Using Kolmogorov Complexity to Prove some Primality Theorems

First, we present one of the results given in class:

Let $p_1, p_2, p_3, \ldots$ be an enumeration of the primes (in order, so that $p_1 = 2, p_3 = 5$, etc.)

Recall that we say that a string x is ℓ-random if $C(x) \geq |x| - \ell$.

Claim: There is a constant c such that, if $p_m | x$ where x is 7-random, then $p_m \leq cm \log^2 m$.

Proof: Let m and x be as above, where $r = |x|$. Since $p_m | x$, let $x = yp_m$.

Thus x is described by the program (of length $c' = O(1)$):

> Take as input a string $\underline{i}my$, and decode this to find i (using the self-delimiting property of the encoding $\underline{i}$, and interpret i as the length of m (where the remainder of the input is the string y). Using m, compute p_m, and output $y \cdot p_m$.

This shows that $C(x) \leq O(1) + 2|i| + |m| + (|x| - |p_m|) = O(1) + 2|\log\log m| + \log m + (r - \log p_m)$ But we also have $r - 7 \leq C(x)$. Thus we obtain $\log p_m \leq O(1) + 2|\log\log m| + \log m$. This is equivalent to the statement of the claim (merely taking logs of both sides of the claim).

Observe that, for any N such that $N = p_m$ for some prime p_m that satisfies the hypothesis of the claim above, we have that $\pi(N) \geq N/c\log^2 N$, where $\pi(N)$ denotes the number of primes less than or equal to N. This is because $\pi(N) = \pi(p_m) = m \geq p_m/(c\log^2 m) = N/(c\log^2 m) \geq N/(c\log^2 N)$. However, we still need to establish that there are many such primes p_m.

Claim: There are infinitely many primes that divide an ℓ-random number (for each value of ℓ).

Proof: Assume, for the sake of a contradiction, that the only primes that divide an ℓ-random number are $q_1, q_2, \ldots q_k$. Thus each ℓ-random number is described by a program of size $O(1)$ that takes as input two tuples $(e_1, \ldots, e_k)$ and $(q_1, \ldots, q_k)$ and outputs $x = \prod_{i=1}^{k} q_i^{e_i}$. If $|x| = n$, then note that each e_i has length at most $\log n$, and thus $C(x) \leq k \log n + \sum_{i=1}^{k} |q_i| + O(1) = k \log n + \log(\prod_{i=1}^{k} q_i) + O(1)$. (Since we're assuming that $q_1, q_2, \ldots q_k$ is a fixed list, it follows that $\log(\prod_{i=1}^{k} q_i)$ is actually $O(1)$, but it will be useful to keep track of the size of this part of the "constant".) Since $n - \ell \leq C(x)$, we thus have $n - \ell \leq k \log n + \log(\prod_{i=1}^{k} q_i) + b$ for some constant b. But this clearly fails for large enough n. (Picking n a bit larger than $k \log(\prod_{i=1}^{k} q_i) + \ell + b$ is sufficient.)

This implies that, for infinitely many lengths n, there are at least $2^n/O(n^2)$ primes of length at most n. In fact this holds for *all* large n, and not merely for

infinitely many such n, but I am not aware of a *simple* proof of this fact, using Kolmogorov complexity.