse and prevalent SARSr-CoVs, and are adding clinic-based syndromic surveillance of SARI and ILI cases in these same regions – both will increase likelihood of finding positive individuals. Second, our serology testing will include a panel of assays for a large diversity of lineage 1 and 2 SARSr-CoVs as well as SADS-CoV, HKU10 and other bat-borne CoVs. *Rhinolophus* spp. bats host all of these (overall bat CoV PCR prevalence, 11.8%; β-CoV, 3.4%; α-CoV, 9.1%). Thus, using this broad serological panel to screen individuals in likely contact with these species increases the potential for detecting spillover with enough power for statistical analyses, and will shed light on behaviors that predispose to CoV spillover from bats. Third, we will include common human CoVs in our panel, so that even if low prevalence of bat CoVs is found, we will be able to conduct a valuable cross-sectional study of the seroprevalence of human CoVs. Finally, we will be able to assess relative measures of human-wildlife contact from our survey work. We will analyze intensity of contact against other risk factors and clinical outcomes to provide useful proxy information for spillover risk. **Patients visiting clinics may have cleared virus, but not yet developed IgG antibodies, reducing seropositive cases.** Our 35 day follow-up sampling should avoid this because the maximum lag time between SARS infection and IgG development was ~28 days (*53*). We also expect that patients in rural communities will only visit clinics when symptoms have progressed, likely coinciding with late illness and onset of IgG. We will also have data from our community study, so won't be completely reliant on hospital data to identify PCR- or seropositives. Finally, the risk is outweighed by the potential public health importance of discovering active spillover of a new SARSr-CoV. **Serological testing may not match known CoVs due to recombination events.** We will use the three-tiered serological testing system outline in **2.6.a** to try to identify these 'novel' CoVs, however, we will also remain flexible on interpretation of data to ensure we account for recombination.

Aim 3: *In vitro* and *in vivo* characterization of SARSr-CoV spillover risk, coupled with spatial and phylogenetic analyses to identify the regions and viruses of public health concern

3.1 Rationale/Innovation: In **Aim 1**, we aim to expand the known diversity of SARSr-CoVs by over 125 strains, targeting 10-25% S protein divergence that we predict infers high spillover risk and evasion of immune therapeutic and vaccine efficacy. In **Aim 3**, we will further characterize the zoonotic potential of a selected group of these novel SARSr-CoVs, using infectious clone technology, *in vitro* and *in vivo* infection experiments and analysis of HKU3r-CoV receptor binding to test the hypothesis that S protein % sequence divergence thresholds predict spillover potential (*18, 55*). We will analyze data from these viral characterization and infection experiments, coupled with bat host distribution, viral diversity and phylogeny, human survey of risk behaviors and illness, and human serology to assess spillover risk of SARSr-CoVs in different bat species across southern China. This will enable future development of public health interventions and enhanced surveillance to prevent the emergence of a novel SARSr-CoV.

3.2 General Approach: We will use S protein sequences to select a range of viral strains that cover the 10-25% S protein divergence we predict as high public health potential and construct chimeric SARSr-CoVs using the WIV1 backbone and these S genes as done previously (*12, 18, 38*). We will rescue of full-length clones and assess infection of non-permissive cells expressing human, bat and civet ACE2 receptors, Vero cells, primary human airway epithelial cells, and CaCo cells for HKU3r-CoVs (which have not been cultured and may use intestinal epithelium in nature). We will conduct experimental infections in hACE2 transgenic mice to assess pathogenicity and clinical signs (*18*). Finally, using a panel of mAbs that neutralize SARS-CoV infection *in vitro* and *in vivo*, and vaccine against SARS-CoV S protein, we will examine the capacity of strains with divergent S protein sequences to evade therapeutics, revealing strains with high public health potential. We will also conduct limited experiments to analyze HKU3r-CoV receptor binding and assess spillover potentia. Using these results, and data from Aims 1 and 2, we will use spatial modeling techniques to identify geographic hotspots in southern China where bat species that harbor high risk SARSr-CoVs inhabit, where communities that have high exposure to bats exist, where serological or PCR evidence of spillover has been identified, and where underlying demographic or environmental trends suggest high risk of future emergence.

3.3 Virus characterization: 3.3.a Construction of chimeric SARSr-CoV viruses: Infectious clones with the S gene of novel SARSr-CoVs and the SARSr-CoV WIV1 genome backbone using the reverse genetic system developed in our previous R01 (*24*). The correct infectious BAC clones will be screened by BAC DNA digestion with appropriate restriction enzyme or PCR amplification. The chimeric viruses will be rescued in Vero cells and then verified by sequence analyses. Our research group is well versed in coronavirus reverse genetics.

3.3.b Cell entry analysis: HeLa cells expressing human ACE2 are cultured on coverslips in 24-well plates incubated with the chimeric bat SARSr-CoVs with different spike proteins at a multiplicity of infection (MOI) = 1.0 for 1h. The inoculum is removed and the cells are washed twice with PBS and supplemented with medium. HeLa cells without ACE2 are used as negative control. Twenty-four hours after infection, cells are rinsed with PBS and fixed with 4% formaldehyde in PBS (pH7.4) at 4℃ for 20 min. ACE2 expression is detected by using goat anti-human ACE2 immunoglobulin followed by FITC-labelled donkey anti-goat immunoglobulin. Virus replication is detected by using rabbit antibody against the nucleocapsid protein of bat SARSr-CoV followed by Cy3-conjugated mouse anti-rabbit IgG. In parallel with the immunofluorescence assay, plaque assay will be conducted to determine the viral titers and growth kinetics in the infected cells at different times post-infection.

3.3.b Primary human airway epithelial cell culture: Primary human ciliated airway epithelial cells (HAE) cultures from the lungs of transplant recipients represent highly differentiated human airway epithelium containing ciliated and non-ciliated epithelial and goblet cells, grown on an air-liquid interface for several weeks prior to use (*18, 55, 56*). We will prepare HAE cultures from three different patient codes in triplicate in collaboration with the tissue procurement facility at the Cystic Fibrosis Center at UNC. Cultures will be inoculated with chimeric bat SARSr-CoVs to assess efficient replication. At 72 hpi, cultures will be fixed for immunofluorescent staining using antisera to the SARS-CoV conserved nucleocapsid protein (N) (*57, 58*). SARSr-CoVs that differ significantly in S protein sequence (11-24%) from epidemic SARS-CoV yet replicate *in vitro*, will also be evaluated for sensitivity to neutralization in Vero cells by PRNT50 assays using broadly SARS-CoV cross reactive human mAbs S227.14, S230.15, S215.17, and S109.8 (*49, 55*). As controls, the S

genes of novel SARSr-CoV will be inserted into VEE 3526 replicon vectors (VRP3526-S), packaged and used to vaccinate mice (*59*). Polyclonal sera will be harvested and tested for ability to cross neutralize SARS-CoV, GD03, WIV-1, SHC014, WIV-16, other novel SARSr-CoV and HKU3-SRBD by PRNT50 assay (*55, 60, 61*). Using PRNT50 titers with sera (n=4 each) among these viruses, antigenic cartography (*62*) will allow comparison of antigenic vs. phylogenetic distance, identifying the transition at which SARSr-CoV strains escape SARS-CoV based vaccines, informing immunotherapeutic and vaccine design strategies (*63-65*).

3.3.c Humanized mouse infection experiments: Briefly, in BSL3, n=5 10- to 20-week old hACE2 transgenic mice will be intranasally inoculated with 1 x 10^4 PFU of wildtype WIV-1 or chimeric bat SARSr-CoVs with different spike proteins, then monitored daily for weight loss, morbidity, and clinical signs of disease. Mice will be euthanized at 2, 4, and 6 dpi (endpoint of 14 dpi), organs harvested and virus quantified by SARS-CoV NP RT-PCR. After 7 days in formalin, tissues for histopathology will be removed from the BSL3 and stained with H&E, and for immunohistochemistry using polyclonal antibodies against the N protein. We will conduct limited evaluation of existing countermeasures using therapeutic monoclonal antibodies *in vitro* and *in vivo*. Existing SARSr-CoV mAbs will be diluted in DMEM starting at 1:20, and serial 2-fold dilutions mixed with an equal volume of 50 PFU of chimeric bat SARSr-CoVs with different spike proteins, then incubated on Vero E6 cells at 37°C for 1 h, then changed to a semi-solid medium (DMEM containing 1% methylcellulose and 2% FBS). Antibody neutralization titers will be evaluated by plaque quantification at 4-5dpi. hACE2 transgenic mice will be injected with SARS-CoV mAbs, and infected with chimeric bat SARSr-CoVs. Clinical signs and morbidity will be assessed and tissue pathology examined and compared with mice without treatment of mAbs to determine the therapeutic effect on SARSr-CoV infection, and protection of SARSr-CoV by wildtype SARS-S based vaccines assessed as described (*56, 66*). We will sequence full length genomes of high risk strains that are antigenically distinct and escape SARS cross neutralization, synthetically reconstruct a small subset (1-2) and evaluate the ability of nucleoside analogues to inhibit growth in HAE cultures and/or *in vivo* (*55, 56*).

3.3.d HKU3 clade cellular receptor: We will screen potential receptor molecules by pull-down analysis on membrane proteins interacting with the spike protein, initially using bat primary intestinal epithelial cell lines and lysates to extract protein, isolate membranes, and proteomically sequence intestinal proteins. The fusion protein of the HKU3 and HKU3r-CoV S proteins containing human Immunoglobulin Fc fragment will be eukaryotically expressed and purified. SARSr-CoV S will be incubated as bait protein with the membrane proteins extracted from *Rhinolophus sinicus* intestinal cells, to capture and precipitate membrane proteins that interact with the S protein. Mass Spectrometry will be performed to screen for the candidate receptor molecules and Co-Immunoprecipitation assay to confirm binding of the SARSr-CoV S protein to the candidate receptor. Alternatively, retroviruses pseudotyped with the SARSr-CoV S protein will be constructed and used to infect cells trans-expressing the candidate receptor molecule. Luciferase activity will be measured to test whether the S protein can bind to the receptor. If successful, this work will allow future research to clone and study human HKU3 receptor ortholog's ability to function as a receptor for other clade 2 strains and will allow better assessment of risk of clade 2 SARSr-CoV spillover to humans.

3.4 Combined spatial risk 'hotspot' analyses: We will use data from **3.3** to identify rank SARSr-CoV strains most likely to infect people and evade therapeutic and vaccine modalities. We will use bat survey and zoological data (*44*) to build species distribution models (*67*) and predict the distribution of bat species that harbor low, medium and high risk viral strains. Stacking these modeled distributions for the ~20-30 *Rhinolophus* and related species that occur in the region will allow estimates of SARSr-CoV diversity for a given locality. We will use machine learning models (boosted regression trees) and spatial 'hotspot' mapping approaches to identify the ecological, socio-economic and other correlates of SARSr-CoV diversity and spillover (from serosurveys) (*21, 68, 69*). We will include data from our human behavioral surveys and sampling to give a direct measure of where risk of spillover to people is likely to be highest in the region.

Potential problems/alternative approaches: We may not be able to glean further information about the **capacity of HKU3r-CoVs to infect human cells, or bind to human cell surface receptors.** If attempts at culture are unsuccessful, and efforts to identify the receptor too costly or time-consuming, we will cease this line of work. In that event, we will focus entirely on filling out the gaps in the 10-25% S protein sequence divergence from SARS-CoV, by working on a greater diversity of lineage 2 SARSr-CoVs.