Technical subtleties:

How to divide the samples initially? Easy, we use the rule of powers of two, that is, let 1/N be the probability that a sample is positive, and let's say for example N=100. Then there is a unique k that is the greatest power of two that is smaller than N (In the example k=64). The number of samples to be merged together in the first iteration of the experiment is then k/2 (In the example 32 samples). This method guarantees two things: 1) the number of samples analyzed per unit of time is the maximum possible 2) at no step in our analysis will we ever find ourselves having more tubes to analyze than the thermal cycler can accommodate*(see note1) (meaning that only in one analysis out of 10 will it happen that indeed the thermal cycler is full and we have to either use another thermal cycler or finish the analysis of 2 out of 1000 patients with a 3-hour delay. If you have 3 thermal cyclers, the number of analyses that delay even one patient by 3 hours drops to 1/50). **What not to do when estimating the a priori probability p that a sample is infected?** The answer is that you should not estimate it downward, otherwise the thermocycler will be full and many analyses (those of positives to boot) will be delayed by 3 hours. So if you are undecided between p=1/100 and p=1/130 you must choose p=1/100 and divide the samples into groups of 32.

How does our procedure for PCR relate to all the preprocessing of samples? There is only one thing to remember, which is what has brought us this far, and that is that bundling the analyses wins out over doing them individually, so for example the genomic extraction step CANNOT be done on a patient-by-patient basis, otherwise this step would be the bottleneck, but must be done in parallel with the PCR processes and with the samples that will be needed in the next PCR step.

Then swabs from all patients (e.g., 1024 patients) arrive at the lab, the swabs are soaked in a solution in which therefore viral particles will be present, aliquots*(see note2) are taken from each sample that are mixed together in groups of 32 for example (obviously 32 depends on the probability of a sample being positive as explained in the appropriate section). Each of these groups will undergo genomic extraction, resulting in 32 tubes (32x32=1024) in triplicate=96, the right number to perform the first PCR. In the 3 hours that the PCR is busy, we take aliquots from all the samples and divide them into groups of 16 so that each group of 32 is exactly divided into 2, and from each of these we extract the genomic, for a total of 64 tubes in triplicate=192 tubes. Fortunately for us, the throughput of the extractors is about twice the throughput of PCR*(see note3), so PCR and sample extraction should about finish together. Now that the PCR is finished some groups of 32 tested negative (more than 60%), and all extracts from 16 swabs that corresponded to tubes of 32 that tested negative can be thrown out instantly. Now it is time to do PCR with the 16 batches that came from 32-positive batches. These will almost never totally occupy the thermal cycler*(but if it happens see note4). So now we do PCR in parallel to the groups of sixteen and extraction to the groups of 8 that were in the groups of 32 that we know had at least one positive. For the same reason as before extraction and PCR end up about together, and you proceed in this way until the PCR is testing on one person, and at the end of that you know exactly which patients are positive and which are not.

Logistics, convenience, and criticality of the method.

Positive sides: There is no question about it, when the probability p of having a positive swab is low our method performs better than the standard method by a factor of 4 if p=1/32, up to a factor of 10 if p=1/128 for example, and moreover after only the first iteration the number of negatives detected are half of the total people analyzed, i.e. for p=1/32 there is a gain in the number of outcomes after 4h 30m equal

to a factor of 8, and for p=1/128 even a factor of 32.
Criticism: our method is cost-effective compared to the current method (in the sense that it increases throughput at the expense of increasing by at least 3 hours the time of analysis of some patients) only if the probability p that a swab is infected is less than 30% *(see footnote 5), so it is not useful for screening patients in serious condition or with respiratory crisis, when fast analysis is needed and we are already almost certain that the patient is positive, but rather for "mass" screening, or of health care personnel, where the percentage of positives is well below 30%. Although it is true that our method would have a much higher throughput than the currently used method, only half of the patients screened could be alerted after the normal 4h 30m wait, (still 600 patients compared to 32 of the standard method in our example with p=1/64) while a quarter of the patients would be alerted after an additional 3 hr, an eighth after another 3h 30m wait, and so on until you get to the positives, the last to be detected, who in our example (1024 patients, prob infected=1/64, initial groups of 32) would be alerted only after about 19h 30m from the start of the analysis. Two other technical critical issues are that one has to keep the swabs of individual patients until the last extraction is performed (and it could take about 20h from the start of the procedure), and finally that the concentration of viral particles could be too low in the early stages of the procedure (when the groups of 32 are formed) and false negatives could occur.
This can be solved, for example, by possible concentration by centrifugation to be carried out along with genome extraction.

Notes
1) Here we used the word thermal cycler as a synonym for the machinery used to do real time RT_PCR for short.
2) The liquid fraction of the swab to be taken should not exceed 1/(N+1), where N is the exponent of two when 2^N swabs are initially processed together (in our 32 samples, N=5) **3)**Customarily in hospitals, the inactivation and extraction phase takes about 1h 30min and the PCR 3h.
4) if at a certain point the PCR machine fills up and not all the tubes could find a place in it it would be enough that no patient would be forced into two rounds in which his genomic is not processed, this is the way to make the delay as short as possible for everyone, so let's say at one point a tube with the genomic of 4 people is left out, at the next round it is processed along with the two-by-two groupings leaving out one group of two, at the next round if it tested positive the 3 groups of 2 are processed with all groups of 1, and in this way you have delayed at most six patients for a time of one PCR run, no patient has waited two extra PCR runs. This eventuality is very rare anyway.
5) 30% is the probability of positivity below which our method has a higher patient throughput than the standard method, even though it takes half of the patients 2 PCR runs to know the result, so actually it would be better to use our method for positivity below 20% and never when timing in patient care is vital.