

Supplementary Materials for

Discovery of ROH diplotype clusters and their associations with

diseases in UK Biobank

Ardalan Naseri^{1,+}, Degui Zhi^{2,*}, and Shaojie Zhang^{1,*}

¹Department of Computer Science, University of Central Florida, Orlando, Florida 32816, USA

²School of Biomedical Informatics, University of Texas Health Science Center at Houston,
Houston, Texas 77030, USA

*Correspondence: degui.zhi@uth.tmc.edu and shzhang@cs.ucf.edu

⁺Present address: School of Biomedical Informatics, University of Texas Health Science Center at
Houston, Houston, Texas 77030, USA

Algorithm 1 Find all blocks of matches while maximizing the number of haplotypes

procedure MAXIMIZE_HAPLOTYPES(x_k, t, a_k, d_k, L, c) $V = []$, $Q = []$ **for** $r = 0$ **to** t **do** $m.append([])$ $pq.append(k+1)$ $V.append([])$ $Q.append([])$ **end for** $i_0 \leftarrow 0$ **for** $i = 0$ **to** $M - 1$ **do****if** $d_k[i] > k - L + 1$ **then** $changed \leftarrow false$ **for** $p = 0$ **to** t **do****for** $q = p + 1$ **to** t **do****if** $len(m[p]) > 0$ **and** $len(m[q]) > 0$ **then** $changed \leftarrow true$ $break$ **end if****end for****end for****if** $changed$ **and** $i - i_0 \geq c$ **then** $blockMatches = []$ $d_{min} \leftarrow 0$ $blockMatches.append(a_k[i_0])$ **for** $i_b = i_0 + 1$ **to** i **do** $blockMatches.append(a_k[i_b])$ **if** $(d_k[i_b] > d_{min})$ **then** $d_{min} = d_k[i_b]$ **end if****end for** $report\ d_{min}\ and\ blockMatches$ **end if** $i_0 \leftarrow i$, $m = []$ **for** $r = 0$ **to** t **do** $m.append([])$ **end for****end if** $updateV_Q(x_k[l], i, t, a_k, d_k, pq, V, Q)$ **end for** $a_{k+1} \leftarrow concatenate[V_0, V_1, ..V_t]$ $d_{k+1} concatenate[Q_0, Q_1, ..Q_t]$ **end procedure**

Algorithm 2 Update intermediate variables for cPBWT

```
procedure UPDATEV_Q(allele, i, t, ak, dk, pq, V, Q)  
  for p = 0 to t do  
    if dk[i] > pq[p] then  
      pq[p] ← dk[i]  
    end if  
  end for  
  V [allele].append(ak[i])  
  for p = 0 to t do  
    if allele == p then  
      Q[allele].append(pq[allele])  
      pq[allele] ← 0  
      mab[p].append(ak[i])  
    end if  
  end for  
end procedure
```

Algorithm 3 Find all blocks of matches larger than L while maximizing the length of match

procedure MAXIMIZE HAPLOTYPES(x_k, t, a_k, d_k, L, c)

```
 $V = [], Q = []$ 
for  $r = 0$  to  $t$  do
   $m.append([])$ 
   $pq.append(k+1)$ 
   $V.append([])$ 
   $Q.append([])$ 
end for
 $i_0 \leftarrow 0$ 
for  $i = 0$  to  $M - 1$  do
  if  $d_k[i] > k - L + 1$  then
     $changed \leftarrow f$  else
      for  $p = 0$  to  $t$  do
        for  $q = p + 1$  to  $t$  do
          if  $len(m[p]) > 0$  and  $len(m[q]) > 0$  then
             $changed \leftarrow true$ 
            break
          end if
        end for
      end for
    end if
  end for
  if  $changed$  and  $i - i_0 \geq c$  then
     $blockMatches = [], d_{min} \leftarrow 0, f_a \leftarrow a_k[i_0], f_d \leftarrow d_k[i_0 + 1]$ 
     $T_d[f_a].append(f_d), all_d.append(f_d)$ 
    for  $i_a = i_0 + 1$  to  $i$  do
       $blockMatches.append([a_k[i_a]])$ 
      if  $(d[i_a] > d_{min})$  then
         $d_{min} = d_k[i_a]$ 
      end if
       $f_a \leftarrow x_k[a_k[i_a]], T_d[f_a].append(d_{min}), all_d.append(d_{min})$ 
    end for
    for  $r = 0$  to  $t$  do
      if  $len[T_d[r]] \geq c$  and  $quick\_select(T_d[r], c - 1) \leq quick\_select(all_d, c - 1)$  then
        skip

         $i_a \leftarrow i_0$ 
        if  $f_d \leq quick\_select(all_d, c - 1)$  then
           $blockMatches.append([a_k[i_0]])$ 
          for  $i_a = i_0 + 1$  to  $i$  do
            if  $d_k[i_a] \leq quick\_select(all_d, c - 1)$  then
               $blockMatches.append(a_k[i_a])$ 
            end if
          end for
           $d_r = quick\_select(all_d, c - 1)$ 
          report  $blockMatches$  and  $d_r$ 

        end if
         $i_0 \leftarrow i, m = []$ 
        for  $r = 0$  to  $t$  do
           $m.append([])$ 
        end for
        end if
         $updateVQ(x_k[L], i, t, a_k, d_k, pq, V, Q)$ 
      end for
       $a_{k+1} \leftarrow concatenate[V_0, V_1, ..V_t]$ 
       $d_{k+1} \leftarrow concatenate[Q_0, Q_1, ..Q_t]$ 
```
